Methods in Molecular Biology 1480

Springer Protocols

Chiara Lanzuolo
Beatrice Bodega *Editors*

Polycomb Group Proteins

Methods and Protocols

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Polycomb Group Proteins

Methods and Protocols

Edited by

Chiara Lanzuolo

IBCN-CNR, Institute of Cell Biology and Neurobiology and Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi", Rome, Italy

Beatrice Bodega

Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy

Humana Press

Editors
Chiara Lanzuolo
IBCN-CNR
Institute of Cell Biology and Neurobiology and
Istituto Nazionale di Genetica Molecolare
"Romeo ed Enrica Invernizzi"
Rome, Italy

Beatrice Bodega
Istituto Nazionale di Genetica Molecolare
"Romeo ed Enrica Invernizzi"
Milan, Italy

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-6378-2 ISBN 978-1-4939-6380-5 (eBook)
DOI 10.1007/978-1-4939-6380-5

Library of Congress Control Number: 2016948855

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC New York

Preface

Investigating the Polycomb Group of Proteins: Technologies à la Carte

In eukaryotic cells, genomic DNA is packaged in a highly regulated conformation inside the nucleus. This ordered shape consists of multiple levels of epigenetic regulation resulting from dynamic interactions between the genome, chromatin modifiers, and various species of noncoding RNAs. The overall set of DNA, histone modification, and chromatin regulators define an epigenome, which is cell specific and regulates the transcriptome, determining the cell identity (reviewed in [1]).

The Polycomb group of proteins (PcG proteins) is one of the most studied families of transcriptional repressors which act on chromatin at various levels of regulation, from modification of histone tails to modulation of DNA-DNA association (reviewed in [2]). To date, several PcG complexes have been purified and characterized in various organisms, revealing that the combinatorial association of PcG proteins and their co-regulators determines their enzymatic functions and target's specificity. In mammals, the best-characterized complexes are Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2) that can act synergistically or independently of each other and are responsible for the H2AK119ub and H3K27me3 histone signature placements, respectively (reviewed in [3]). In addition to histone modifications, PcG proteins modulate the folding of chromatin in specific higher order structures, which favor the maintenance of genes repression (reviewed in [2, 4]). Hence, in the nuclear space, PcG proteins are organized into aggregates called PcG bodies, mediated by intrinsic and extrinsic protein-protein interactions [1] whose assembly mirrors the chromatin architecture of PcG clustered targets. Interestingly, recent findings have shown that PcG proteins are also able to crosstalk with the nuclear components [5], suggesting that the positioning of PcG bodies in the nucleus could be highly regulated.

Another important aspect of PcG proteins is their dynamism. In fact, PcG proteins are extremely important for lineage commitment, development, and cell differentiation, when a proper timing of gene expression is needed. Upon differentiation stimuli, PcG proteins leave lineage-specific promoters and bind genes important for stemness maintenance [6]. These processes require a highly regulated, coordinated, and fast re-localization of PcG proteins inside the nucleus followed by chromatin remodeling.

Being involved in various biochemical dynamic processes and working at different chromatin levels, PcG biology has inspired several new technical approaches aimed at dissecting the complex molecular mechanisms which together determine their function. Many of these experimental approaches have provided paradigms for the study of chromatin structure and epigenetics in general.

For instance, Chromosome Conformation Capture (3C) technology [7] and its derivative technologies (reviewed in [8]), such as Chromosome Conformation Capture on chip (4C) [9] and High resolution Capture (Hi-C) [10], have been applied in the PcG-related research to shed light on the chromatin contacts occurring in the nucleus and to allow the high-throughput mapping of the genome conformation. The use of these technologies has

led to important advances in understanding PcG functions, demonstrating that the coordinated action of PcG proteins is required to form multi-looped structures where all the major PcG targets are gathered together by *cis* and *trans* interactions [11–14]. These findings were recently confirmed and corroborated by super-resolution microscopy-based studies showing that PcG protein interactions mediate the formation of a characteristic repressive chromatin folding, with a high degree of chromatin intermixing and exclusion of neighboring active chromatin [15, 16].

Hence, our knowledge concerning PcG mode of action is steadily increasing while PcG research is inspiring the development of novel technologies and the appearance of several variations on pre-existing protocols. The current special issue provides a snapshot of the most recent technologies used in the PcG field; scientists working on Polycomb have been invited to contribute with state-of-the-art detailed methods, so as to create a unique and comprehensive reference source for investigating Polycomb function in the nucleus.

Rome, Italy *Chiara Lanzuolo*
Milan, Italy *Beatrice Bodega*

References

1. Bianchi A, LanzuoloC (2015) Into the chromatin world: role of nuclear architecture in epigenome regulation. AIMS Biophys 2(4): 585–612. doi://dx.doi.org/10.3934/biophy.2015.4.585
2. Lanzuolo C, Orlando V (2012) Memories from the polycomb group proteins. Annu Rev Genet . doi: 10.1146/annurev-genet-110711-155603
3. Simon JA, Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 10(10): 697–708
4. Bantignies F, Cavalli G (2011) Polycomb group proteins: repression in 3D. Trends Genet 27(11):454–464.doi:S0168-9525(11)00099-0 [pii] 10.1016/j.tig.2011.06.008
5. Cesarini E, Mozzetta C, Marullo F, Gregoretti F, Gargiulo A, Columbaro M, Cortesi A, Antonelli L, Di Pelino S, Squarzoni S, Palacios D, Zippo A, Bodega B, Oliva G, Lanzuolo C (2015) Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. J Cell Biol 211(3): 533–551. doi: 10.1083/jcb.201504035
6. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 20(9): 1123–1136. doi: gad.381706 [pii] 10.1101/gad.381706
7. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558): 1306–1311
8. de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26(1): 11–24. doi:26/1/11 [pii] 10.1101/gad.179804.111
9. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38(11): 1348–1354. doi: ng1896 [pii] 10.1038/ng1896
10. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950): 289–293. doi:326/5950/289 [pii] 10.1126/science.1181369
11. Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, Cavalli G (2011) Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell 144(2): 214–226. doi: S0092-8674(10)01485-6 [pii] 10.1016/j.cell.2010.12.026
12. Lanzuolo C, Roure V, DekkerJ, Bantignies F, OrlandoV (2007) Polycomb response elements mediate the formation of chromosome higher-

order structures in the bithorax complex. Nat Cell Biol 9(10): 1167–1174. doi: ncb1637 [pii] 10.1038/ncb1637

13. Lo Sardo F, Lanzuolo C, Comoglio F, De Bardi M, Paro R, Orlando V (2013) PcG-mediated higher-order chromatin structures modulate replication programs at the Drosophila BX-C. PLoS Genet 9(2): e1003283. doi: 10.1371/journal.pgen.1003283 PGENETICS-D-12-01096 [pii]
14. Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M, Simonis M, de Laat W, van Lohuizen M, van Steensel B (2011) Interactions among polycomb domains are guided by chromosome architecture. PLoS Genet 7(3): e1001343. doi: 10.1371/journal.pgen.1001343
15. Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu CT, Zhuang X (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature. doi: 10.1038/nature16496
16. Wani AH, Boettiger AN, Schorderet P, Ergun A, Munger C, Sadreyev RI, Zhuang X, Kingston RE, Francis NJ (2016) Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat Commun 7: 10291. doi:10.1038/ncomms10291

Contents

Preface . . . . . *v*
Contributors . . . . . *xiii*

PART I

1 Mapping the Function of Polycomb Proteins 3
Diego Pasini

2 Chromatin Immunoprecipitation *7*
Laura Wiehle and Achim Breiling

3 Chromatin preparation and Chromatin Immuno-Precipitation from *Drosophila* embryos 23
Eva Löser, Daniel Latreille, and Nicola Iovino

4 ChIP-seq Data Processing for PcG Proteins and Associated Histone Modifications. 37
Ozren Bogdanović and Simon J. van Heeringen

5 Analysis of Single-Locus Replication Timing in Asynchronous Cycling Cells. 55
Lo Sardo Federica

PART II

6 Noncoding RNA Interplay with the Genome 69
Davide Gabellini

7 RIP: RNA Immunoprecipitation 73
Miriam Gagliardi and Maria R. Matarazzo

8 Capture Hybridization Analysis of DNA Targets 87
Alec N. Sexton, Martin Machyna, and Matthew D. Simon

9 Identification of RNA–Protein Interactions Through In Vitro RNA Pull-Down Assays 99
Claire Barnes and Aditi Kanhere

10 Understanding RNA-Chromatin Interactions Using *Ch*romatin *I*solation by *R*NA Purification (ChIRP). 115
Ci Chu and Howard Y. Chang

11 Analysis RNA-seq and Noncoding RNA 125
Alberto Arrigoni, Valeria Ranzani, Grazisa Rossetti, Ilaria Panzeri, Sergio Abrignani, Raoul J.P. Bonnal, and Massimiliano Pagani

PART III

12 The Dynamics of Polycomb Complexes. 139
Daniela Palacios

13 Fluorescence Resonance Energy Transfer Microscopy for Measuring Chromatin Complex Structure and Dynamics 143
Alessandro Cherubini and Alessio Zippo

14 Analysis of Endogenous Protein Interactions of Polycomb Group of Proteins in Mouse Embryonic Stem Cells 153
Lluis Morey and Luciano Di Croce

15 Determination of Polycomb Group of Protein Compartmentalization Through Chromatin Fractionation Procedure 167
Federica Marasca, Fabrizia Marullo, and Chiara Lanzuolo

16 An Automatic Segmentation Method Combining an Active Contour Model and a Classification Technique for Detecting Polycomb-group Proteins in High-Throughput Microscopy Images. 181
Francesco Gregoretti, Elisa Cesarini, Chiara Lanzuolo, Gennaro Oliva, and Laura Antonelli

PART IV

17 Polymer Physics of the Large-Scale Structure of Chromatin. 201
Simona Bianco, Andrea Maria Chiariello, Carlo Annunziatella, Andrea Esposito, and Mario Nicodemi

18 Chromosome Conformation Capture in Drosophila 207
Hua-Bing Li

19 Chromosome Conformation Capture in Primary Human Cells 213
Alice Cortesi and Beatrice Bodega

20 Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq) 223
Mélody Matelot and Daan Noordermeer

21 Chromosome Conformation Capture on Chip (4C): Data Processing 243
Benjamin Leblanc, Itys Comet, Frédéric Bantignies, and Giacomo Cavalli

PART V

22 In Vivo Models to Address the Function of Polycomb Group Proteins 265
Frédéric Bantignies

23 A Rapid TALEN Assembly Protocol 269
Arslan Akmammedov, Tomonori Katsuyama, and Renato Paro

24 Following the Motion of Polycomb Bodies in Living Drosophila Embryos . . . 283
Thierry Cheutin and Giacomo Cavalli

25 Reprogramming of Somatic Cells Towards Pluripotency by Cell Fusion. 289
Andrzej R. Malinowski and Amanda G. Fisher

26 Imaginal Disc Transplantation in Drosophila 301
Tomonori Katsuyama and Renato Paro

27 Isolation and Culture of Muscle Stem Cells 311
Chiara Mozzetta

Erratum to: E1

Index *323*

Contributors

Sergio Abrignani • *Istituto Nazionale Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy; Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy*

Arslan Akmammedov • *Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, Milan, Switzerland*

Carlo Annunziatella • *Dipartimento di Fisica, Università di Napoli Federico II and INFN, SPIN-CRN, Naples, Italy*

Laura Antonelli • *Institute for High Performance Computing and Networking, ICAR-CNR, Naples, Italy*

Alberto Arrigoni • *Istituto Nazionale Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*

Frédéric Bantignies • *Institut de Génétique Humaine, CNRS UPR 1142, Montpellier, France*

Claire Barnes • *Department of Biochemistry, University of Leicester, Leicester, UK*

Simona Bianco • *Dipartimento di Fisica, Università di Napoli Federico II and INFN, SPIN-CRN, Naples, Italy*

Beatrice Bodega • *Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy*

Ozren Bogdanović • *ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, Perth, Australia*

Raoul J.P. Bonnal • *Istituto Nazionale Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*

Achim Breiling • *Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg, Germany*

Giacomo Cavalli • *Institut de Génétique Humaine, Montpellier, France*

Elisa Cesarini • *Institute of Cell Biology and Neurobiology, IRCCS Santa Lucia Foundation, Rome, Italy*

Howard Y. Chang • *Center for Personal Dynamic Regulomes, Stanford University, Stanford, USA*

Alessandro Cherubini • *Fondazione Istituto Nazionale di Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*

Thierry Cheutin • *Institut de Génétique Humaine, CNRS UPR 1142, Montpellier, France*

Andrea Maria Chiariello • *Dipartimento di Fisica, Università di Napoli Federico II and INFN, SPIN-CRN, Naples, Italy*

Ci Chu • *Center for Personal Dynamic Regulomes, Stanford University, Stanford, USA*

Itys Comet • *Institut de Génétique Humaine, CNRS UPR 1142, Montpellier, France; BRIC, University of Copenhagen, Copenhagen N, Denmark*

Alice Cortesi • *Genome Biology Unit, Istituto Nazionale di Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*

Luciano Di Croce • *Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG) and Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain*
Andrea Esposito • *Dipartimento di Fisica, Università di Napoli Federico II and INFN, SPIN-CRN, Naples, Italy*
Lo Sardo Federica • *Oncogenomic and Epigenetic Unit, Regina Elena Cancer Institute, Rome, Italy*
Amanda G. Fisher • *Lymphocyte Development Group, MRC Clinical Sciences Centre, Imperial College School of Medicine, Hammersmith Hospital Campus, London, UK*
Davide Gabellini • *Dulbecco Telethon Institute and Division of Regenerative Medicine, San Raffaele Scientific Institute, Milan, Italy*
Miriam Gagliardi • *Institute of Genetics and Biophysics CNR, Naples, Italy*
Francesco Gregoretti • *Institute for High Performance Computing and Networking, ICAR-CNR, Naples, Italy*
Simon J. van Heeringen • *Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands*
Nicola Iovino • *Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany*
Aditi Kanhere • *School of Biosciences, University of Birmingham, Birmingham, UK*
Tomonori Katsuyama • *Department of Genetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan*
Chiara Lanzuolo • *IBCN-CNR, Institute of Cell Biology and Neurobiology and Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy*
Daniel Latreille • *Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany*
Benjamin Leblanc • *Institut de Génétique Humaine, Montpellier, France; BRIC, University of Copenhagen, Copenhagen N, Denmark*
Hua-Bing Li • *Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA*
Eva Löser • *Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany*
Martin Machyna • *Department of Molecular Biophysics & Biochemistry, and Chemical Biology Institute, Yale University, West Haven, USA*
Andrzej R. Malinowski • *Lymphocyte Development Group, MRC Clinical Sciences Centre, Imperial College School of Medicine, Hammersmith Hospital Campus, London, UK*
Federica Marasca • *Istituto Nazionale di Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*
Fabrizia Marullo • *Institute of Cell Biology and Neurobiology, IRCCS Santa Lucia Foundation, Rome, Italy*
Maria R. Matarazzo • *Institute of Genetics and Biophysics, CNR, Naples, Italy*
Mélody Matelot • *Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris Sud, Gif sur Yvette, France*
Lluis Morey • *Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain; Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, FL, USA*
Chiara Mozzetta • *CNR Institute of Cellular Biology and Neurobiology, IRCCS Santa Lucia Foundation, Rome, Italy; Department of Biology and Biotechnology, University of Sapienza, Rome, Italy*

Mario Nicodemi • *Dipartimento di Fisica, Università di Napoli Federico II and INFN, SPIN-CRN, Naples, Italy*
Daan Noordermeer • *Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris Sud, Gif sur Yvette, France*
Gennaro Oliva • *Institute for High Performance Computing and Networking, ICAR-CNR, Naples, Italy*
Massimiliano Pagani • *Istituto Nazionale Genetica Molecolare, Romeo ed Enrica Invernizzi', Milan, Italy; Department of Medical Biotechnology and Translational Medicine, Università degli Studi di Milano, Milan, Italy*
Daniela Palacios • *IRCCS Fondazione Santa Lucia, Rome, Italy*
Ilaria Panzeri • *Istituto Nazionale Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*
Renato Paro • *Department of Biosystems Science and Engineering, University of Basel, Basel, Switzerland*
Diego Pasini • *Department of Experimental Oncology, European Institute of Oncology, Milan, Italy*
Valeria Ranzani • *Istituto Nazionale Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*
Grazisa Rossetti • *Istituto Nazionale Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*
Alec N. Sexton • *Department of Molecular Biophysics & Biochemistry, and Chemical Biology Institute, Yale University, West Haven, CT, USA*
Matthew D. Simon • *Department of Molecular Biophysics & Biochemistry, and Chemical Biology Institute, Yale University, New Haven, CT, USA*
Laura Wiehle • *Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg, Germany*
Alessio Zippo • *Fondazione Istituto Nazionale di Genetica Molecolare, Romeo ed Enrica Invernizzi, Milan, Italy*

Part I

Chapter 1

Mapping the Function of Polycomb Proteins

Diego Pasini*

Abstract

Polycomb group (PcG) proteins are master regulators of proliferation and development that play essential roles in human pathologies including cancers. PcGs act as gatekeepers of cellular identity, maintaining repression of a multitude of target genes. However, these properties have only been recently uncovered thanks to technological advances, first of all chromatin immunoprecipitations (ChIP), that allowed a systematic characterization of the activity of these factors in an unbiased manner at a genome-wide level. Using PcG protein as example, this chapter introduces the readers to the use of chromatin analysis (ChIP assays and replication timing) and how to move these approaches to a level of genome-wide interpretation.

Key words Polycomb, Chromatin, ChIP, Next-generation sequencing, Replication

Polycomb group proteins (PcG) were first identified many years ago via genetic screens in *Drosophila melanogaster* as essential proteins for flies' development. Loss of PcG function resulted in spatiotemporal deregulation of homeotic genes, which result in an aberrant activation of gene expression along the anterior–posterior axis of the developing embryo. For this, PcG proteins were rapidly classified as transcriptional repressors [1, 2].

Taking advantage of Drosophila genetics, several laboratories were able to identify elements (PRE), often placed at long distances from promoters, at which PcG proteins were directly recruited to maintain transcriptional repression. The isolation of these genetic elements became very useful in the pre-genomic era to characterize the means by which PcG proteins are recruited to chromatin and how they repress transcription [3]. However, based on specific staining of polytene chromosomes, it also became immediately clear that PcGs were not simply bound to the few identified PREs, but likely had a much broader occupancy along the fly's genome. This prompted several laboratories to identify consensus DNA sequences within PREs that could predict PcG recruitment. However, these attempts were relatively unsuccessful and immediately suggested that the mechanisms by which PcG proteins are recruited to specific genomic loci were likely more complicated than the simple picture

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_1,

of transcription factors recognizing particular DNA sequences [2, 4]. Although in the last 20 years we have made tremendous improvements in understanding the biological role, the biochemistry and the activity of PcG proteins in different organisms, the recruitment mechanisms still remains an open issue [5]. Indeed, the methodologies described in details in this book describe several approaches by which multiple mechanisms involved in PcG recruitment were in part uncovered.

Another milestone discovery in the PcG field came from the work of Maarten van Lohuizen in the lab of Anton Berns in 1991 that identified the drosophila orthologue of PSC, BMI1 (PCGF4), as a proto-oncogene that cooperates with MYC in inducing lymphomas in mice [6]. This pushed several laboratories to identify and study PcG orthologues in mammalian cells. From this work, it became immediately evident that PcGs functions were highly conserved throughout evolution and this field exploded once it became clear that PcG proteins were controlling development and human pathologies, such as cancer, via posttranslational modification of histone proteins [5, 7].

A great technological boost to study the activity of chromatin-associated proteins in living cells also came from the Polycomb field in the early 90s from the work of Valerio Orlando in Renato Paro's laboratory. They were the first to develop and apply immunoprecipitations of formaldehyde cross-linked chromatin (ChIP), a technique that allows locating in vivo the direct association of proteins with a specific regions of the genome [8]. The power of this technique became immediately clear and still remains one of the most important approaches to study protein association and their activity with chromatin. Although the release of the human and mouse genomic sequences gave a great improvement to study the mechanism of recruitment and action of PcG proteins at several genomic loci, it was only in the post-genomic era, first using chip-based hybridization and more recently with next-generation sequencing (NGS), that we obtained a comprehensive genome-wide view of the activity of PcG proteins. These experiments opened to the understanding that, although PcG multiprotein complexes are highly conserved between flies and mammals, their recruitment mechanisms are likely different as mammalian PcGs exclusively associates with transcriptionally silent (poised) CpG-rich promoter elements. These experiments clearly showed that PcG proteins act as key developmental regulators allowing cells to acquire specific identities by contributing to the establishment of cell-type specific transcription programs [9–11]. These activities act broadly at a genome-wide level and must be precisely coordinated with the cell cycle events that allow the duplication of the genetic information and its repartition upon mitosis to maintain the correct epigenetic information in daughter cells. For this reason, PcG proteins retain different biochemical proprieties that

coordinate the replication timing of the bound loci and allow the rapid reestablishment of PcG-dependent chromatin modifications upon each round of DNA synthesis [12–14].

In this chapter, while the contribution from Wiehle and Breiling thoroughly addresses the basic procedures to perform accurate ChIP assays, the contribution from Nicola Iovino presents a detailed methodology to perform ChIP assays directly on developing drosophila embryos. The outcome of both approaches can be easily used for downstream NGS analyses in order to obtain a genome-wide view of PcG activity. For this, the contribution from Bogdanovic and van Heeringen describes in precise detail the analytical pipelines that are commonly used to analyze ChIPseq results for chromatin associated protein such as PcGs as well as for the products of their enzymatic activity using Histone H3 lysine 27 tri-methylation (H3K27me3) as a representative example of a histone posttranslational modification. Finally, the contribution from Federica Lo Sardo describes the methodology that allows studying the replication timing of a given genomic loci in a bulk population of exponentially growing cells based on the possibility to couple FACS-sorting of cells in different stages of the S-phase with the isolation of newly replicated DNA fragments after a short pulse of BrdU incorporation.

Overall, these techniques provide a comprehensive overview of technical approaches that can be used to study the direct activity of PcG proteins on chromatin at a genome-wide level in relation to the progression of proliferating cells through the cell cycle. Although this chapter specifically focuses on PcG proteins, the same procedures can also be applied to any type of chromatin-bound protein or any kind of chromatin modification.

References

1. Bantignies F, Cavalli G (2006) Cellular memory and dynamic regulation of polycomb group proteins. Curr Opin Cell Biol 18(3):275–283. doi:10.1016/j.ceb.2006.04.003
2. Orlando V (2003) Polycomb, epigenomes, and control of cell identity. Cell 112(5):599–606
3. Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 134(2):223–232. doi:10.1242/dev.02723
4. Ringrose L, Rehmsmeier M, Dura JM, Paro R (2003) Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. Dev Cell 5(5):759–771
5. Scelfo A, Piunti A, Pasini D (2015) The controversial role of the Polycomb group proteins in transcription and cancer: how much do we not understand Polycomb proteins? FEBS J 282(9):1703–1722. doi:10.1111/febs.13112
6. van Lohuizen M, Verbeek S, Scheijen B, Wientjens E, van der Gulden H, Berns A (1991) Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. Cell 65(5):737–752
7. Piunti A, Pasini D (2011) Epigenetic factors in cancer development: polycomb group proteins. Future Oncol 7(1):57–75. doi:10.2217/fon.10.157
8. Orlando V, Paro R (1993) Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. Cell 75(6):1187–1198
9. Di Croce L, Helin K (2013) Transcriptional regulation by Polycomb group proteins.

Nat Struct Mol Biol 20(10):1147–1155. doi:10.1038/nsmb.2669

10. Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469(7330):343–349. doi:10.1038/nature09784
11. Ringrose L (2007) Polycomb comes of age: genome-wide profiling of target sites. Curr Opin Cell Biol 19(3):290–297. doi:10.1016/j.ceb.2007.04.010
12. Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, Rappsilber J, Lerdrup M, Helin K (2008) A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol 10(11):1291–1300. doi:10.1038/ncb1787
13. Lanzuolo C, Lo Sardo F, Diamantini A, Orlando V (2011) PcG complexes set the stage for epigenetic inheritance of gene silencing in early S phase before replication. PLoS Genet 7(11), e1002370. doi:10.1371/journal.pgen.1002370
14. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ 3rd, Voigt P, Martin SR, Taylor WR, De Marco V, Pirrotta V, Reinberg D, Gamblin SJ (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461(7265):762–767. doi:10.1038/nature08398

Chapter 2

Chromatin Immunoprecipitation

Laura Wiehle* and Achim Breiling

Abstract

Chromatin immunoprecipitation (ChIP) is a valuable method to investigate protein-DNA interactions in vivo. Since its discovery it has been indispensable to identify binding sites and patterns of a variety of DNA-interacting proteins, such as transcription factors and regulators, modified histones, and epigenetic modifiers. The Polycomb repressors were the first proteins that have been mapped using this technique, which provided the mechanistic basis for the understanding of their biological function. Cross-linked (XChIP) or native (NChIP) chromatin from tissues or cultured cells is fragmented and the protein of interest is immunoprecipitated using a specific antibody. The co-precipitated DNA is then purified and subjected to analysis by region-specific PCR, DNA microarray (ChIP-on-chip), or next-generation sequencing (ChIP-seq). The assay can therefore produce information about the localization of the analyzed protein at specific candidate loci or throughout the entire genome. In this chapter, we provide a detailed protocol of the basic standard ChIP assay and some remarks about variations.

Key words Chromatin, Cross-linking, DNA-protein interactions, Repressive protein complexes, Immunoprecipitation, ChIP

1 Introduction

The aim of chromatin immunoprecipitation (ChIP) is to analyze interactions of proteins or protein complexes with chromosomal DNA in vivo. The method (schematically summarized in Fig. 1) allows the dynamic visualization of chromosomal proteins in their natural context, through the characterization of their association with specific genomic targets. To this end, the biological material of interest is fixed in vivo with formaldehyde, the cells are lysed, and the chromatin is cut and solubilized. Fixation preserves associations between macromolecules (DNA and proteins and/or proteins and proteins) that would otherwise be lost or disturbed during lysis of the cells, and in particular during the shearing of the chromatin. The resulting chromatin suspension is subjected to immunoprecipitation with an antibody against the protein(s) of interest and the co-immunoprecipitated DNA fragments are analyzed. If the protein under investigation is associated with a specific

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_2,

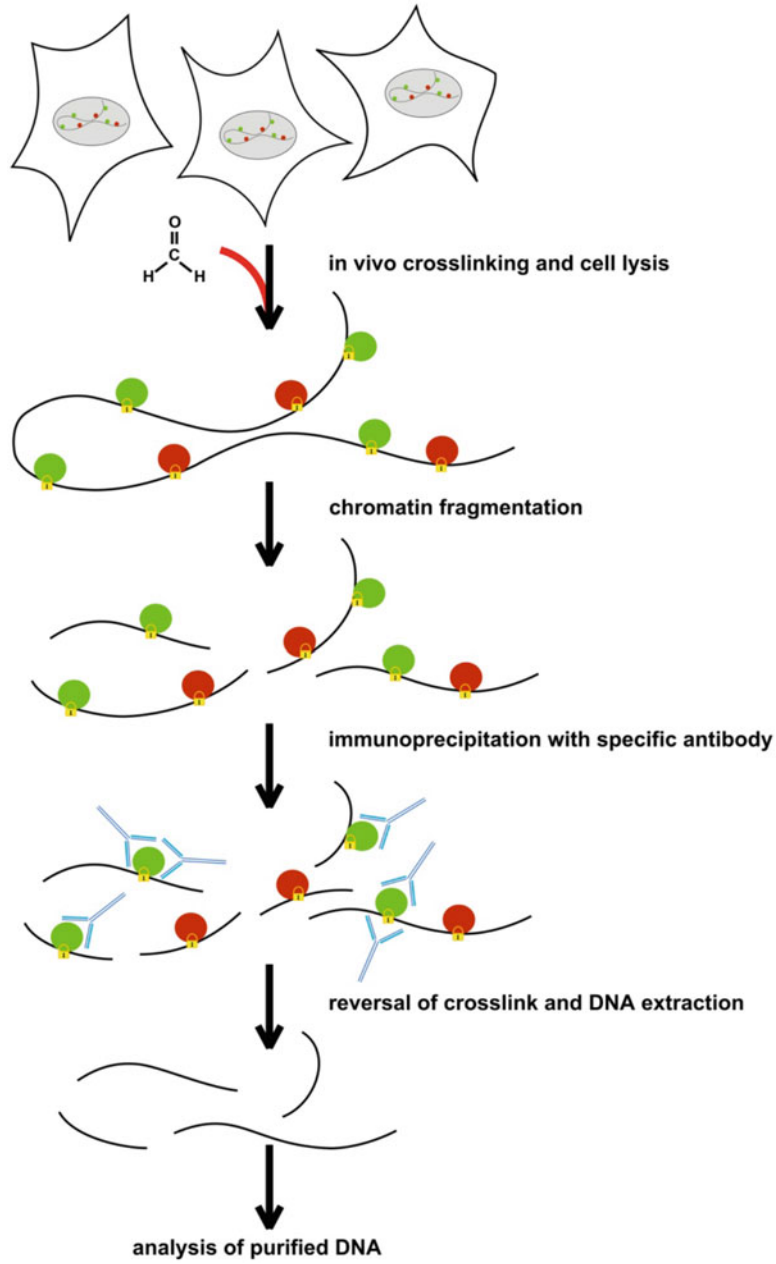


Fig. 1 Scheme of the assay

genomic region in vivo, DNA fragments from this region should be found enriched in the immunoprecipitate. In principle, this method can be applied to any chromosomal protein, as the only prerequisite is the existence of a highly specific antibody. It should be taken into account, though, that it is difficult to determine precise binding sites for a factor, because of the limited resolution of the method.

Importantly, the significance of a ChIP result must be validated by other functional experiments. The occupancy of specific genomic sites by a given protein does not always allow simple extrapolations of protein function. The presence of a factor at a specific DNA sequence just indicates that this protein has access to a particular site, not necessarily what it is doing there. The relevance of any interaction for gene regulation and function has to be confirmed and evaluated by other means.

In a seminal paper published in 1993 Polycomb (PC) was the first protein that was mapped by ChIP to a genomic region, using cultured cells from *Drosophila* [1]. Later on the ChIP technique has been applied to a variety of biological systems like yeast, *Tetrahymena*, *Schistosoma*, mouse tissue culture cells and embryos, various human tissue culture cells, and plant tissues, to map a broad variety of proteins, including several Polycomb group (PcG) members [2]. To initiate a ChIP analysis, cells or tissues are fixed with formaldehyde for an empirically predetermined time (XChIP). Formaldehyde is a very reactive agent that interacts, through its nucleophilic core, with amino and imino groups of proteins (e.g., the epsilon-amino group of lysine) and DNA (side chains in adenine, cytosine, and guanine). No special conditions are required for fixation, as formaldehyde is a small water-soluble molecule that easily penetrates biological membranes. Thus, fixation can be done in vivo, by adding a concentrated stock directly to the living tissue suspended in a standard buffer system, or directly into the culture medium of cultured cells. It is important, though, to avoid buffers that contain primary amines, such as Tris base, as the formaldehyde will react also with these, causing incomplete fixation.

The formaldehyde cross-linking step is the most empirical part of the protocol. Little is known about its specificity and efficiency and for each protein analyzed cross-linking conditions may vary and need to be optimized. For some proteins and chromatin components cross-linking and ChIP analysis has turned out to be difficult or even not possible [3]. Formaldehyde treatment has also been reported to induce changes to the chromatin composition itself, which might lead to reduced ChIP efficiency [4]. Finally, highly expressed genes at least in yeast seem to be vulnerable to nonspecific enrichments of the immunoprecipitated proteins [5]. Therefore, as mentioned above, it is extremely important to confirm the biological relevance of any interaction found using XChIP with independent experiments.

The cross-linking reaction is stopped by adding glycine, which provides an excess of amino groups, thus terminating the fixation (*see* also **Note 3**). The cells are then lysed in a buffer containing physiological salt concentrations and the detergent NP40. This step removes cytoplasmic and membrane proteins, perforates the nuclear membrane, and washes the chromatin to remove non-cross-linked proteins. Depending on the starting material and the cell type

employed, the use of a Dounce homogenizer may be required to aid cell lysis. Finally, the nuclei are pelleted, and then resuspended in a small volume of lysis buffer, containing a high (0.8–1 %) concentration of the detergent sodium dodecyl sulfate (SDS), to induce complete lysis of the nuclei. In addition, SDS helps to efficiently shear the DNA in the next step.

After lysis, a suspension of soluble chromatin is obtained by sonication. This is a very efficient method to cut the chromatin to easily precipitated pieces of 0.3–1 kb in length (*see* Fig. 2). The shorter the DNA fragments, the higher the resolution of the final protein mapping. However, fragments should have a sufficient length allowing the endpoint analysis to work properly, e.g., fragments should not be shorter than the PCR amplicon in case of quantitative PCR readout. Sonication is not the only method that can be used to cut chromatin. Protocols have been developed that employ specific DNA endonucleases, such as micrococcal nuclease, to achieve the same effect (*see*, e.g., [6]), or combinations of sonication and nuclease treatment. A similar approach to DNA shearing is employed in a ChIP method that completely omits the cross-linking step. In native chromatin IP (NChIP—opposed to XChIP including cross-linking), cells are homogenized without prior cross-linking and the chromatin is digested with micrococcal nuclease to mononucleosome resolution. This native chromatin preparation is then used directly for IP, and the co-purified DNA is analyzed [7, 8]. The NChIP method has been used successfully for the analysis of histone modifications. Its advantages are a better chromatin and protein recovery due to higher antibody specificity. Nevertheless, it is mostly not applicable to non-chromatin components, such as transcription factors, regulators, or repressor proteins (including PcG members), as their interactions with the DNA are not stable enough to survive the procedure without cross-linking.

After sonication, the chromatin solution is cleared of debris and insoluble material by centrifugation. Great care must be taken to remove all insoluble material. Any remaining contamination can be pelleted during subsequent steps of the immunoprecipitation, and thus be carried through the rest of the experimental procedure, potentially creating false-positive results. The buffer composition during immunoprecipitation and washing determines the stringency of the analysis. The detergent concentration of SDS in the nuclear lysis buffer is typically too high to allow an efficient interaction between the antibody and the epitope. The solution must therefore be diluted to reduce the SDS concentration. In addition, the salt concentration is increased during this step, bringing the whole solution to IP-buffer conditions. The buffer employed in the following protocol (RIPA-buffer—[9]) usually works well (especially for rabbit polyclonal antibodies) as it contains several ionic and nonionic detergents, and is considered to be very stringent. In general, RIPA buffer seems to prevent aggregation and keeps most

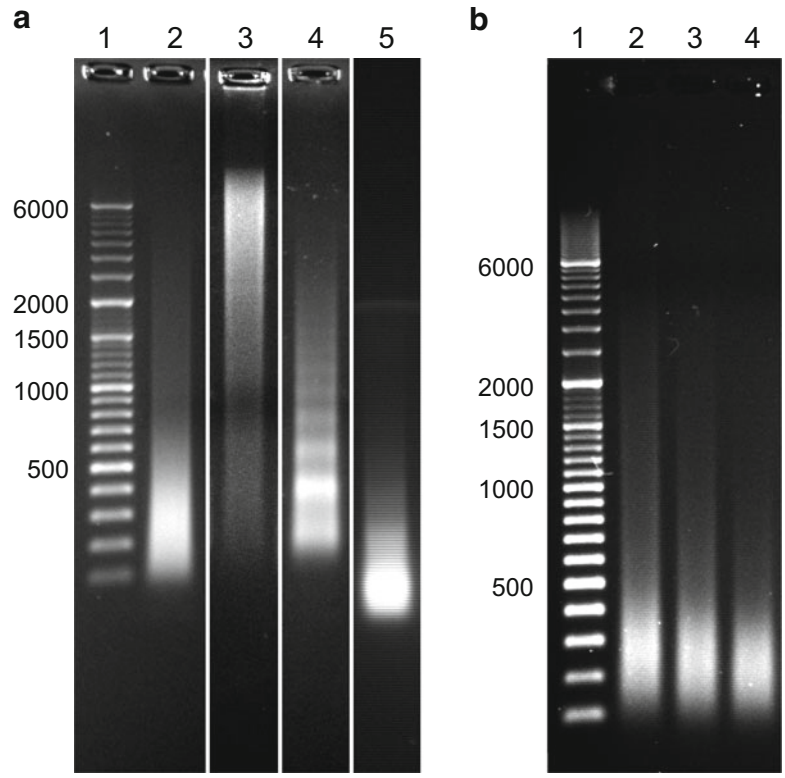


Fig. 2 Chromatin quality control. (**a**) Examples of different chromatin preparations from primary mouse embryonic fibroblasts. Sheared chromatin fragments should show a length between 300 and 1000 bp. The agarose gels show correctly sheared chromatin with most fragments having a size of 200–300 bp (lane 2), insufficiently sheared chromatin (lane 3), properly fragmented chromatin using micrococcal nuclease digestion (lane 4—a nucleosome ladder is clearly visible), and an example of oversheared chromatin (lane 5—most fragments are smaller than 200 bp). Note the high-molecular-weight smear at around 6000 bp and unsheared genomic DNA, seen as a sharp band close to the pocket in lane 3. A molecular weight marker is shown in the first lane. (**b**) Effect of different shearing times on chromatin fragmentation. The longer the chromatin is sheared, the smaller is the resulting fragment size. The agarose gels show DNA fragmented using 20 cycles (alternating on/off shearing for 30 s and 20 s, respectively) on a Covaris S2 sonifier with 5 % duty cycle, intensity 4, 200 cycles per burst (lane 2), 30 cycles (lane 3), and 40 cycles (lane 4). A molecular weight marker is shown in the first lane

proteins in solution. Nevertheless, prior to immunoprecipitation of fixed chromatin, it might be necessary to test the antibodies for their compatibility with the IP buffer. This is done by carrying out an immunoprecipitation of a nuclear extract in RIPA buffer, and analyzing the precipitated material on a Western blot. If the antibody works well under these conditions, it will probably also perform well in this buffer during XChIP. In some cases it may be

necessary to reduce the detergent and salt concentrations in the IP buffer for optimal precipitation results.

The specificity of the antibody used determines the quality and reliability of the result. Antibodies should have a high affinity for the antigen under very stringent conditions, and should not cross-react with other proteins. Thus, the use of affinity-purified antibodies is strongly recommended. Negative control precipitations in the absence of antibody, or with an irrelevant antibody of the same class, are strongly recommended to evaluate the specificity of the results obtained. If no good antibody is available ("chip grade"), a peptide tag can be added to the protein of interest, which is then recombinantly expressed in the system used for the chromatin preparation. The tag should not interact with cellular proteins and should remain accessible to commercially available antibodies raised against it. Nevertheless, it should be noted that over-expression of the tagged protein may cause nonspecific interactions with genomic sites or protein complexes, due to its high, nonphysiological concentration.

There are basically two ways to analyze the ChIP products. Firstly, the immunopurified DNA can serve as a template for fine mapping by quantitative-PCR using primers for specific regions of interest (which will be described in this protocol—*see* Fig. 3). Secondly, the immunopurified material can be used for high-throughput analysis, either by hybridizing to a DNA microarray covering a subset or the complete genome of the model organism (ChIP on chip) or by directly sequencing the immunopurified DNA (ChIP-seq). If the sequence of the target regions is known, the immunopurified DNA can be used directly as template for quantitative PCR, using primer pairs that span the potential binding sites. Amplification of specific DNA fragments will take place only if protein binding occurs; otherwise no amplification will be observed (Fig. 3). Putative "non-binding DNA targets" could be included as controls (Fig. 3). Nevertheless, as PcG proteins tend to spread over larger genomic domains, it might turn out to be difficult to find such a region near the potential genomic target. To increase confidence in the results, regardless of the method of data analysis chosen, the whole ChIP procedure should be performed at least twice for each antibody starting with independent biological replicates (input chromatin). Due to the different precipitation efficiencies of different antibodies, it is difficult to quantitatively compare immunoprecipitations with different antibodies. Therefore, conclusions or comparisons regarding the absolute abundance of various proteins in the same region are not feasible.

The following protocol has been established for adherent mammalian cultured cells. If other tissues are used, certain steps of the protocol may need to be optimized or adjusted, although the basic rationale would be the same. The main difference is likely to be the cross-linking step. Cells growing in suspension can be easily cross-linked by adding a concentrated stock of the cross-linking

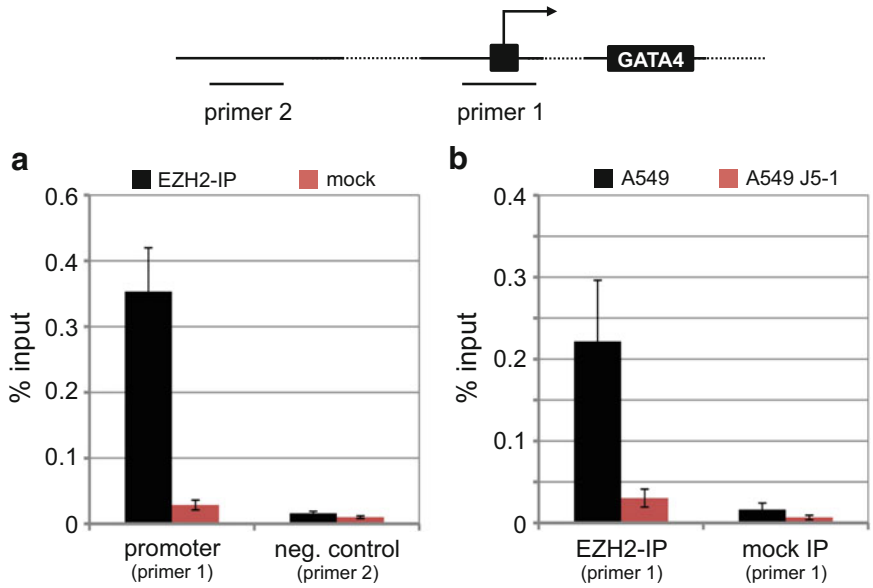


Fig. 3 Example for a ChIP experiment using an antibody for the human PcG protein Enhancer of zeste homolog 2 (EZH2) analyzing the human GATA4 promoter. On top a schematic drawing of the genomic locus and the primer positions are shown. (**a**) ChIP was performed using a commercial EZH2 antibody and chromatin from the adenocarcinomic human alveolar basal epithelial cell line A549 (unpublished data, Breiling). Purified DNA was analyzed using primer pairs amplifying a part of the promoter region of GATA4 (primer 1) and a negative control primer amplifying an intergenic region 34 kB upstream (primer 2). The enrichment of DNA fragments was calculated as percentage of the input. The promoter region showed an enrichment of roughly 0.8 % of the input, whereas the negative control showed no enrichment (no clear difference to the mock). The mock control represents an IP without antibody addition and shows that only minor unspecific enrichment of DNA occurs. (**b**) ChIP assay as in A monitoring the EZH2 occupancy of the GATA4-promoter in A549 cells and in MYC-expressing A549 J5-1 cells that have activated GATA4 [20]. EZH2 is strongly reduced at the active GATA4 promoter in A549 J5-1. Error bars represent standard deviations ($n=3$)

buffer directly to the growth medium as described below. When using more compact material, such as tissues, embryos, or imaginal discs, cross-linking conditions have to be more vigorous and might include initial homogenization steps and treatments with detergents or polar solvents, in order to allow the formaldehyde to better penetrate the sample. For additional protocols, *see*, e.g., [10] (imaginal discs from *D. melanogaster*), [11] (embryos from *D. melanogaster*), [12] (tobacco shoots), [13] (mammalian tissues), or [14] (mouse brain). Finally, specific protocols for low numbers of input cells have been developed, especially carrier ChIP [15] or micro-ChIP [16, 17].

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents.

1. Phosphate-buffered saline (PBS) pH 7.4. Stable at room temperature.
2. Fixation solution: 11% Formaldehyde (from a 37% stock equilibrated with methanol), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 50 mM HEPES (pH 8), proteinase inhibitors, e.g., Complete Protease Inhibitor Cocktail (Roche), 1×. Prepare fresh before use.
3. Glycine (powder, or 1 M stock solution).
4. Cell lysis buffer: 5 mM PIPES (pH 8), 85 mM KCl, 0.5% NP40, proteinase inhibitors, e.g., Complete Protease Inhibitor Cocktail (Roche), 1×. Stable at room temperature; add proteinase inhibitors and put on ice before use.
5. Nuclear lysis buffer: 50 mM Tris–HCl (pH 8), 10 mM EDTA, 0.8% sodium dodecyl sulfate (SDS), proteinase inhibitors, e.g., Complete Protease Inhibitor Cocktail (Roche), 1×. Stable at room temperature; add proteinase inhibitors and put on ice before use.
6. Dilution buffer: 10 mM Tris–HCl (pH 8.0), 0.5 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 140 mM NaCl, proteinase inhibitors, e.g., Complete Protease Inhibitor Cocktail (Roche), 1×. Stable at room temperature; add proteinase inhibitors and put on ice before use.
7. RIPA buffer: 10 mM Tris–HCl (pH 8.0), 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, 140 mM NaCl, proteinase inhibitors, e.g., Complete Protease Inhibitor Cocktail (Roche), 1×. Stable at room temperature; add proteinase inhibitors and put on ice before use.
8. LiCl buffer: 0.25 M LiCl, 0.5% NP40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris–HCl (pH 8.0). Stable at room temperature.
9. TE: 1 mM EDTA, 10 mM Tris–HCl (pH 8.0). Stable at room temperature.
10. Protein A/G agarose beads (50%), pre-swollen and blocked (Santa Cruz, sc-2003).
11. 10% SDS.
12. 20 mg/ml Proteinase K stock. Store at –20 °C.
13. 10 mg/ml RNase-, DNase-free stock. Store at –20 °C.
14. Phenol/chloroform/isoamyl alcohol (25:24:1). Store at 4 °C.
15. Chloroform/isoamyl alcohol (24:1). Prepare fresh before use.
16. 3 M Sodium acetate (pH 5.2).
17. 15 mg/ml Glycogen. Store at –20 °C.
18. 100 and 70% ethanol.
19. Glass beads (150–200 μm, acid-washed).
20. Sonicator (e.g., Sanyo Soniprep 150, exponential microprobe).

3 Methods

3.1 Cross-Linking and Chromatin Shearing

1. Grow cultured cells in an appropriate medium to a near-confluent density or, if suspension cells are used, to a near stationary density (*see* **Note 1**).
2. Add the fixation solution (1/10th of the volume of culture medium—the final formaldehyde concentration should be 1%) directly to the flask or cell culture plate and mix. Incubate fixation reaction for 15 min at room temperature on a shaker (*see* **Note 2**).
3. Stop the fixation by adding glycine powder or a concentrated stock solution to a final concentration of 125 mM (*see* **Note 3**). Mix well and incubate for 5 min at room temperature.
4. Scrape off the cells and transfer them to a conical tube of appropriate size. Pellet cells by centrifuging at 800 × *g* for 5 min at 4 °C. Wash the cells once with ice-cold PBS. Pellet cells again.
5. Resuspend the cell pellet in 15 ml of ice-cold cell lysis buffer, pipetting up and down until all the cells have been resuspended. Place them on ice for 10 min.
6. Collect the nuclei by centrifugation at 2000 × *g* for 5 min at 4 °C. Carefully discard the supernatant and resuspend the pellet in 2 ml of ice-cold nuclear lysis buffer by pipetting up and down. Transfer the suspension to a 15 ml conical tube. Leave on ice for 10 min.
7. Add ~0.5 ml of glass beads to the cell suspension (*see* **Note 4**). Store on ice or sonicate immediately. Sonicate the sample with six 30-s pulses (output near microtip limit), using a high-power sonicator. The sonicator tip should be immersed roughly 1/4 into the liquid. Avoid foaming. If foaming occurs, centrifuge the tube briefly, to reduce the foam layer. Leave on ice for some minutes and sonicate again. Always keep the tube cool by holding it in a beaker containing an ice/water mix (*see* **Note 5**).
8. Transfer the sonicated suspension to a new tube (leaving most of the glass beads behind) and centrifuge for 10 min at 12,000–14,000 × *g* at 4 °C. Dilute the supernatant with dilution buffer to a final volume of 8 ml (i.e., four times dilution, *see* **Note 6**).
9. Rotate the tubes on a wheel for 10 min at 4 °C. Take a 50 µl aliquot to check the average size of the DNA fragments (**steps 10–17**). From the remaining sample, prepare aliquots and store at −80 °C, or use the chromatin directly for immunoprecipitation.
10. To analyze sonication efficiency add 50 µl of TE to the aliquot taken in **step 9** and de-cross-link. Incubate overnight at 65 °C.
11. Add proteinase K to 500 µg/ml and SDS to 0.5% (w/v). Incubate at 50 °C for 3 h. Centrifuge briefly.

12. Add one volume of phenol-chloroform-isoamyl alcohol, vortex for 2 min, and centrifuge at 12,000–14,000 × *g* for 8 min.
13. Transfer the aqueous supernatant to a new tube. Add one volume of chloroform-isoamyl alcohol, vortex for 2 min, and centrifuge at 12,000–14,000 × *g* for 8 min.
14. To the second aqueous supernatant, add 1/10 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol and mix well. Leave at −20 °C for at least 30 min.
15. Centrifuge at 12,000–14,000 × *g* for 15 min at 4 °C. Carefully discard the supernatant and wash the pellet in 800 µl of 70% ethanol. Centrifuge again and allow the pellet to air-dry for 5 min.
16. Dissolve the pellet in 10 µl of TE. Add to each sample 0.5 µg DNase-free RNase, and incubate for 30 min at 37 °C.
17. Add an appropriate amount of gel loading solution. Run the sample on a 0.8% agarose gel and view on a UV transilluminator (*see* Fig. 2). The average length of the DNA should be around 500 bp (most DNA should run as a smear between 300 and 1000 bp).

3.2 Immunoprecipitation and Reversal of Cross-Links

1. Measure the DNA concentration (OD_{260}) of the chromatin using a spectrophotometer (assuming that the absorbance reflects mostly the DNA content). Typically, 100–500 µg of chromatin is used in one immunoprecipitation (IP). For abundant proteins like histones also less chromatin can be used (30–50 µg, *see* **Note 7**). For each IP, the mock-control, and the input-control, dilute the appropriate amount of chromatin in 500 µl of RIPA buffer.
2. Add 25 µl of protein A/G agarose beads (*see* **Note 8**) using a cut pipette tip (wide aperture). Incubate for 1–2 h at 4 °C for pre-clearing and centrifuge in a microfuge at 12,000–14,000 × *g* for 10 min at 4 °C.
3. Transfer the supernatant to a new tube and add the appropriate amount of antibody (*see* **Note 9**), usually 1 µg of affinity-purified antibody (dilutions of 1:100–1:500). Use the same amount of pre-cleared chromatin in the controls, without the addition of antibody (for mock- and input-control), preimmune serum, or an appropriate nonspecific antibody. Incubate the samples from 2 to 3 h to overnight at 4 °C on a rotator or rocker.
4. Centrifuge the samples in a benchtop centrifuge at 12,000–14,000 × *g* for 10 min at 4 °C. Transfer the IPs to new tubes.
5. Add 25 µl of the 50% protein A/G agarose beads (*see* **Note 8**) and incubate for a further 2–4 h.
6. Pellet the beads by centrifugation (500 × *g* for 2 min) in a benchtop centrifuge. Transfer the supernatant of the no-antibody con-

trol to a new tube and leave on ice. This material will serve as total input control. Discard the other supernatants.

7. Wash the beads five times with 600 µl of RIPA buffer, once with 600 µl LiCl buffer and once with 600 µl TE (pH 8.0), collecting the beads between washes with brief centrifugations. Finally, resuspend the beads in 100 µl of TE.
8. Add 1 µg of DNase-free RNase (also to the input-control) and incubate samples overnight at 65 °C.
9. The next day, adjust samples to 0.5% SDS and 0.5 mg/ml proteinase K and incubate for a further 3 h at 50 °C.
10. Add one volume of phenol-chloroform-isoamyl alcohol, vortex for 2 min, and centrifuge at 12,000–14,000 × *g* for 8 min. Back-extract the phenol phase by adding an equal volume of TE (pH 8.0) and vortex. Combine the aqueous phases and add one volume of chloroform-isoamyl alcohol, vortex for 2 min, and centrifuge at 12,000–14,000 × *g* for 8 min.
11. Precipitate the DNA by adding GlycoBlue to 100 µg/ml as carrier, 1/10 volume of 3 M sodium acetate pH 5.2, and 2.5 volumes of 100% ethanol. Incubate at –20 °C for 2 h to overnight.
12. Collect the DNA by centrifugation at 12,000–14,000 × *g* for 15 min at 4 °C, and wash the pellet in 800 µl of 70% ethanol. Repeat centrifugation and discard the supernatant (*see* **Note 10**).
13. Allow the pellet to air-dry for 5 min. Redissolve the precipitated DNA in 50 µl of TE and store at 4 °C (to avoid DNA precipitation, do not freeze).

3.3 PCR Analysis

1. Real-time PCR is the preferred method for analyzing known target regions. PCR amplification is easiest performed using a ready-to-use SYBR green mix in 10 µl reactions on 384-well PCR plates, using 0.25 µl of 10 µM forward and reverse primers per well (see Note 11). Start the analysis using 1 µl of the purified immunoprecipitated DNA and from the 1:100 and 1:1000 diluted input. If too high Ct values are obtained with the ChIP samples, more template should be used per reaction. Results are usually presented by calculating the "percent input," indicating how much of the initial input DNA representing the target region was co-purified during the procedure. Examples for finished analyses are shown in Fig. 3.

4 Notes

1. The efficiency of immunoprecipitation (IP) is normally quite low. Therefore, in order to obtain enough material to analyze by PCR or subsequent high-throughput analyses, each ChIP

should start with a minimum of 1×10^6 (up to 1×10^7) mammalian cells. Nevertheless, the best starting number can vary for the proteins analyzed (depending on their abundance and frequency of interaction) and should be determined empirically and optimized for each specific factor.

2. Usually, a final formaldehyde concentration of 1% works well for fixation. Use of this concentration and varying the cross-linking time and temperature are likely to yield optimal results. Excessive cross-linking, or insufficient exposure to formaldehyde, may result in failure to detect specific interactions. If cells are cross-linked for too short time, fixation of the protein to its target will be incomplete and the interaction will be lost during the subsequent steps of the protocol. An overly long fixation time increases the number of protein-protein cross-links, which reduces the efficiency of chromatin shearing and might block the epitope of the antibody used for ChIP. In principle, longer cross-linking is desirable to investigate proteins that do not bind DNA directly or as part of a bigger complex (like PcG proteins), whereas shorter incubation with formaldehyde is favorable if the protein contacts the DNA directly (like histones). Usually, cross-linking times of 5–15 min at room temperature are sufficient. In any case, trial experiments with different cross-linking times (e.g., 5, 10, 15, and 30 min) and temperatures (room temperature, 4 °C—cross-linking will be considerably slower at 4 °C) are recommended when starting a ChIP analysis or using a new antibody.
3. Quenching using glycine has been found to partly inhibit downstream enzymatic reactions, e.g., nuclease treatments [18]. Glycine can therefore be replaced by Tris (e.g., Tris, pH 8.0, at a final concentration of 250 mM), which has also been shown to be more efficient in stopping formaldehyde cross-linking reactivity [19]. Interestingly, the major mode of action of both quenching agents seems to be the removal of unreacted formaldehyde from the solution [19].
4. To aid sonication, and to reduce the number of pulses necessary to obtain short DNA pieces, it is helpful (but not necessary) to add 1/4 volume of glass beads (acid washed, 200 μm or less in diameter) prior to sonication. For sonication also other devices than a standard sonicator can be used, e.g., a Bioruptor (bath sonicator) or a Covaris ultrasonicator (*see* Fig. 2b). The volume of nuclear lysis buffer the nuclei are taken up prior to sonication has then to be reduced. In case, the nuclei pellet can be split into separate aliquots. It is vitally important to cool the sample during sonication, as the probe heats up during emission and exposure to heat might result in partial de-cross-linking of the material.

5. The sonication time, amplitude, and number of pulses required will vary depending on the type of sonicator used, the cell type, and the extent of cross-linking. In general, the longer the material has been cross-linked, the more difficult it is to sonicate the DNA into small pieces (*see* Fig. 2). It is always recommended to check the average size of the chromatin prior to continuing with the immunoprecipitation. To do this, the sample is de-cross-linked and the DNA length is analyzed on a conventional agarose gel with ethidium bromide staining (**step 7**, in Subheading 3.1—Fig. 2). The fragmentation profile can also be analyzed easily and very precisely using an Agilent Bioanalyzer or a TapeStation, which requires less material. To establish the best sonication conditions for the cell type used and the cross-linking time employed, trial experiments can be performed: After each sonication pulse, an aliquot is taken from the chromatin suspension. The samples are de-cross-linked and the average DNA length is analyzed. The minimum number of sonication pulses is chosen to shear most of the DNA to pieces in the range of 300–1000 bp.
6. If dilution is not sufficient to achieve an SDS concentration allowing for the antibody to work, samples can be dialyzed (into RIPA buffer) or small centrifugal ultrafilters (e.g., from amicon, 30 kDa) can be used to exchange buffers.
7. To avoid spectrophotometric DNA concentration measurement of cross-linked chromatin, which might be imprecise, the amount of DNA in the chromatin preparation can also be extrapolated from the aliquot taken for analysis of chromatin fragmentation. After reversal of cross-link and DNA purification measure the concentration using a microvolume spectrophotometer and calculate the amount in the total chromatin preparation. Moreover, the amount of used chromatin should also be adjusted by counting the cells before cross-linking and using chromatin preparations isolated from the same number of cells for each IP.
8. The antibody and any associated DNA-protein complex is trapped with protein A, G, or L agarose or sepharose beads. To prevent nonspecific interactions, the conjugated beads can be pre-blocked with BSA (1 mg/ml) and also with a nonspecific DNA (e.g., salmon sperm DNA, 0.25 mg/ml). It is also possible to use magnetic particles coated with protein A, G, or L, which simplifies washing steps. Beads coated both with protein A and G, which will bind most polyclonal or monoclonal antibodies, might be the best choice. In any case, a pre-clearing step with the particles of choice to fish out any proteins or DNA that nonspecifically interact with the beads before antibody addition should be performed.

9. Immunofluorescence-grade antibodies are preferred, as their epitopes are unlikely to be affected by formaldehyde fixation. Many companies also offer "chip-grade" antibodies that have already been tested for the application. The amount of chromatin and antibody used and the length of incubation must be determined empirically (though overnight incubation is mostly more than sufficient). The final signal strength is proportional to the quantity of antibody and chromatin used. Thus, the ideal amount of both that leads to optimal signals without wasting antibody or chromatin should be determined.
10. To avoid phenol-chloroform extraction and ethanol precipitation after the proteinase K treatment, the DNA can also be purified using spin columns or magnetic bead-based systems (like AMPure beads).
11. Primer pairs (melting temperature 64–66 °C, around 25 bp in length), amplifying 150–300 bp fragments in the target region of interest, are designed with the appropriate software. Care must be taken when designing real-time PCR primers, so that only the correct PCR product is synthesized. In particular, the formation of primer dimers must be avoided, as these would compromise Ct values and lead to false positives.

References

1. Orlando V, Paro R (1993) Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. Cell 75:1187–1198
2. Orlando V (2000) Mapping chromosomal proteins in vivo by formaldehyde-cross-linked-chromatin immunoprecipitation. Trends Biochem Sci 25:99–104
3. Gavrilov A, Razin SV, Cavalli G (2014) In vivo formaldehyde cross-linking: it is time for black box analysis. Brief Funct Genomics pii: elu037
4. Beneke S, Meyer K, Holtz A, Hüttner K, Bürkle A (2012) Chromatin composition is changed by poly(ADP-ribosyl)ation during chromatin immunoprecipitation. PLoS One 7, e32914
5. Teytelman L, Thurtle DM, Rine J, van Oudenaarden A (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. Proc Natl Acad Sci U S A 110:18602–18607
6. Cirillo LA, Zaret KS (2004) Preparation of defined mononucleosomes, dinucleosomes, and nucleosome arrays in vitro and analysis of transcription factor binding. Methods Enzymol 375:131–158
7. O'Neill LP, Turner BM (2003) Immunoprecipitation of native chromatin: NChIP. Methods 31:76–82
8. Cosseau C, Grunau C (2011) Native chromatin immunoprecipitation. Methods Mol Biol 791:195–212
9. Orlando V, Strutt H, Paro R (1997) Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods 11:205–214
10. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298:1039–1043
11. Orlando V, Jane EP, Chinwalla V, Harte PJ, Paro R (1998) Binding of trithorax and Polycomb proteins to the bithorax complex: Dynamic changes during early Drosophila embryogenesis. EMBO J 17:5141–5150
12. Chua YL, Mott E, Brown AP, MacLean D, Gray JC (2004) Microarray analysis of chromatin-immunoprecipitated DNA identifies specific regions of tobacco genes associated with acetylated histones. Plant J 37:789–800
13. Turner FB, Cheung WL, Cheung P (2006) Chromatin immunoprecipitation assay for mammalian tissues. Methods Mol Biol 325:261–272
14. Sailaja BS, Takizawa T, Meshorer E (2012) Chromatin immunoprecipitation in mouse hippocampal cells and tissues. Methods Mol Biol 809:353–364

15. O'Neill LP, VerMilyea MD, Turner BM (2006) Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. Nat Genet 38:835–841
16. Acevedo LG, Iniguez AL, Holster HL, Zhang X, Green R, Farnham PJ (2007) Genome-scale ChIP-chip analysis using 10,000 human cells. Biotechniques 43:791–797
17. Dahl JA, Collas P (2009) MicroChIP: chromatin immunoprecipitation for small cell numbers. Methods Mol Biol 567:59–74
18. Wu CH, Chen S, Shortreed MR, Kreitinger GM, Yuan Y, Frey BL, Zhang Y, Mirza S, Cirillo LA, Olivier M, Smith LM (2011) Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification. PLoS One 6, e26217
19. Sutherland BW, Toews J, Kast J (2008) Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. J Mass Spectrom 43:699–715
20. Castro I, Breiling A, Luetkenhaus K, Ceteci F, Hausmann S, Kress S, Lyko F, Rudel T, Rapp UR (2013) MYC-induced epigenetic activation of GATA4 in lung adenocarcinoma. Mol Cancer Res 11:161–172

Chapter 3

Chromatin Preparation and Chromatin Immuno-precipitation from *Drosophila* Embryos

Eva Löser*, Daniel Latreille*, and Nicola Iovino*

Abstract

This protocol provides specific details on how to perform Chromatin immunoprecipitation (ChIP) from *Drosophila* embryos. ChIP allows the matching of proteins or histone modifications to specific genomic regions. Formaldehyde-cross-linked chromatin is isolated and antibodies against the target of interest are used to determine whether the target is associated with a specific DNA sequence. This can be performed in spatial and temporal manner and it can provide information about the genome-wide localization of a given protein or histone modification if coupled with deep sequencing technology (ChIP-Seq).

Key words Chromatin immunoprecipitation, *Drosophila* embryos, ChIP-Seq

1 Introduction

Chromatin is a complex of macromolecules found in the cell nucleus, consisting of DNA, protein, and RNA. The primary protein components of chromatin are histones that compact the DNA and control the underlying genomic sequences' transcriptional output. ChIP is an invaluable method for studying interactions between specific proteins or modified forms of proteins and a genomic DNA region. This technique has been very useful to study the mechanisms of gene expression, histone modification and transcriptional regulation. The first ChIP assay was developed by Gilmour and Lis [1] as a technique to monitor the association of RNA polymerase II with transcribed and poised genes in *Drosophila*.

This protocol is intended to provide general guidelines, experimental settings and conditions for ChIP in *Drosophila* embryos.

The original version of this chapter was revised. The erratum to this chapter is available at: DOI 10.1007/978-1-4939-6380-5_28

*Author contributed equally to this work.

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_3, © Springer Science+Business Media New York 2016

Briefly, protein–DNA interactions are stabilized by formaldehyde cross-linking freshly collected embryos. Chromatin is isolated from purified nuclei and subsequently sheared using sonication. A pre-clearing step prior to immunoprecipitation reduces nonspecific background. Immunoprecipitation is performed with a specific antibody followed by the addition of antibody-affinity beads to capture the antibody–antigen complex. Finally, the DNA fragments are purified and analyzed by qPCR (quantitative PCR) and/or by deep sequencing.

ChIP can also be performed in the absence of cross-linking (referred to as native ChIP or N-ChIP) to examine proteins that remain stably associated with DNA during chromatin processing and immunoprecipitation, a protocol not described in this method section [2].

Specific optimization might be required if antibody and instruments differ from those described here. Specifically, the antibody to chromatin ratio, as well as sonication conditions have a strong influence on the quality of later analysis and might require optimization.

2 Material and Reagents

See **Note 1**.

2.1 Embryo Collection

Standard equipment for fly husbandry and maintenance is required.

1. Embryo collection cages (compatible with 60 mm petri dishes are suitable for our approach). For large-scale collection, cages compatible with 100 mm petri dishes are recommended.
2. Petri dishes diameter fitting with fly cages.
3. Magnetic/heating stirrer.
4. Apple juice agar plates:
   Dissolve 12.5 g sucrose in 125 mL apple juice. Dissolve 15 g agar in 375 mL dH_2O by autoclaving (apply magnetic stirrer). Directly after autoclaving pour apple juice into the hot agar under constant stirring. Distribute into petri dishes. Let them dry with closed lid overnight on the bench.
5. Yeast paste:
   Mix dry baker's yeast with a bit of water in a beaker and stir with a metal spatula into a smooth paste. The consistency of the paste should not be too watery otherwise flies will stick to it. Keep the yeast paste at 4 °C.
6. 50% "bleach" solution using sodium hypochlorite (6–14% active chlorine).
7. 70-μm pore size sieve or homemade mesh baskets for embryo collection.
8. Soft paint-brush.

9. Distilled water.
10. Paper towel.
11. Stereo microscope with transmitted light.
12. Cooling/heating chamber for microscopes.
13. External temperature control system (heating–cooling circulator).
14. Glass chamber slide, single well.
15. Aspiration hand set with adjustable tip holder to accept a variety of tips and thumb-wheel control for microliter aspiration.

2.2 Protein–DNA Cross-Link

1. Buffer A: 15 mM HEPES pH 7.9, 60 mM KCl, 15 mM NaCl, and 4 mM $MgCl_2$.
2. Buffer A-TX: Buffer A supplemented with Triton X-100 to 0.1% (v/v) final.
3. Two-phase fixing solution (prepare freshly before use): (*see* **Note 2**)

   Dilute formaldehyde 16% methanol-free solution in Buffer A to 1% (v/v) final. Add equal volume of heptane on top.
4. 2.5 M of Glycine solution.
5. Screw cap glass tube 24×100, conical.
6. Glass Pasteur pipette.
7. Orbital Shaker equipped with tension rollers attachment.

2.3 Chromatin Extraction

1. Buffer A-TBP (prepare freshly before use): Buffer A supplemented to a final concentration with 0.5% (v/v) Triton X-100, 0.5 mM DTT, 10 mM sodium butyrate, 1× EDTA-free protease inhibitor cocktail.
2. Lysis Buffer basic (prepare freshly before use): 15 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA pH 8, 0.5 mM EGTA, 1% (v/v) Triton X-100, 0.5 mM DTT, 10 mM sodium butyrate, 0.1% (w/v) sodium deoxycholate, 1× EDTA-free protease inhibitor cocktail.
3. Lysis Buffer S1: Lysis Buffer basic supplemented with 0.1% (w/v) SDS, 0.5% (w/v) *N*-lauroylsarcosine.
4. Lysis Buffer S2: Lysis Buffer basic supplemented with 0.05% (w/v) SDS.
5. Dounce tissue grinder with "tight" pestle (1 mL/7 mL).
6. Protein Low Binding tubes.
7. Refrigerated tabletop centrifuge.
8. Rotating wheel.
9. E220 Focused-ultrasonicator (Covaris).
10. Tube AFA Fiber & Cap 12×12 mm (Covaris) compatible tubes for E220 Focused-ultrasonicator.

11. Refrigerated tabletop centrifuge.
12. Qubit fluorometer.
13. 2100 Bioanalyzer.

2.4 Affinity Purification of Protein–DNA Complexes

1. Agarose beads (Protein A or Protein G).
2. Antibodies validated for immunoprecipitation of cross-linked material.

   This protocol is optimized for the rabbit anti H3K27me3 polyclonal antibody-premium from Diagenode C15410195.
3. Isotype matched control immunoglobulin.
4. TE-0.1 Buffer: 10 mM Tris pH 8, 0.1 mM EDTA.
5. Elution Buffer 1: 50 mM Tris pH 8, 10 mM EDTA, 1% (w/v) SDS.
6. Elution Buffer 2: TE-0.1 Buffer containing 0.67% (w/v) SDS.
7. Thermomixer.

2.5 DNA Isolation

1. DNA Low Binding tubes.
2. Proteinase K solution 20 mg/mL.
3. RNase A (DNase and protease free).
4. Phenol–chloroform–isoamyl alcohol (25:24:1).
5. Phase Lock Gel 1.5 mL, Heavy.
6. Sodium acetate Buffer solution 3 M, pH 5.2.
7. 100% EtOH p.a.
8. 70% (v/v) EtOH p.a.
9. Glycogen 5 mg/mL.
10. High Sensitivity DNA Kit, Agilent.
11. Primer sets:

    5′engrailed GGCTTGTTAGGCAGCAATATGAC
    3′engrailed TGAACAGTGCCGCTATATGACC
    5′antp-PRE TGGCCGAGTTTATATCGAAGCG
    3′antp-PRE CGGCCAACTTGTGTTGTTGTTC
    5′vg-prom GTTTCCAGGTTTCCAACTAACG
    3′vg-prom AGAAACAGCAGGAGTTTCGTCT
    5′RpL32-TSS TCCAGCTTCAAGATGACCATCC
    3′RpL32-TSS CTTGTTCGATCCGTAACCGATG
    5′PGRP-LE-TSS TGAGCTTTCCAACACTCTTGC
    3′PGRP-LE-TSS GGTTTTGGTGGTTTATCTGAG

3 Methods

ChIP experiments aim to address the question whether a protein—or a modified form of a protein—has a direct interaction with DNA. This is achieved by quantifying the interaction between a protein and a specific locus (ChIP-qPCR) or mapping its interaction genome wide by ChIP followed by deep-sequencing (ChIP-seq) of DNA-bound to it. Two key steps of a ChIP protocol have to be considered and optimized before conducting an experiment: First, the starting amount of nuclei/chromatin has to be determined based on antibody IP efficiency. Second the size of DNA fragments released after sonication should be in a range of 200 to 600 bp. The flowchart is presented in Fig. 1.

3.1 Embryo Collection

Stage 5 Drosophila embryos contain around 4000 nuclei. Usually, up to 1 μg of chromatin can be recovered from 1 million nuclei. This amount of starting material is enough to perform

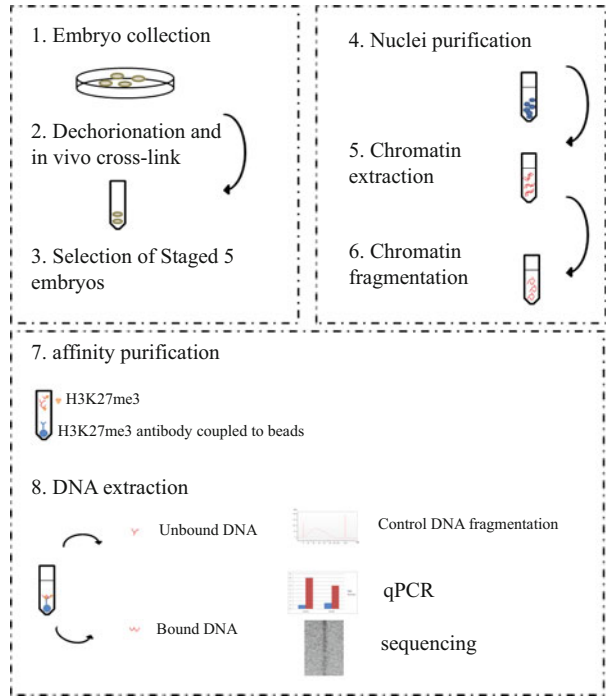


Fig. 1 Flowchart of Chromatin Immuno-Precipitation in *Drosophila* embryos. The scheme highlights the major steps of the protocol, from the selection of starting material (e.g., stage 5 embryos), through the extraction of size-specific chromatin fragments, to their specific enrichment after immunoprecipitation. Eluted DNA fragments are further analyzed by qPCR to characterize an interaction with a locus of interest or by deep-sequencing to assess the genome-wide chromatin binding profile of a protein of interest

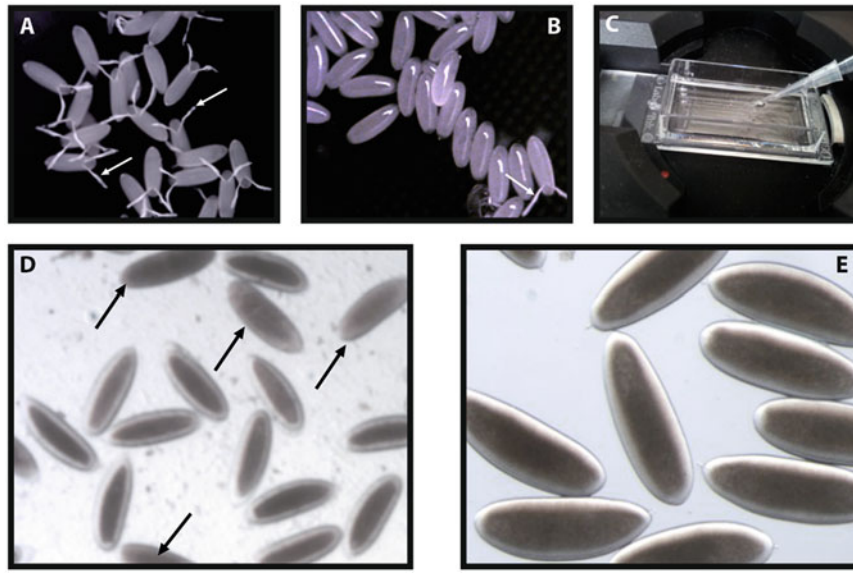


Fig. 2 Selection of stage 5 embryos. Collection of embryos before dechorionation (**a**) and after dechorionation (**b**) under light microscope. The efficiency of dechorionation can be monitored by the loss of appendages on embryos (*white arrow*, **a** and **b**). (**c**). Embryos are transferred to a cooling system to discard earlier and older embryos (**d**. *black arrows*). (**e**). Typical enrichment of stage 5 embryos after selection can be observed under light microscope

H3K27me3-IPs (and appropriate controls) for both qPCR and deep-sequencing.

Embryos are hand-sorted after fixation and stage 5 embryos are identified by their characteristic morphology [3, 4] according to Campos-Ortega [5] (Fig. 2).

1. To reach the maximum egg production, maintain flies for 2 days in egg-laying cages at 25 °C and 70% humidity. Change apple juice plates with a spot of fresh yeast paste placed in the middle two to three times a day (egg-laying is stimulated by continuous feeding) (*see* **Note 3**).
2. Before collection, change the apple juice plates with fresh yeast paste every hour for three times to achieve synchronized egg-laying (*see* **Note 3**).
3. To obtain stage 5 embryos, collect embryos for 30 min at 25 °C. Incubate the collected plate for additional 130 min at 25 °C (*see* **Note 4**).

   The collection and dechorination procedure is performed in a single step (*see* **Note 5**).
4. Remove dead flies and the excess of yeast from the apple juice plate.

   Add 50% "bleach" directly onto the plate for maximum 2 min (dechorionation efficiency can be checked under a stereo microscope) (*see* **Note 6**).

5. Collect the dechorionated embryos by pouring the suspension through a collection sieve while actively rinsing the plate with water. (Repeat until all embryos are collected and remaining yeast is removed).
6. The collected embryos are extensively washed with a stream of water from a squirt bottle to remove all traces of bleach.
7. The sieve is finally blotted on a paper towel to remove any remaining liquid.

3.2 Fixation

1. Prepare 10 mL of two-phase cross-linking solution in a screw-cap glass tube (*see* **Note 2**).
2. Immediately transfer the embryos from the sieve to the heptane phase with a soft paintbrush soaked in heptane. The embryos will gather at the interphase between PFA-solution and heptane.
3. Cap the tube tightly, shake briefly by hand and then clip the tube horizontally on a platform orbital shaker to increase the heptane–fixative interface boundary, agitate vigorously for 15 min at 250 to 300 rpm at room temperature. Do not exceed 15 min cross-linking (*see* **Note 7**).
4. Stop the cross-linking reaction by addition of glycine to a final concentration of 0.25 M to quench the formaldehyde. Incubate for 5 min while rotating at room temperature.
5. Remove the two-phases cross-linking solution using a Pasteur pipette starting with the lower aqueous phase and then the organic phase.
6. Add 10 mL Buffer A-Tx to the embryos. Incubate for 5 min while rotating at low speed to avoid foaming. Let embryos settle to the bottom of the tube. (The embryos tend to stick to the inside of the cap - make sure to recover them.) Remove supernatant. Repeat wash with 10 mL Buffer A-Tx. Remove supernatant.
7. Transfer embryos in Buffer A-Tx to a precooled chamber slide fitting in a precooled chamber for microscopes connected to an external temperature control system.
8. Identify incorrect stages by morphology and remove contaminating embryos with an aspiration hand set with thumb-wheel control aspiration (Fig. 2).
9. Transfer the selected embryos with glass Pasteur pipette to a 1 mL-Wheaton Dounce tissue grinder precooled on ice. Let settle embryos down.

3.3 Chromatin Extraction

From now on, all the procedure should be performed at 4 °C with precooled Buffers except if other recommendations are mentioned. This paragraph includes embryo lysis, nuclei purification as well as chromatin extraction and fragmentation.

1. Remove Buffer and wash with 1 mL ice-cold A-TBP Buffer, let embryos settle down.
2. Remove Buffer and add 500 µL ice-cold A-TBP Buffer.
3. Homogenize with strokes (10 times) until all embryos are destroyed. From this step onward, use 1.5 mL Protein Low Binding tubes.
4. Transfer homogenate to precooled 1.5 mL tube.
5. Wash Dounce homogenizer and pestle with another 500 µL of ice-cold Buffer A-TBP.
6. Combine with homogenate to final volume of 1 mL.
7. Pellet the nuclei by centrifugation at 3200 × *g* for 10 min at 4 °C.
8. Carefully remove supernatant and resuspend the pellet in Lysis Buffer basic.
9. Repeat this step once or twice until the pellet appears clear.
10. Finally resuspend pellet in 1 mL of Lysis Buffer S1. Incubate for 1 h on rotating wheel at 4 °C.
11. Chromatin shearing is performed using the Covaris E220 Focused-ultrasonicator (*see* **Note 8**).
    1. Settings: Peak power 140 W; Duty factor 5 %; Cycle/burst 200 for 15 min.
12. Centrifuge samples for 10 min at 20,000 × g at 4 °C. Transfer supernatant to a new precooled 1.5 mL tube referred to as Chromatin Extracts.

3.4 DNA Fragment Sizes Control

Confirm sonication efficiency prior to IP in case of first time ChIP, before proceeding. For optimal results compare DNA size of RNase treated samples (*see* **Note 9**)

Once the chromatin shearing step is standardized, we always use the supernatant after IP (S-IP) to perform the DNA fragment size control *see* Subheading 3.6, **step 14**

1. Add Proteinase K to a final concentration of 200 µg/mL. Incubate at 65 °C shaking at 300 rpm overnight to reverse the formaldehyde cross-link.
2. Allow samples to cool down to room temperature before proceeding with RNase treatment.
3. Add RNase A 50 µg/mL final to the samples. Incubate for 30 min at 37 °C.
4. Go to 3.7 to perform DNA extraction and precipitation.
5. Control sheared chromatin size either on a 1.5 % agarose gel or on Bioanalyzer 2100 using High Sensitivity DNA chip according to manufacturer's protocol. The optimal size range of DNA for ChIP-seq analysis should be between 200 and 600 base pairs (Fig. 3).

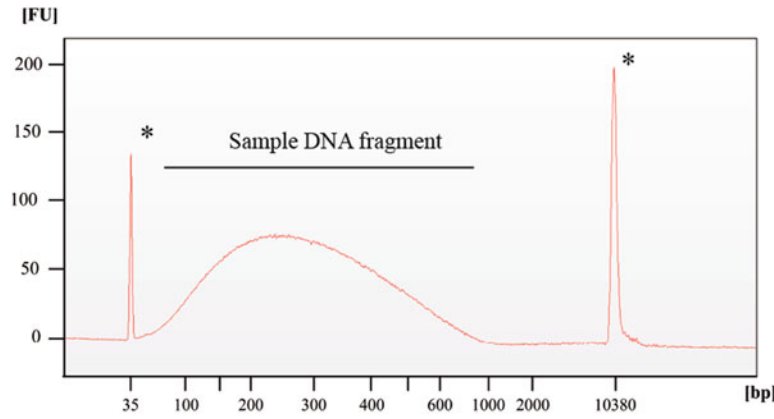


Fig. 3 Typical electropherogram of DNA fragments from input sample analyzed by Bioanalyzer 2100. DNA fragments are electrophoretically driven by a voltage gradient. Dye molecules coupled to DNA are detected by fluorescence. The size of DNA fragments is proportional to the migration time. Two internal references are included (indicated as a *star*). The average size of DNA fragments is between 150 bp and 500 bp

3.5 Overnight Pre-clearing

1. Calculate 50 µL of homogeneous 50% agarose slurry (25 µL beads volume) per mL of chromatin extracts (*see* **Note 10**).
2. Equilibrate beads with 1 mL (at least 10 bead volumes) of Lysis Buffer S1. Pellet beads by centrifugation 1500×*g* for 3 min at 4 °C. Repeat twice. Resuspend beads in an equal volume of S1 Buffer (50 µL).
3. Add 50 µL of beads per mL of chromatin extract. Incubate at 4 °C on a rotator wheel overnight.
4. Pellet the beads by centrifugation 1500×*g* for 3 min at 4 °C. Transfer supernatant (1 mL) to a fresh precooled 1.5 mL tube referred to as pre-cleared chromatin.

3.6 Affinity Purification of Protein–DNA Complexes

For this approach we first incubate the antibody with the chromatin extract followed by the addition of affinity beads to capture the antibody-antigen complex.

A negative control is necessary to determine the background level. We use isotype matched control immunoglobulin to reveal the (background) nonspecific DNA enrichment.

1. Split pre-cleared chromatin extract into 2 fractions: 100 µL referred to as mock-IP and 900 µL as IP.
2. Add antibody to samples, e.g., for H3K27m3-IP we use 1 µg antibody per 1 mL chromatin extract (based on our starting material of 1 million nuclei) and 0.1 µg of normal rabbit IgG per 100 µL Mock-IP sample respectively. Incubate at 4 °C for up to 7 h on a rotating wheel (*see* **Note 11**).

3. Equilibrate beads in Lysis Buffer S1 as in Subheading 3.5, **steps 1** and **2**. Mix slurry well and add 50 µL to the sample and 5 µL (1/10) to the Mock-IP respectively. Incubate overnight at 4 °C on a rotating wheel.
4. After incubation centrifuge the tubes 400 × *g* for 3 min at 4 °C.
5. Transfer supernatants to new tubes renamed as Supernatant-IP (S-IP) and Supernatant-mock referred to as INPUT (*see* **Note 12**). The S-IP can be later used to control the chromatin shearing (go to Subheading 3.6, **step 14**) Samples can be either snap-frozen in liquid nitrogen then stored at −80 °C or kept on ice until Subheading 3.6, **step 14**.
6. Wash beads with 1 mL of Lysis Buffer S2. Incubate for 10 min at 4 °C rotating wheel.
7. Centrifuge at 400 × *g* for 3 min at 4 °C. Discard supernatant.
8. Repeat three times **steps 6**–**7**.
9. Wash the beads two times with TE-0.1 Buffer (same procedure as above). Finally, remove as much wash buffer as possible from the beads.
10. Add 250 µL of Elution Buffer 1 to the beads. Incubate at 65 °C for 15 min under vigorous shaking (1300 rpm).
11. Spin at 400 × *g* for 3 min at RT. Transfer supernatant to a new tube referred to as IP-Elu.
12. Add 250 µL of Elution Buffer 2 to the beads. Incubate at 65 °C for 15 min under vigorous shaking (1300 rpm).
13. Centrifuge at 10,000 × *g* for 2 min at RT. Combine supernatant with IP-Elu to a final volume of 500 µL.
14. Add Proteinase K to a final concentration of 200 µg/mL to all samples including INPUT and S-IP. Incubate at 65 °C shaking at 300 rpm overnight to reverse the formaldehyde cross-link.

3.7 DNA Extraction and Precipitation

1. Cool down samples to room temperature.
2. Immediately prior to use, pellet Phase Lock Gel at 12,000–16,000 g for 30 s.
3. Add an equal volume of Phenol-Chloroform (PCl) to your sample. Vortex for 1 min. Transfer to pre-spun PLG tube. Spin 14,000 × *g* for 5 min to separate the phases (*see* **Note 13**).
4. Transfer the nucleic-acid-containing aqueous upper phase to a DNA low binding tube. Add 1/10 the volume of 3 M sodium acetate solution, 2 volumes of 100% EtOH and glycogen 40–50 µg/mL final (*see* **Note 14**). Incubate the mixture overnight at −20 °C.
5. Spin 1 h at 20,000 × *g* at 4 °C. Carefully remove supernatant.

6. Wash with 1 mL 70% EtOH (precooled at −20 °C), spin for 10 min 20,000 × *g* at 4 °C. Remove supernatant.
7. Air-dry the pellet.
8. Resuspend the pellet in a small volume, e.g., 10 µL dH_2O according to your input or expectations.
9. Determine DNA concentration using Qubit according to manufacturer's protocol.

3.8 Quality Control of ChIP and Deep-Sequencing Analysis

3.8.1 qPCR Analysis

The specificity and efficiency of the ChIP can be evaluated using real-time quantitative PCR. DNA samples from Input, Mock-IP, and IP-material will be assayed with positive and negative control primer sets. To compare the enrichment of a known target gene (e.g., engrailed (en), Antennapedia (antp), and Vestigial (vg) for H3K27m3) in the specific IP-fraction versus the Mock-IP-fraction will determine the background level. The enrichment of a known target gene in the specific IP-fraction versus the input will determine the percentage of input recovered. The comparison of the expected enriched region (known target gene) over a control gene locus (e.g., RpL32, PGRP-LE) will provide information about the fold change enrichment over a control (Fig. 4).

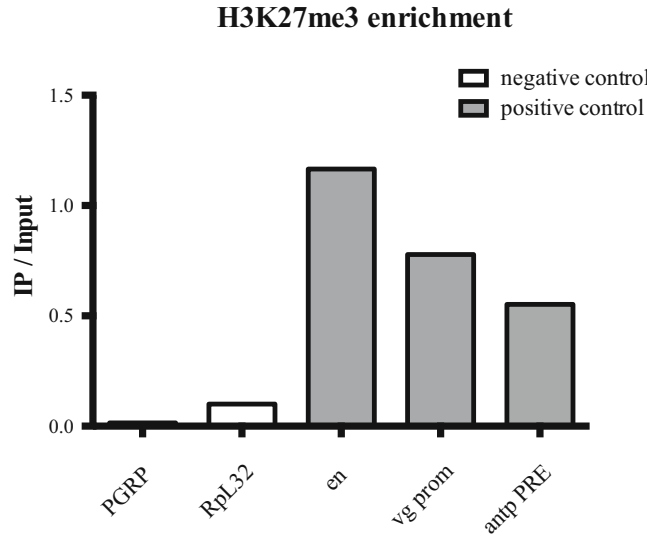


Fig. 4 Example of H3K27me3 enrichment on selected genes by ChIP followed by q-PCR. The specificity of the ChIP is controlled by analyzing the enrichment of H3K27me3 on two non-bound loci (negative control, PGRP and RpL32). We typically recover up to 1 % of input material on known H3K27me3 bound loci (positive control, en, vg prom and antp PRE). Results are presented as the percentage of input material recovered after immunoprecipitation. (PGRP: Peptidoglycan-recognition protein LE; RpL32: 60S ribosomal protein L32; en: engrailed; vg prom: vestigial promoter; antp PRE: antennapedia Polycomb Responsive Element)

3.8.2 Deep Sequencing Analysis

For ChIP followed by sequencing, input DNA will be the reference. The INPUT and the IP material will be required for library construction and DeepSeq. For this analysis the amount of DNA, the average length of the DNA fragments as well as the sequencing technology used will determine the depth and the coverage of your experiment and should be carefully defined prior sequencing.

4 Notes

1. Use molecular biology grade water (w/18 MΩ water; no detectable DNase, RNase, or protease activities) in all recipes and protocol steps.

   Low Retention Filter tips are recommended at all protocol steps.

2. Always use fresh formaldehyde. Ampule-sealed solutions of formaldehyde are highly recommended. Ampule packaging stabilizes the preparation by protecting the solution from both air oxidation and light, allowing access to "fresh" formaldehyde every time. Formaldehyde is highly toxic and has to be handled under a fume hood.

3. Set up cages using freshly hatched flies; ratio of male to female 1:3. Females are within their peak laying period 5 days after hatching and lay eggs quickly (~1 every 3 min). Flies tend to hold fertilized eggs for some time in their oviducts before laying them. Providing fresh food supports to release eggs held by the female overnight and stimulates egg-laying. For very early embryo collection set up several cages to collect enough embryos for ChIP.

4. Early development of *Drosophila* embryo has been widely described. At cycle 14 (stage 5), cellularization of the 4096 nuclei occurs. At 25 °C, cycle 14 occurs between 2h10 and 2h50 [5].

5. The *Drosophila* embryo is protected by a chorion (eggshell) and an impermeable vitelline membrane. For fixation the chorion needs to be removed. Treatment with sodium hypochlorite dissolves the chorion (dechorination), the heptane in the 2-phase fixing solution is essential to generate holes in the vitelline membrane that allow the formaldehyde to enter the embryo.

6. As the strength of bleach can vary, observation of dechorination is recommended. If dechorionation is complete embryos will look like smooth prolonged ovoids lacking the characteristic dorsal appendages (Fig. 2). The dechorionated embryos are very sensitive to dehydration; prolonged bleaching increases the frequency of artifacts.

7. The advantages of using formaldehyde as cross-linking reagent are the fast cross-linking kinetics, the stabilization of transient interactions and the short cross-linker length that allows only closely associated proteins to be cross-linked [6]. Since, Cross-linking is a time-dependent process it should not exceed 15 min. Extended cross-linking could result in reduced sonication efficiency, reduced antibody accessibility to antigen and epitope masking.
8. For the first time it is strongly recommended to establish the chromatin shearing step with your settings to ensure the optimal quality of your sheared chromatin, before continuing with the experiment. Later on, the quality check will be performed after IP using the S-IP sample (*see* Subheading 3.4).
9. Treatment of samples with RNase prior to agarose gel analysis significantly improves visual evaluation of shearing quality. For "Bioanalyzer" application it is also recommended. Mind that the dye used in the High Sensitivity DNA chips can also stain RNA at high concentrations.
10. Since beads tend to stick to the sides of the tip, try to minimize the movement in the pipette and use a tip cut 5 mm from the top.
11. The amount of antibody per IP and the incubation time depend on 2 major parameters, the abundance of the protein and the affinity properties of the antibody. For each new antibody a range of concentrations and different incubation settings should be tested to determine optimal conditions.
12. We assume that most DNA fragments remain in the supernatant after pre-clearing in the Mock-IP. Therefore, we use this supernatant as INPUT for further analysis as well as for library preparation.
13. Caution: phenol–chloroform is highly toxic, avoid inhalation and skin contact. Wear protective clothing and gloves, work under the fume hood.
14. Glycogene is useful as a nucleic acid co-precipitant; it increases the pellet's mass and therefore the quantitative recovery of low concentrations (ng/mL) of nucleic acids.

References

1. Gilmour DS, Lis JT (1985) In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. Mol Cell Biol 5(8):2009–2018
2. O'Neill LP, Turner BM (2003) Immunoprecipitation of native chromatin: NChIP. Methods 31(1):76–82
3. Li XY et al. (2014) Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. Elife 3
4. Lott SE et al (2011) Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed

through single-embryo RNA-seq. PLoS Biol 9(2), e1000590
5. Hartenstein V, Camposortega JA (1985) Fate-mapping in wild-type drosophila-melanogaster 1. The spatio-temporal pattern of embryonic-cell divisions. Wilhelm Rouxs Arch Dev Biol 194(4):181–195
6. Gavrilov A, Razin SV, Cavalli G (2015) In vivo formaldehyde cross-linking: it is time for black box analysis. Brief Funct Genomics 14(2):163–165

Chapter 4

ChIP-seq Data Processing for PcG Proteins and Associated Histone Modifications

Ozren Bogdanović and Simon J. van Heeringen*

Abstract

Chromatin Immunoprecipitation followed by massively parallel DNA sequencing (ChIP-sequencing) has emerged as an essential technique to study the genome-wide location of DNA- or chromatin-associated proteins, such as the Polycomb group (PcG) proteins. After being generated by the sequencer, raw ChIP-seq sequence reads need to be processed by a data analysis pipeline. Here we describe the computational steps required to process PcG ChIP-seq data, including alignment, peak calling, and downstream analysis.

Key words Polycomb, ChIP-seq, ChIP-sequencing, PRC1, PRC2, H3K27me3

1 Introduction

The Polycomb group complexes (PcG) have initially been described as major regulators of Hox gene silencing in Drosophila, able to exert their function in a spatially controlled manner [1]. Nowadays, we know that the PcG proteins display remarkable evolutionary conservation and that their regulatory reach extends well beyond the Hox genes. These proteins are known to form two major complexes, PRC1 [2] and PRC2 [3]. These complexes can cooperatively bind target genes in order to establish transcriptional repression; however, PRC1 also has independent roles [4–6]. In Drosophila, the canonical PRC1 complex consists of four core components: Polycomb (Pc), Posterior sex combs (Psc), Polyhomeotic (Ph), and Sex combs extra (Sce/RING) [2, 7, 8]. The Pc protein can bind histone H3 trimethylated at lysine 27 (H3K27me3), a repressive histone mark deposited by the PRC2 complex [9]. Additionally, the Sce subunit can ubiquitylate H2AK118 that participates in chromatin compaction and further promotes transcriptional silencing. In vertebrates, many homologues of the Drosophila PRC1 complex have been identified for Pc (CBX2, CBX4, CBX6, CBX7, and CBX8), Psc (BMI1, PCGF2), Ph (PHC1, PHC2, PHC3), and Sce (RING1, RNF2). The

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins*: *Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_4,

composition of PRC1 in vertebrates is highly variable, with distinct roles linked to different complex variants [10].

The PRC2 complex is composed of Enhancer of zeste (E(z)) that di- and trimethylates H3K27 [11, 12], Suppressor of zeste 12 (Su(z)12), and Extra sex combs (Esc) [10]. The canonical mammalian PRC2 complex contains the homologue proteins EZH1/2, SUZ12 and EED [13]. In addition, other proteins such as JARID2, AEBP2, and RBBP4/7 interact with PRC2 in a substoichiometric manner [14]. A number of recent studies have demonstrated that the repertoire of PcG complexes is far more elaborate than previously thought and that such complexity can have profound implications on embryonic development and disease [10].

In Drosophila the PRC1/PRC2 complexes are recruited to their target genes through sequences called Polycomb Response Elements (PREs) [15–17]. In vertebrates, the mechanism of recruitment is less clear. PcG proteins are recruited to unmethylated CpG dinucleotides within the context of regions with high GC content, such as CpG islands [18–23]. Other potential mechanisms include recruitment by transcription factors [24, 25] or RNA [26–30].

The studies aimed at unraveling the mechanisms of recruitment of PcG to chromatin were greatly facilitated by the development of immunoprecipitation approaches combined with next-generation sequencing, known as ChIP-seq (Chromatin Immunoprecipitation Sequencing) [31]. ChIP-seq studies have allowed researchers to generate profiles of genomic PcG complex binding sites as well as H3K27me3 and H2AK118/119 histone mark enrichments. Such genome-wide studies, carried out throughout the last decade, have greatly transformed our understanding of PcG regulation. Myriad studies conducted in various model systems have demonstrated the importance of PcG patterning for embryonic development [22, 32, 33] and disease [34]. The role of PcG complexes in transcription regulation is much more diverse than previously thought [6, 35, 36]. Furthermore, the PcG proteins were found to be a key determinant of the 3D genome architecture [37] whereas the chromatin mark H3K27me3 deposited by the PRC2 complex could be a hallmark of inactive and poised enhancers [38, 39]. An overview of some key genome-wide PcG studies and associated publicly available data is provided in Table 1.

The first step in ChIP-seq library preparation consists of cross-linking of protein-protein and protein-DNA interactions by formaldehyde fixation. Afterwards, the cells are lysed and the fixed chromatin is fragmented by sonication. Specific antibodies against a target protein or a histone modification are then used to immunoprecipitate the bound DNA, resulting in a sample enriched in fragments bound by the protein or a histone modification of interest. The immunoprecipitated DNA fragments are then sequenced by massively parallel sequencing technologies and the genome-wide binding profiles are obtained through bioinformatics

Table 1
A selection of PcG ChIP-seq studies

Reference	Species	Experiments	GEO accession
Bowman [67]	*Drosophila*	Ph, Pc, Suz12, H3K27me3	GSE55257
Herz [36]	*Drosophila*	Jarid2, Suz12, H3K27me3	GSE33546
Orsi [68]	*Drosophila*	Adf1, GAF, Pc, Pho, H3K27me3	GSE47829
Schuettengruber [17]	*Drosophila*	DSP1, Ph, Pc, H3K27me3	GSE60428
Cao [69]	Human	BMI1, EED, RING1A, RING1B, H2AK119ub, H3K27me3	GSE42566
ENCODE [59]	Human	CBX2, CBX8, EZH2, H3K27me3	ENCODE[a]
Gao [4]	Human	BMI1, CBX2, RING1B, RYBP, H2AK119Ub1	GSE34774
Pemberton [70]	Human	CBX6, CBX7, CBX8, RING1A, RING2, H3K27me3	GSE40740
Roadmap Epigenomics [71]	Human	H3K27me3	Roadmap[b]
Ku [18]	Human, mouse	EZH2, RING1B, SUZ12, H3K27me3	GSE13084
Blackledge [48]	Mouse	EZH2, KDM2B, RING1B, SUZ12	GSE55698
ENCODE [59]	Mouse	EZH2, H3K27me3	ENCODE[a]
Farcas [72]	Mouse	KDM2B, RING1B	GSE41267
Frangini [73]	Mouse	CBX7, EZH2, RING1B, H2AK119Ub1,H3K27me3	GSE42706
Gao [6]	Mouse	BMI1, RING1B, H2AK119Ub1, H3K27me3	GSE60409
Pasini [74]	Mouse	JARID2, SUZ12	GSE19365
Peng [75]	Mouse	EZH2, JARID1A, JARID2, SUZ12	GSE18776
Tavares [5]	Mouse	RING1B, H3K27me3	GSE23716
Wu [37]	Mouse	KDM2B, RING1B	GSE37930
Irimia [33] De la Calle Mustienes [61]	*D. rerio* *D. rerio*	H3K27me3 H3K27me3	GSE35050 GSE70847
van Heeringen [22]	*X. tropicalis*	Jarid2, Ezh2, H3K27me3	GSE41161

[a]http://www.ncbi.nlm.nih.gov/geo/info/ENCODE.html
[b]http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/

approaches. In this chapter we will outline the general guidelines for ChIP-seq data analysis of PcG proteins and associated histone modifications.

2 Materials

2.1 Experimental Design

While this protocol is mainly focused on the processing and computational analysis of ChIP-seq data, we still would like to briefly touch upon experimental design. The ENCODE ChIP-seq guidelines are a good start for a successful experiment [31]. The major considerations are:

1. Use a characterized antibody or make sure to characterize your antibody to detect poor reactivity and/or cross-reactivity with other proteins. The first aspect will result in an unsuccessful ChIP-seq experiment. The second, however, is potentially much more influential as the ChIP-seq profile will not represent the correct protein. In the absence of a specific antibody, proteins with an epitope tag such as V5 or HA are a viable alternative.
2. Obtain at least two biological replicates for each experiment.
3. Sequence a control sample. This can be either DNA that was subjected to the same protocol (cross-linking and fragmentation) or a "mock" IP with a control antibody, "IgG."
4. The read length for ChIP-seq experiments is not of significant importance; 50 cycles is enough to provide meaningful data in most cases.
5. The minimum amount of reads necessary for a ChIP-seq experiment depends on the enrichment profile of the factor or modification and genome size. The H3K27me3 modification and most of the PcG proteins are enriched in broad domains. For these types of profiles, ENCODE recommends 20 million mapped reads in mammalian sized genomes, while the equivalent modENCODE guideline for broad domains in Drosophila is 5 million reads.

2.2 Required Software

1. Install the following programs:
2. SRA Toolkit
   http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software
3. FastQC
   http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
4. Trimmomatic [40]
   http://www.usadellab.org/cms/?page=trimmomatic

5. bwa [41]
   http://bio-bwa.sourceforge.net/
6. SAMtools [42]
   http://samtools.sourceforge.net/
7. Picard Tools
   http://broadinstitute.github.io/picard/
8. bedtools [43]
   http://bedtools.readthedocs.org/en/latest/
9. phantompeakqualtools [44, 45]
   http://code.google.com/p/phantompeakqualtools/
10. MACS2 [46]
    https://github.com/taoliu/MACS/
11. wigToBigWig, bedGraphToBigWig, and fetchChromSizes
    http://hgdownload.cse.ucsc.edu/admin/exe
12. IGV [47]
    http://www.broadinstitute.org/igv/
13. fluff
    http://github.com/simonvh/fluff

2.3 Required Hardware

The minimum system requirements concern mainly the memory, which should be at least 8GB RAM. In addition, having a multi-core CPU will increase the speed of, for instance, read alignment. We recommend 8 cores, but this is not absolutely required.

3 Methods

3.1 Introduction and Sample Data

The data used as an example are from "Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation" [48], NCBI GEO accession GSE55697. This dataset contains single-end 51 bp ChIP-seq data for RNF2 (RING1B), SUZ12, EZH2, and H3K27me3. The first step of the protocol illustrates how to obtain data from the Sequence Read Archive (SRA) at NCBI (http://www.ncbi.nlm.nih.gov/sra) and is only applicable for published, publicly available experiments. It will be assumed that the protocol is run from the Bash shell on a Linux machine with 8 CPUs and 16GB of internal memory.

3.2 Data Download from SRA via NCBI GEO

1. Download the SRA file (*see* **Note 1**). Go to the GEO page for accession GSE55697 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55697) and select a sample. Here we will select GSM1341951. At the bottom of the page there is a link called "(ftp)" to an SRA Experiment. Follow this link and click on the directory that starts with SRR, SRR1185473 in this case. Download the .sra file SRR1185473.sra.

2. Convert the SRA file to a gzipped FASTQ using fastq-dump (*see* **Note 2**).

```
$ fastq-dump --gzip SRR1185473.sra
```

3. Remove the SRA file and rename the FASTQ file to reflect the GSM ID.

```
$ rm SRR1185473.sra && mv SRR1185473.fastq.gz\
GSM1341951.fastq.gz
```

3.3 Quality Control

1. Check data quality using FastQC (*see* **Note 3**). Create a quality report of all FASTQ files in a folder with the following command.

```
$ mkdir fastqc.out
$ fastqc *.fastq.gz -o fastqc.out
```

3.4 Adapter Clipping

1. Remove the adapters with Trimmomatic [40] (*see* **Notes 4** and **5**). Set TRIM_PATH to the location where Trimmomatic is installed on your computer.

```
$ export TRIM_PATH=/opt/trimmomatic/
$ java -jar $TRIM_PATH/trimmomatic-0.32.jar SE \
GSM1341951.fastq.gz GSM1341951.trimmed.fastq.
gz \
ILLUMINACLIP:$TRIM_PATH/adapters/TruSeq3-SE.
fa:2:30:10
```

2. Repeat step 1 for GSM1341937, GSM1341938, and GSM1341952.

3.5 Read Mapping

1. Obtain the reference genome, which will be used for mapping. Download the mouse genome (Dec. 2011 assembly, GRCm38/mm10) (*see* **Notes 6** and **7**).

```
$ mkdir mm10
$ cd mm10
$ wget http://hgdownload.cse.ucsc.edu/golden-
Path/mm10/\
bigZips/chromFa.tar.gz
$ tar cvzf chromFa.tar.gz
$ cat *.fa mm10.fa
```

2. Optionally delete separate chromosome FASTA files.

```
$ rm chr*.fa
```

3. Prepare the genome using bwa index.

```
$ bwa index mm10.fa
```

4. Map reads to the reference genome using bwa aln (*see* **Notes 8** and **9**).

```
$ bwa aln -t 12 mm10/mm10.fa \\
bwa samse mm10/mm10.fa - GSM1341951.trimmed.
fastq.gz | \
samtools view -Sb - > GSM1341951.mm10.bam
```

5. Sort the bam file by coordinate (*see* **Note 10**).

```
$ samtools sort -m 12G GSM1341951.mm10.bam \
GSM1341951.mm10.sorted
```

6. Mark duplicate reads using MarkDuplicates from Picard Tools.

```
$ java -jar MarkDuplicates.jar ASSUME_
SORTED=true \
VALIDATION_STRINGENCY=LENIENT \
INPUT=GSM1341951.mm10.sorted.bam \
OUTPUT=GSM1341951.mm10.sorted.markdup.bam \
METRICS_FILE=GSM1341951.mm10.sorted.markdup.txt
```

7. Index the BAM file

```
$ samtools index GSM1341951.mm10.sorted.mark-
dup.bam
```

8. Get an estimation of library complexity (*see* **Note 11**).

```
$ samtools flagstat GSM1341951.mm10.sorted.
markdup.bam
```

9. Repeat **steps 3–6** for GSM1341937, GSM1341938, and GSM1341952.

3.6 Visualization of ChIP-seq Data

The procedure described in this section will generate a bigWig file that can be used for visualization, from a sorted, duplicate marked BAM file (*see* **Note 12**).

1. Create a genome file with chromosome sizes (*see* **Note 13**).

```
$ fetchChromSizes mm10 > mm10.genome
```

2. Filter duplicates and non-uniquely mapped reads with samtools [42] (*see* **Note 14**).

```
$ samtools view -b -F 1024 -q 30 \ GSM1341951.
mm10.sorted.markdup.bam > \
GSM1341951.mm10.filtered.bam
```

3. Run phantompeakqualtools to estimate fragment length (*see* **Note 15**).

```
$ Rscript /opt/phantompeakqualtools/run_spp.R \
-c=GSM1341951.mm10.bam -savp -out=GSM1341951.
ppqt.txt
$ FSIZE=`cat GSM1341951.ppqt.txt | cut -f3 |
cut -f1 -d,`
```

4. Determine scaling factor.

```
$ SCALE=`samtools flagstat GSM1341951.mm10.fil-
tered.bam \
|head -n 1 | awk '{print 1e6 / $1}'`
```

5. Generate a genome coverage bedGraph file using bedtools [43]. The reads will be extended to the estimated fragment size, and the bedGraph will be normalized by million mapped reads.

```
$ bedtools bamtobed -i GSM1341951.mm10.fil-
tered.bam \
```

```
bedtools flank -l 0 -r 1 -s -g mm10.genome \
bedtools slop -l 0 -r $(($FSIZE - 1)) -g mm10.
genome \
-s |bedtools genomecov -i - -g mm10.genome -bg \
> GSM1341951.mm10.bg
```

6. Convert bedGraph to bigWig (*see* **Note 16**).

```
$ bedGraphToBigWig GSM1341951.mm10.bg mm10.genome \
GSM1341951.mm10.bw
```

7. Upload the bigWig file to a web accessible location (http or ftp).
8. Repeat **steps 2–7** for GSM1341937, GSM1341938, and GSM1341952.
9. Create a track definition file. Open a text editor and add one or more track line(s) as follows (*see* **Note 17**):

```
track type=bigWig name=H3K27me3_repl1 \
bigDataUrl=http://url/to/files/GSM1341951.mm10.
bw \
color=228,26,28 autoScale=on visibility=full \
maxHeightPixels=128:32:11
```

10. You can add one or more lines, representing tracks that will be visualized. Note that the above example should all be written on one line, without the \ characters at the end of these lines.
11. Upload track and view in the UCSC genome browser [49] (*see* **Note 18**). Go to http://genome.ucsc.edu/cgi-bin/hgGateway and select the mouse mm10 assembly. Click on "add custom tracks," select "choose" to upload the file you have created, and click on "Submit."
12. Visually inspect the ChIP-seq track to assess the quality (*see* **Note 19**).

3.7 Peak Calling

1. Call peaks using MACS2 [46] (*see* **Notes 20–22**).

```
$ macs2 callpeak -t GSM1341951.mm10.filtered.
bam \
-c GSM1341937.mm10.filtered. bam \
-f BAM -g mm -n H3K27me3_untreated_rep1 --ext-
size $FSIZE \
--broad -q 0.01
```

2. Create a BED file of the peaks with the -log10(q-value) in the fourth column.

```
$ cut -f1-3,9 H3K27me3_untreated_rep1.broad-
Peak \ > H3K27me3_untreated_rep1.broad.bed
```

3. Add the track line to the top of the BED file.

```
$ sed -i
\'1s/^/track name=H3K27me3_untreated_rep1_
peaks\n/' \
H3K27me3_untreated_rep1_peaks.broad.bed
```

4. Upload the BED file to the UCSC Genome Browser with "add custom tracks" (*see* Subheading 3.6, **step 10**).
5. Visually compare the peaks to the ChIP-seq profile (*see* Subheading 3.6) and evaluate the peak calls.
6. Repeat step 1 and increase or decrease the q-value cutoff to change the selection criteria for the peaks (*see* **Note 23**).
7. Repeat **steps 1–6** for replicate 2 (GSM1341952 and GSM1341938) (*see* **Note 24**).
8. Create a set of reproducible peaks.

```
$ bedtools intersect \
-a H3K27me3_untreated_rep1.broad.bed \
-b H3K27me3_untreated_rep2.broad.bed > \
H3K27me3_untreated.peaks.bed
```

3.8 Follow-Up Analysis

The steps covered in the protocol until this point represent the most common steps in ChIP-seq data processing. Many of the follow-up experiments and analysis depend on the specific type of question. Here, we will detail a few possible options that cover functional annotation, creating heatmaps and clustering.

3.8.1 Conversion Between Genome Builds

1. Convert peaks to mm9 for tools that don't support mm10 using UCSC liftOver. Go to http://genome-euro.ucsc.edu/cgi-bin/hgLiftOver. Select mm10 as "Original Assembly" and mm9 as "New Assembly." Go to the "upload data from a file" section, select the peak BED file, and click "Submit File" (*see* **Note 25**). Click on "View Conversions" and save the file as "H3K27me3_untreated.peaks.mm9.bed".

3.8.2 Functional Annotation of Peaks

We define functional annotation here as any analysis that aims to derive biological meaning from the potential target genes obtained from a ChIP-seq experiment. This includes various methods such as Gene Ontology analysis and pathways or disease enrichment.

1. Analyze peaks with GREAT [50]. Go to http://great.stanford.edu/ and select "Mouse: NCBI build 37 (UCSC mm9, Jul/2007)." Select the "Test regions BED file" and upload the peaks. Click on "Submit."
2. Create a file transcription start sites of mm10 genes. Go to http://genome.ucsc.edu/cgi-bin/hgTables and select the mouse mm10 assembly. Select group "Genes and Gene Predictions" and track "RefSeq Genes." On the region line select "genome" and on the "output format" line select "BED." Click on "get output." Create one BED record per "Upstream by 1 bases" and click on get BED. Save the file as `RefSeq.tss.mm10.bed`.

3. Overlap the transcription start sites with peaks.

```
$ bedtools intersect -a RefSeq.tss.mm10.bed \
-b H3K27me3_untreated.peaks.bed | cut -f4 | \
cut -f1-2 -d\_ | sort -u > RefSeq.tss.mm10.
overlap.txt
```

4. Analyze overlapping genes with DAVID [51] (*see* **Note 26**). Go to http://david.abcc.ncifcrf.gov/. Select "Functional Annotation" and click on "Upload." Select the file "`RefSeq.tss.mm10.overlap.txt`" in Step 1, "B:Choose From a File," and in Step 2 select "REFSEQ_MRNA." Select "Gene List" and click on "Submit List." Not all RefSeq IDs might be recognized, select Option 1 to "Continue to Submit the IDs That DAVID Could Map" (*see* **Note 27**). Click on "Functional Annotation Clustering" link, and once again on the "Functional Annotation Clustering" button (*see* **Note 28**). Note the Enrichment Score and Benjamini (the p-value corrected for multiple testing) columns to get an idea of the significance of the results.

3.8.3 Creating Heatmaps and Clustering

Create a heatmap using fluff (http://github.com/simonvh/) (*see* **Note 29**).

1. Follow **steps 1**–7, Subheading 3.5 for the following accessions: GSM1341937, GSM1341939, GSM1341943, GSM1341947, GSM1341951.
2. Create symlinks for easier naming (*see* **Note 30**).

```
$ ln -s GSM1341937.mm10.sorted.markdup.bam
Input.1.bam
$ ln -s GSM1341937.mm10.sorted.markdup.bam.bai
Input.1.bam.bai
$ ln -s GSM1341939.mm10.sorted.markdup.bam
RING1B.1.bam
$ ln -s GSM1341939.mm10.sorted.markdup.bam.bai
RING1B.1.bam.bai
$ ln -s GSM1341943.mm10.sorted.markdup.bam
SUZ12.1.bam
$ ln -s GSM1341943.mm10.sorted.markdup.bam.bai
SUZ12.1.bam.bai
$ ln -s GSM1341947.mm10.sorted.markdup.bam
EZH2.1.bam
$ ln -s GSM1341947.mm10.sorted.markdup.bam.bai
EZH2.1.bam.bai
$ ln -s GSM1341951.mm10.sorted.markdup.bam
H3K27me3.1.bam
$ ln -s GSM1341951.mm10.sorted.markdup.bam.bai
H3K27me3.1.bam.bai
```

åå 3. Create the heatmap using k-means clustering with 6 clusters (*see* **Note 31**).

```
$ fluff_heatmap.py -f RefSeq.tss.mm10.bed -d \
```

```
Input.1.bam,RING1B.1.bam,SUZ12.1.bam, \
EZH2.1.bam,H3K27me3.1.bam \
-c grey,blue,green,purple,red -C kmeans -k 6
-s 95% \
-o heatmap.png
```

4 Notes

1. Newer versions of fastq-dump allow to download FASTQ files directly without going through an intermediate SRA file. However, the Toolkit should then be configured correctly, see http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc&f=std.
2. This protocol is for single-end ChIP-seq data. In the case of paired-end data add the --split-files option to fastq-dump.
3. The data from the sequencer comes in a format called FASTQ that contains both sequence and quality information [52]. Before mapping, the quality of the FASTQ files can be checked with FastQC, which produces several graphs and tables that give an impression of the sequence quality. FastQC, like most programs that read FASTQ files, can directly read gzip compressed FASTQ files, which saves disk space. The most informative measures from FastQC are the "Per base sequence quality" and the "Sequence duplication levels." Other interesting features are the "Overrepresented sequences" and "Adapter Content," which can help identify adapters, and the "Per sequence GC content," which would allow to spot putative contaminants such as bacterial sequences.
4. The Illumina TruSeq adapters used in this step are included with Trimmomatic. For other adapters, replace the file $TRIM_PATH/adapters/TruSeq3-SE.fa with a FASTA file that contains the adapter sequences.
5. If there are low-quality bases at the start or end of the sequence, Trimmomatic can also be used to remove these.
6. In this case we obtain the reference genome from the UCSC http://hgdownload.cse.ucsc.edu/downloads.html. Alternatives are Ensembl http://www.ensembl.org/info/data/ftp or iGenomes http://support.illumina.com/sequencing/sequencing_software/igenome.html.
7. The steps preparing the reference genome need to be performed only once for a specific reference.
8. There are many different aligners that can be used for ChIP-seq data. Two of the most popular alignment programs are bwa [41] and bowtie [53, 54]. For this protocol we will use bwa, but bowtie is a good alternative.

9. There is a new alignment algorithm available for bwa, bwa mem [55], which is recommended for 70 base pair and longer reads. In this protocol we use bwa aln as the reads are 51 bp.
10. Optionally the samtools sort command can be combined with the alignment command. In some versions of samtools this will generate a warning, which can be ignored. The full command would then look like this:
```
$ bwa aln -t 6 mm10/mm10.fa GSM1341951.fastq.
gz | \
bwa samse mm10/mm10.fa - in.fq | \
samtools view -Sb - | \
samtools sort -m 12G - GSM1341951.mm10.sorted
```
Make sure to set the sort buffer size to a reasonable value, given that the genome index used by bwa aln will also consume quite some memory.
11. The proportion of duplicate reads relative to the total sequence reads can be used as a measure of ChIP library quality. The ENCODE consortium has defined the fraction of nonredundant mapped reads as the Non-Redundant Fraction (NRF), and suggests an NRF of at least 0.8 for 10 million mapped reads as the minimum library quality [31]. Alternative to samtools flagstat, the duplicate statistics can be retrieved from the Picard MarkDuplicates metrics output file. Another method, Preseq, can be used to estimate the molecular complexity of a sequencing library and predict the added benefit of additional sequencing [56].
12. Other tools, such as MACS [46], enable creation of Wiggle visualization track. However, this procedure allows for normalization of read depth in the bedtools genomecov step.
13. This command only needs to be executed once for a specific reference genome.
14. The -q parameter selects a minimum mapping quality. As nonuniquely mapping reads get a quality score of 0, they will be filtered out. The -F 1024 parameter selects all reads without the flag 1024, which marks reads as being duplicates. See https://broadinstitute.github.io/picard/explain-flags.html for combinations of different flags.
15. PhantomPeakQualTools computes the most likely fragment length based on the strand cross-correlation profile of read start density on the + and – strand. The strongest peak, next to a potentially strong peak at the read length, should be the predominant fragment length. Note that this is not guaranteed to return the correct fragment length in all cases; it is recommended to examine the cross-correlation plots [45].
16. Alternatively, wigToBigWig can be used in a pipe to directly create a bigWig file without storing the intermediate bedGraph file.

However, this will consume more memory. The bedGraphToBigWig command can not be used in a pipe as it has to read the input file in multiple passes to limit memory consumption.

17. The ColorBrewer website http://colorbrewer2.org/ can be used to pick color schemes for visualization.
18. Alternatively, the bigWig file can be loaded by IGV (the Integrative Genomics Viewer) [47, 57].
19. Check for occurrence of peaks, or broadly enriched domains, at known targets. While more formal methods can be used to define the quality of a ChIP-seq track [31], visual inspection is a very informative step.
20. Reliable and consistent peak calling remains somewhat challenging. A variety of peak callers have been developed, using different methods; however, it remains unclear which method performs best across a wide range of experimental conditions. MACS2 [46] and PeakRanger [58] which have been used by the ENCODE [59] and modENCODE [60] consortia, respectively, generally yield reliable results. Peaks for the PcG proteins and associated histone modifications are generally broad peaks, and not point-source, but can still mostly be identified by programs such as MACS2 with the --broad option. We have recently used this algorithm to identify H3K27me3 peaks in zebrafish embryos [61]. However, some cell types, such as a significant proportion of human (cancer) cell lines, show extremely broad regions of very low enrichment. In these cases alternative methods might be more suitable. For instance, RSEG [62] which has yielded good results in our hands [63] and EDD [64] could be used.
21. This step assumes that the input track GSM1341937 has been mapped to the mm10 genome and that the corresponding BAM file GSM1341937.mm10.bam exists.
22. The $FSIZE variable represents the fragment size obtained from phantompeakqualtools in Subheading 3.6, **step 3**.
23. The default settings are not always suitable. If desired, peaks for a specific combination of parameters can be evaluated experimentally using ChIP-qPCR validation. Select at least 20 random peaks for qPCR validation to obtain an approximate experimental False Discovery Rate (FDR).
24. Use the same q-value cutoff for replicate experiments.
25. UCSC liftOver supports very specific formats. The BED file needs to be between 4 and 6 columns and cannot contain track or header lines.
26. In this protocol we have used the RefSeq symbols. However, other gene symbols, such as official gene names are possible and can be selected in DAVID.

27. If gene names are used, the species will need to be selected in the list of species located at the left of the page.
28. It is possible to use separate annotations such as Gene Ontology or Pathways; however the clustering tool gives a nice comprehensive overview.
29. Heatmaps can be created using a variety of methods. Other good alternatives are ngs.plot [65], the gplots package in R (http://cran.r-project.org/package=gplots), or the metaseq package for Python (https://pythonhosted.org/metaseq/).
30. The file naming is mainly a cosmetic change, as fluff uses the file names to name the different tracks in the visualization.
31. For an overview of clustering methods, see D'haeseleer [66]. Most of the observations and guidelines are also applicable to clustering of ChIP-seq profiles. Experiment with a number of different clusterings and cluster numbers in the case of k-means. While there are ways to computationally define the optimal number of clusters, this is not straightforward and does not work well in all cases.

Acknowledgments

O.B. is supported by an Australian Research Council Discovery Early Career Researcher Award—DECRA (DE140101962); S.J.v.H. is supported by the Netherlands Organization for Scientific Research (NWO-ALW grant 863.12.002).

References

1. Lewis EB (1978) A gene complex controlling segmentation in Drosophila. Nature 276:565–570
2. Shao Z, Raible F, Mollaaghababa R et al (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell 98:37–46
3. Cao R, Wang L, Wang H et al (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (New York, NY) 298:1039–1043
4. Gao Z, Zhang J, Bonasio R et al (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell 45:344–356
5. Tavares L, Dimitrova E, Oxley D et al (2012) RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell 148:664–678
6. Gao Z, Lee P, Stafford JM et al (2014) An AUTS2-Polycomb complex activates gene expression in the CNS. Nature 516:349–354
7. Francis NJ, Saurin AJ, Shao Z et al (2001) Reconstitution of a functional core polycomb repressive complex. Mol Cell 8:545–556
8. Saurin AJ, Shao Z, Erdjument-Bromage H et al (2001) A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. Nature 412:655–660
9. Steffen PA, Ringrose L (2014) What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat Rev 15:340–356
10. Schwartz YB, Pirrotta V (2013) A new world of Polycombs: unexpected partnerships and emerging functions., Nature reviews. Genetics 14:853–864
11. Czermin B, Melfi R, McCabe D et al (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111:185–196

12. Müller J, Hart CM, Francis NJ et al (2002) Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 111:197–208
13. Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469:343–349
14. Smits AH, Jansen PWTC, Poser I et al (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res 41:e28
15. Simon J, Chiang A, Bender W et al (1993) Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev Biol 158:131–144
16. Chan CS, Rastelli L, Pirrotta V (1994) A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. EMBO J 13:2553–2564
17. Schuettengruber B, Oded Elkayam N, Sexton T et al (2014) Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila. Cell Rep 9:219–233
18. Ku M, Koche RP, Rheinbay E et al (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet 4:e1000242
19. Mendenhall EM, Koche RP, Truong T et al (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet 6:e1001244
20. Lynch MD, Smith AJH, De Gobbi M et al (2012) An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J 31:317–329
21. Long HK, Sims D, Heger A et al (2013) Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. Elife 2, e00348
22. van Heeringen SJ, Akkers RC, van Kruijsbergen I et al (2014) Principles of nucleation of H3K27 methylation during embryonic development. Genome Res 24:401–410
23. Wachter E, Quante T, Merusi C et al (2014) Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. Elife 3, e03397
24. Dietrich N, Lerdrup M, Landt E et al (2012) REST-mediated recruitment of polycomb repressor complexes in mammalian cells. PLoS Genet 8:e1002494
25. Arnold P, Schöler A, Pachkov M et al (2013) Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. Genome Res 23:60–73
26. Rinn JL, Kertesz M, Wang JK et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323
27. Zhao J, Sun B, Erwin J et al (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 215
28. Tsai M-C, Manor O, Wan Y et al (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science (New York, NY) 329:689–693
29. da Rocha ST, Boeva V, Escamilla-Del-Arenal M et al (2014) Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Mol Cell 53: 301–316
30. Kaneko S, Bonasio R, Saldaña-Meyer R et al (2014) Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol Cell 53:290–300
31. Landt SG, Marinov GK, Kundaje A et al (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22:1813–1831
32. Akkers RC, van Heeringen SJ, Jacobi UG et al (2009) A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev Cell 17:425–434
33. Irimia M, Tena JJ, Alexis MS et al (2012) Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res 22:2356–2367
34. Richly H, Aloia L, Di Croce L (2011) Roles of the Polycomb group proteins in stem cells and cancer. Cell Death Dis 2, e204
35. Enderle D, Beisel C, Stadler M (2011) Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res 21(2):216–226
36. Herz H-M, Mohan M, Garrett AS et al (2012) Polycomb repressive complex 2-dependent and -independent functions of Jarid2 in transcriptional regulation in Drosophila. Mol Cell Biol 32:1683–1693
37. Wu X, Johansen JV, Helin K (2013) Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Mol Cell 49:1134–1146
38. Bonn S, Zinzen RP, Girardot C et al (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44:148–156
39. Rada-Iglesias A, Bajpai R, Swigut T et al (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470:279–283

40. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30:2114–2120
41. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 25:1754–1760
42. Li H, Handsaker B, Wysoker A et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 25:2078–2079
43. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) 26:841–842
44. Kharchenko PV, Tolstorukov MY, Park PJ (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol 26:1351–1359
45. Marinov GK, Kundaje A, Park PJ et al (2014) Large-scale quality analysis of published ChIP-seq data. G3 (Bethesda, MD) 4:209–223
46. Zhang Y, Liu T, Meyer CA et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137
47. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192
48. Blackledge NP, Farcas AM, Kondo T et al (2014) Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell 157:1445–1459
49. Kent W, Sugnet C, Furey T (2002) The human genome browser at UCSC. Genome Res 12(6):996–1006
50. McLean CY, Bristor D, Hiller M et al (2010) GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28:495–501
51. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57
52. Cock PJA, Fields CJ, Goto N et al (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 38:1767–1771
53. Langmead B, Trapnell C, Pop M et al (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
54. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
55. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, 00, 3
56. Daley T, Smith A (2013) Predicting the molecular complexity of sequencing libraries. Nat Methods 10:325–327
57. Robinson J, Thorvaldsdóttir H (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26
58. Feng X, Grossman R, Stein L (2011) PeakRanger: A cloud-enabled peak caller for ChIP-seq data. BMC Bioinformatics 12:139
59. ENCODE Project Consortium, Bernstein BE, Birney E et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
60. modENCODE Consortium, Celniker SE, Dillon LA et al (2009) Unlocking the secrets of the genome. Nature 459:927–930
61. de la Calle Mustienes E, Gómez-Skarmeta JL, Bogdanović O (2015) Genome-wide epigenetic cross-talk between DNA methylation and H3K27me3 in zebrafish embryos. Genomics Data 6:79
62. Song Q, Smith AD (2011) Identifying dispersed epigenomic domains from ChIP-Seq data. Bioinformatics (Oxford, England) 27:870–871
63. Brinkman AB, Gu H, Bartels SJJ et al (2012) Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res 22:1128–1138
64. Lund E, Oldenburg AR, Collas P (2014) Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. Nucleic Acids Res 42:92
65. Shen L, Shao N, Liu X et al (2014) ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 15:284
66. D'haeseleer P (2005) How does gene expression clustering work? Nat Biotechnol 23: 1499–1501
67. Bowman SK, Deaton AM, Domingues H et al (2014) H3K27 modifications define segmental regulatory domains in the Drosophila bithorax complex. Elife 3, e02833
68. Orsi G, Kasinathan S (2014) High-resolution mapping defines the cooperative architecture of Polycomb response elements. Genome Res 24(5):809–820
69. Cao Q, Wang X, Zhao M et al (2014) The central role of EED in the orchestration of polycomb group complexes. Nat Commun 5:3127
70. Pemberton H, Anderton E, Patel H et al (2014) Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. Genome Biol 15:R23
71. Bernstein BE, Stamatoyannopoulos JA, Costello JF et al (2010) The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 28:1045–1048

72. Farcas AM, Blackledge NP, Sudbery I et al (2012) KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife 1, e00205
73. Frangini A, Sjöberg M, Roman-Trufero M et al (2013) The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. Mol Cell 51:647–661
74. Pasini D, Cloos PAC, Walfridsson J et al (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature 464:306–310
75. Peng JC, Valouev A, Swigut T et al (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 139: 1290–1302

Chapter 5

Analysis of Single-Locus Replication Timing in Asynchronous Cycling Cells

Lo Sardo Federica*

Abstract

In higher eucaryotes, not all the genome is replicated simultaneously: there are parts of the genome that replicate at the beginning of S-phase (early S-phase), others that are replicated later. In each cell, early replicating genomic regions are alternated to late-replicating regions. In general, eucaryotic genomes are organized into structural domains where genes showing the same epigenetic state replicate at the same time.

Here, we will describe the protocol that we routinely used for the analysis of replication timing of specific loci in Drosophila embryonic cell lines (S2 and S3) based on BrdU labeling and FACS sorting of different S-phase fractions (early, mid, late) of asynchronous cycling cells.

Key words Replication timing, Early, Mid, Late, S-phase, BrdU, FACS sorting

1 Introduction

The notion that the eucaryotic genome is replicated with a specific temporal program during S-phase is well established [1, 2]. The choice of replication origin and the time of firing are plastic and cell-type specific [3–5] and can change dynamically during differentiation and development [6–9]. It is now well established that replication timing is highly correlated to the chromatin state of specific genomic segments and to the three-dimensional folding of the genome in the nucleus (reviewed in [10]). In our lab, we have recently shown in Drosophila that Polycomb Group proteins (PcG) are involved in the epigenetic maintenance of repressed transcriptional state of targeted genomic loci through replication and contribute to the definition and maintenance of the three-dimensional structure and the replication program of targeted loci. By focusing our attention on the PcG repressed homeotic gene cluster Bithorax complex (BX-C), we showed that downregulation of multiple PcG complex subunits disrupts not only PcG-mediated transcriptional repression but also three-dimensional chromatin

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_5,

folding and replication timing program at the BX-C. In agreement with this result, two Drosophila embryonic cell lines differing for BX-C gene expression states, PcG distribution, and chromatin domain conformation revealed a cell-type-specific replication program that mirrors lineage-specific higher-order structures [11–14].

Methods aimed to measure the timing of replication of specific chromosomes, chromosomal segments, loci, or alleles have been developed since 1960s [1, 2, 15–19]. In general, growing cells are labeled during S-phase by using dNTP analogues that are incorporated into newly replicating DNA. Then, cells from different stages of S-phase can be enriched by synchronization or by sorting with flow cytometry (FACS = Fluorescence-Activated Cell Sorting) based on total DNA content. Finally, the replication timing of specific loci is determined by quantifying the relative abundance of their sequences labeled with dNTP analogues in the different stages of S-phase.

In the described protocol, the relative abundance of nascent DNA synthesized during different time points of the S-phase is determined by bromodeoxyuridine triphosphate (BrdU) labeling of asynchronous growing cells and FACS sorting. This method has been adapted with minor modification to that described firstly by Schubeler et al. in 2002 [20] (see also [16]). In this protocol, BrdU is incorporated in cycling cells into newly replicating DNA. Then, asynchronous cells are fixed with ethanol and stained with Propidium Iodide (PI) that is intercalated into cellular DNA. Cells from different stages of the S-phase, (early, mid, late) are characterized by different amount of total DNA content that is progressively doubled from G1 to G2 phase. The more the total DNA content in one cell, the more its PI intensity. FACS analysis plots the distribution of cells with increasing PI intensity and gives a specific cell cycle profile for each cell population as a function of PI intensity. Based on total DNA content, cells from different S-phase fractions are sorted (Fig. 1a). Total DNA is then prepared from an equal number of cells representing the different stages of the S-phase. BrdU-labeled DNA is sonicated and immunoprecipitated from these S-phase-specific fractions to enrich for genomic sequences that replicate during the labeling period. Finally, quantitative real-time PCR is used for quantification of specific sequences in the different S-phase fractions.

2 Materials

1. 5-Bromo-2′-deoxyuridine (BrdU): Make aliquots of 1 M BrdU in deionized water and store at −20 °C.
2. Cell culture medium appropriate for cell type: in our experiments, we used serum-free insect culture medium supplemented

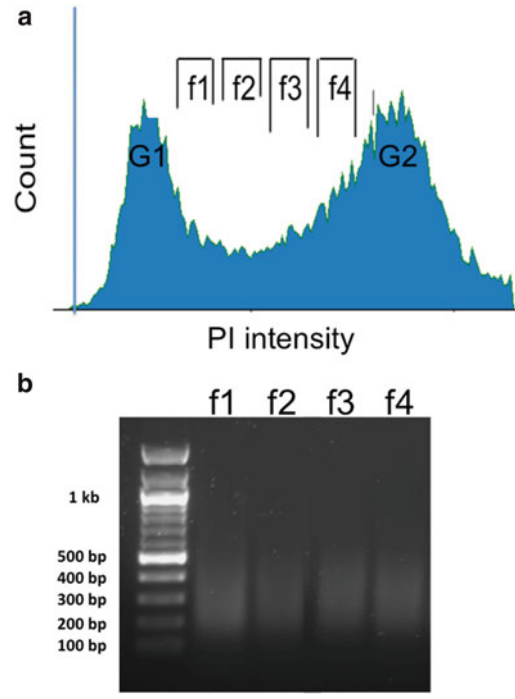


Fig. 1 *Drosophila* S2 cells sorting and DNA sonication. (**a**) Cell-cycle profile of *D. melanogaster* S2 cells obtained by flow cytometry after ethanol fixation and propidium iodide (PI) staining. The variable in the *x*-axis is PI intensity, while the variable in the *y*-axis is the frequency of cells showing specific PI intensity values. Peaks in the distribution represent G1 and G2 phases with a 2 N and 4 N DNA content, respectively. Cells between the G1 and G2 peaks are in S-phase. Gates indicate the sorted fractions: f1 represents the earliest and f4 the latest S-phase fraction. (**b**) Representative gel image showing sonicated DNA of the four sorted S-phase fractions (f1–f4)

with penicillin/streptomycin for *Drosophila* S2 cells and Schneider's medium supplemented with penicillin/streptomycin for *Drosophila* S3 cells.

3. 10× Phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl,10 mM Na_2HPO_4, 2 mM KH_2PO_4.
4. Propidium iodide (PI): make aliquots of 1 mg/ml PI in deionized water, filter, and store at +4 °C in the dark.
5. RNase A: dissolve RNase in deionized water at 10 mg/ml concentration. Make aliquots and store at −20 °C.
6. Proteinase K: dissolve proteinase K in deionized water at 20 mg/ml concentration. Make aliquots and store at −20 °C.
7. 20 mg/ml Glycogen.
8. Phenol-chloroform-isoamyl alcohol 25:24:1.
9. 1× TE: 10 mM Tris–HCl, 1 mM EDTA.
10. 100% Ethanol.

11. 70% Ethanol.
12. 3 M pH 5,2 Na acetate.
13. Lysis buffer: 50 mM Tris–HCl pH8, 10 mM EDTA, 0,8% SDS. Make fresh and store at room temperature.
14. 10× Immunoprecipitation buffer (IP) buffer: 0.1 M Sodium phosphate buffer pH 7.0, 1.4 M NaCl, 0.5% Triton X-100. Make fresh and store at room temperature.
15. Adjusting buffer: 110 mM Sodium phosphate buffer pH 7.0, 1.54 M NaCl, 0.55% Triton X-100. Make fresh and store at room temperature.
16. Digestion buffer: 50 mM Tris–HCl pH 8, 10 mM EDTA, 0.5% SDS, 250 μg/ml proteinase K. Make fresh and store at room temperature.
17. Mouse anti-BrdU DNA monoclonal antibody.
18. Protein AG plus agarose beads.
19. Wheel for rocking of tubes.
20. Nylon mesh 37 μm.
21. 5 ml Round-bottom polystyrene tube.
22. Vortex.
23. Sonicator.
24. Flow cytometer for fluorescence-activated cell sorting (FACS).

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 BrdU Labeling and Collection of Asynchronous Cycling Cells

1. Culture exponentially growing cells in the presence of 50 μM bromodeoxyuridine (BrdU) for a time that depends on the extension of S-phase (60 min is enough, for *Drosophila* S2 and S3 cells) (*see* **Note 1**).
2. Aliquot 5×10^6 cells per tube and pellet cells by centrifugation at $2000 \times g$ (*see* **Note 2**).
3. Wash twice with cold PBS.

3.2 Ethanol Fixation and PI Staining

1. Resuspend each aliquot of cells in 0.5 ml of cold 1× PBS and fix with 5 ml of 70% cold ethanol added drop by drop while gently vortexing.
2. Incubate for 1 h on ice or store at –20 °C. PAUSE POINT (*see* **Note 3**).
3. Centrifuge at approximately $2000 \times g$ for 5 min at 4 °C and remove supernatant carefully.

4. Add 2 ml of cold 1× PBS .
5. Repeat **steps 3** and **4**.
6. Centrifuge at approximately 2000×*g* for 5 min at 4 °C and remove supernatant carefully.
7. Resuspend pellets in 500 µl of PBS with RNase A at a final concentration of 1 mg/ml and incubate for 30 min at 37 °C.
8. Add propidium iodide at a final concentration of 20 µg/ml and incubate for 30 min in the dark at 4 °C or in ice.
9. Filter cells by pipetting them through 37 µm nylon mesh into a 5 ml polystyrene round-bottom tube (*see* **Note 4**).
10. Keep samples on ice in the dark and proceed directly to FACS sorting.

3.3 FACS Sorting

1. At the flow cytometer, select two or four gates containing a comparable number of cells and corresponding to S-phase fractions (Fig. 1a).
2. Sort equal numbers of cells (50,000–200,000) from different S-phase fractions into collecting tubes containing 2× lysis buffer supplemented with 0.2 mg of proteinase K per ml (*see* **Note 5**).
3. Incubate the aliquots at 50 °C for 2 h in lysis buffer and then store at –20 °C. PAUSE POINT (*see* **Note 6**).

3.4 DNA Extraction

1. Add one volume of phenol-chloroform, vortex, centrifuge at 13,000×*g* for 10 min at room temperature, and collect the upper phase (containing the DNA) in a fresh 1.5 ml tube.
2. Add to the lower phase one additional volume of TE 1×, vortex, centrifuge at 13,000×*g* for 10 min at room temperature, and add the upper phase to the one collected before. This step increases the yield of extracted DNA.
3. Precipitate DNA by adding 1 µl of glycogen, 1/10 volume of 3 M sodium acetate, and 2 volumes of 100% ethanol and storing overnight at –20 °C or 1 h at –80 °C.
4. Pellet DNA by centrifugation at 16,000×*g* for 60 min at 4 °C.
5. Discard the supernatant and add 70% ethanol to the pellet.
6. Centrifuge at 16,000×*g* for 5 min at 4 °C, remove all ethanol, and air-dry the pellet.
7. Resuspend the pellet in 500 µl of 1× TE at 37 °C for 5–10 min with mild agitation.

3.5 DNA Sonication and Immunoprecipitation

1. Sonicate the extracted DNA to an average size of 300 bp (*see* **Note 7**), Fig. 1b.
2. Use an aliquot of 25–50 µl to check sonication on agarose gel. Add 1× TE to the aliquot to reach 500 µl of final volume.

Precipitate with ethanol and sodium acetate as indicated in Subheading 3.4, **step 3**, and resuspend in an appropriated volume of water for agarose gel loading (*see* **Note 8**).

3. Heat denature the rest of sonicated DNA for 10 min at 95 °C and cool on ice for 2 min.
4. Add 50 μl of adjusting buffer and 700 μg of mouse anti-BrdU DNA monoclonal antibody to each tube.
5. Incubate tubes on the wheel with constant rocking at room temperature in the dark for 2 h.
6. Add 30–50 μl of protein AG plus agarose beads and incubate for an additional hour at room temperature with rocking in the dark.
7. Pellet DNA-protein complexes by centrifuging at max speed for 5 min at 4 °C in a microfuge.
8. Remove supernatant completely. Sometimes more than one centrifugation step are required to avoid removing of the pellet.
9. Wash DNA-protein complexes by adding 750 μl of 1× IP buffer previously chilled in ice.
10. Incubate for 5 min on the wheel at room temperature.
11. Centrifuge at 16,000 × *g* for 5 min at 4 °C.
12. Remove supernatant.
13. Resuspend pellets in 200 μl of digestion buffer and incubate samples overnight at 37 °C.
14. Next day, add 100 μl of fresh digestion buffer freshly supplemented with proteinase K and digest for 1 h at 50 °C.
15. Extract and precipitate as indicated in **steps 1–6** in Subheading 3.4. Let the pellet dry, resuspend in 40 μl of 1× TE, and store at 4 °C. DNA is now ready for replication timing analysis for up to 1 month.

3.6 Replication Timing Analysis by Real-Time Quantification

1. Design real-time PCR primers specific for each analyzed locus.
2. As control of the quality of S-phase-sorted fraction, it is important to calculate the relative abundance of sequences that have been previously characterized as early and late replicating. Moreover, in order to have an internal control for normalization, it is important to use a sequence which abundance is constant in the different S-phase fractions (*see* **Note 9**). Listed below are the sequences of primers that can be used as early and late S-phase control as well as mitochondrial control in *Drosophila* S2 and S3 embryonic cell lines [12]:

   Early ctrl forward: 5′GGCGTGGCCTCATCGGATGG3′.

   Early ctrl reverse: 5′ACGAGTCCTGCCGCAAAGCC3′.

   Grim forward: 5′AAAGGCCTGGTTCGGCTGGC3′.

   Grim reverse: 5′TTGCTACTTGCCGTGCGCGA3′.

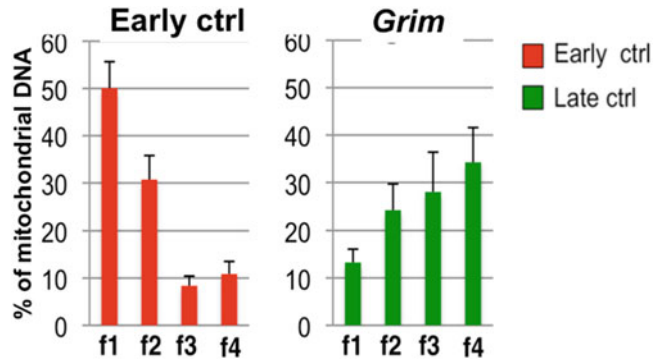


Fig. 2 Histograms representing replication timing of an early- and a late-replicating control locus in *Drosophila* S2 cells. Enrichment of BrdU-containing DNA in the four FACS-sorted S-phase fractions as quantified by real-time PCR with primers specific for positive controls for the early and late S-phase (early ctrl and *Grim* locus, respectively). The relative abundance of locus-specific DNA in each cell-cycle fraction is calculated from the average values of threshold cycle (Ct), normalized to the Ct of a mitochondrial sequence as internal control, using the following equation: $(2^{-(Ct_{mit}-Ct_i)}/\Sigma\, 2^{-(Ct_{mit}-Ct_i)}) \times 100$ (*see* **Notes 9** and **10**). Data points are generated from an average of four independent experiments

Mitochondria forward : 5′AGCAACAGGATTCCACGGAATTC3′.

Mitochondriareverse: 5′ATCATGCAGCTGCTTCAAAACCA3′.

3. Perform real-time PCR. The relative abundance of locus-specific DNA in each cell-cycle fraction can be calculated with the average values of threshold cycle (Ct), normalized to the Ct of a unique mitochondrial sequence as internal control using the following equation: $(2^{-(Ct_{mit}-Ct_i)}/\Sigma\, 2^{-(Ct_{mit}-Ct_i)}) \times 100$ (*see* **Note 10** for a practical example).

In Fig. 2 it is represented the relative abundance of *Drosophila* S2 early- and late-replicating S-phase sequences after sorting of four S-phase fractions (f1–f4). Early- and late-replicating sequences should show the highest enrichment, respectively, in the earliest (f1) and in the latest (f4) S-phase fractions.

4 Notes

1. For adherent cells, replace cell culture medium with fresh medium supplemented with 50 μM BrdU for a more homogeneous BrdU incorporation.
2. A total number of 50×10^6 starting cells would in general be enough for your analysis, but you should evaluate the number of required starting cells empirically. The final recovery of DNA at the end of the protocol would vary depending on the efficiency of the several protocol steps (*see* also **Note 9**).

3. 70% Ethanol fixes and permeabilizes cells for subsequent PI staining of DNA. At this step, fixed cells can be stored at −20 °C in the dark for up to 1 month.
4. Propidium iodide (PI) binds to DNA by intercalating between the bases with little or no sequence preference and with a stoichiometry of one dye per 4–5 base pairs of DNA. By flow cytometry it is possible to visualize the cell-cycle profile of a cell population on the basis of total DNA content that is a function of PI intensity (Fig. 1a). This profile can change in the same population based on cell concentration and PI staining efficiency. When processing a high number of cells, in order to reduce variability between each aliquot containing 5×10^6 cells, pool together all aliquots from the same population during filtration and divide them again before sorting. Perform sorting as soon as possible in order to reduce variability due to PI intensity bleaching and pay attention to keep cells in the dark and in ice.
5. To block DNA synthesis in specific S-phase fractions, cells are collected in 2× lysis buffer. Since cells are resuspended in 1× PBS during sorting the 2× lysis buffer will be diluted during cell collection and at the end of sorting you will have your DNA in 1× lysis buffer.
6. At this step, lysates can be stored at −20 °C for up to 1 month.
7. This step is crucial. You should be able to obtain fragments that are long enough to be amplified with primers specific for the locus of interest. On the other hand, they should be short enough to be easily immunoprecipitated. In our experimental conditions, an average fragment size of 300 bp works well. The time length and the intensity of sonication should be determined empirically. For example with a Diagenode Bioruptor sonicator UCD-200 sonication is performed at low intensity with ten cycles of 30-s on and 30-s off.
8. Control of sonicated DNA on agarose gel gives information about sonication efficiency and about the amount of DNA in the different S-phase fractions. If DNA is not enough sonicated, continue with sonication and check again on agarose gel before proceeding with the next steps. All the aliquots should be comparable in terms of sonication quality and total DNA amounts. If your samples show different DNA quantities, take equal amounts of DNA, adjust volumes, and proceed with the next step. In Fig. 1b there is a representative agarose gel showing sonicated DNA extracted from the four S-phase fractions (f1–f4) in S2 cells.
9. The described protocol consists of several steps and a lot of variables can affect the final result. In order to have an internal control for normalization, it is important to use a sequence which abundance is constant in the different S-phase fractions. Mitochondrial sequences are valid internal control because

they are replicated throughout S-phase and are equally represented in early and late S-phase fractions. Moreover, sequences that have been previously characterized as early and late replicating are important controls necessary to be sure that S-phase fractions have been sorted and processed correctly. A good early-replicating control should show the highest enrichment in the earliest S-phase fraction (f1) while a late-replicating control should show the highest enrichment in the latest fraction (f4). In our system, we used as early- and late-replicating control a previously characterized early genomic segment (Early ctrl) and the *Grim* locus, respectively (Fig. 2).

10. Mitochondrial sequences are abundant in the cell and are amplified more than ten cycles earlier with respect to other immunoprecipitated sequences. For normalization, it would be correct to use an appropriated dilution of immunoprecipitated DNA as template when amplifying with primers specific for mitochondrial sequences in order to obtain comparable ct values. Below there is a practical example of calculation of replication timing of early- and late-replicating controls in S2 cells. By using as real-time PCR template a 1:1 dilution of BrdU-immunoprecipitated DNA from four S-phase fractions (f1–f4) we obtained the following average ct values for early ctrl and *Grim*-specific primer pairs, respectively:

ct Early ctrl f1 = 33,47; ct Early ctrl f2 = 33,92; ct Early ctrl f3 = 35,71; ct Early ctrl f4 = 36,11

ct *Grim* f1 = 33,91; ct *Grim* f2 = 32,70; ct *Grim* f3 = 31,16; ct *Grim* f4 = 29,78.

With a 1:100 dilution of BrdU-immunoprecipitated DNA from f1 to f4 fractions as template we obtained the following average ct values for mitochondrial control:

ct mit f1 = 25,87; ct mit f2 = 26,03; ct mit f3 = 26,21; ct mit f4 = 26,31.

By applying the following equation $(2^{-(Ct_{mit}-C_i)} / \Sigma\ 2^{-(Ct_{mi}-C_i)}) \times 100$ for each S-phase fractions we obtained the following enrichment values for early ctrl locus:

$$f1 = (2^{-(33,47-25,87)} / (2^{-(33,47-25,87)} + 2^{-(33,92-26,03)} + 2^{-(35,71-26,21)} + 2^{-(36,11-26,31)})) \times 100 = 43,54$$
$$f2 = (2^{-(33,92-26,03)} / (2^{-(33,47-25,87)} + 2^{-(33,92-26,03)} + 2^{-(35,71-26,21)} + 2^{-(36,11-26,31)})) \times 100 = 35,39$$
$$f3 = (2^{-(35,71-26,21)} / (2^{-(33,47-25,87)} + 2^{-(33,92-26,03)} + 2^{-(35,71-26,21)} + 2^{-(36,11-26,31)})) \times 100 = 11,62$$
$$f4 = (2^{-(36,11-26,31)} / (2^{-(33,47-25,87)} + 2^{-(33,92-26,03)} + 2^{-(35,71-26,21)} + 2^{-(36,11-26,31)})) \times 100 = 9,44$$

and the following enrichment values for *Grim* locus:

$$f1 = (2^{-(33,91-25,87)} / (2^{-(33,91-25,87)} + 2^{-(32,70-26,03)} + 2^{-(31,16-26,21)} + 2^{-(29,78-26,31)})) \times 100 = 2,80$$
$$f2 = (2^{-(32,70-26,03)} / (2^{-(33,91-25,87)} + 2^{-(32,70-26,03)} + 2^{-(31,16-26,21)} + 2^{-(29,78-26,31)})) \times 100 = 7,20$$
$$f3 = (2^{-(31,16-26,21)} / (2^{-(33,91-25,87)} + 2^{-(32,70-26,03)} + 2^{-(31,16-26,21)} + 2^{-(29,78-26,31)})) \times 100 = 23,81$$
$$f4 = (2^{-(29,78-26,31)} / (2^{-(33,91-25,87)} + 2^{-(32,70-26,03)} + 2^{-(31,16-26,21)} + 2^{-(29,78-26,31)})) \times 100 = 66,16$$

In Fig. 2 these values have been put in average with those of other three biological replicates.

References

1. Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190:372–373, Epub 1961/04/22
2. Taylor JH (1960) Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. J Biophys Biochem Cytol 7:455–464, Epub 1960/06/01. PubMed Central PMCID: PMCPMC2224838
3. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE et al (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 107:139–144, Epub 2009/12/08. doi: 0912402107 [pii]10.1073/pnas.0912402107. PubMed PMID: 19966280; PubMed Central PMCID: PMC2806781
4. Schwaiger M, Stadler MB, Bell O, Kohler H, Oakeley EJ et al (2009) Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. Genes Dev 23:589–601, Epub 2009/03/10. doi: 23/5/589 [pii]10.1101/gad.511809. PubMed PMID: 19270159; PubMed Central PMCID: PMC2658520
5. Zhou J, Ermakova OV, Riblet R, Birshtein BK, Schildkraut CL (2002) Replication and subnuclear location dynamics of the immunoglobulin heavy-chain locus in B-lineage cells. Mol Cell Biol 22:4876–4889, Epub 2002/06/08. PubMed PMID: 12052893; PubMed Central PMCID: PMC133899
6. Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M et al (2010) Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res 20:155–169, Epub 2009/12/03. doi: gr.099796.109 [pii]10.1101/gr.099796.109. PubMed PMID: 19952138; PubMed Central PMCID: PMC2813472
7. Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M et al (2008) Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol 6, e245, Epub 2008/10/10. doi: 08-PLBI-RA-1977 [pii]10.1371/journal.pbio.0060245. PubMed PMID: 18842067; PubMed Central PMCID: PMC2561079
8. Perry P, Sauer S, Billon N, Richardson WD, Spivakov M et al (2004) A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction. Cell Cycle 3:1645–1650, Epub 2004/12/22. doi: 1346 [pii]. PubMed PMID: 15611653
9. Simon I, Tenzen T, Mostoslavsky R, Fibach E, Lande L et al (2001) Developmental regulation of DNA replication timing at the human beta globin locus. EMBO J 20:6150–6157, Epub 2001/11/02. doi: 10.1093/emboj/20.21.6150. PubMed PMID: 11689454; PubMed Central PMCID: PMC 125288
10. Renard-Guillet C, Kanoh Y, Shirahige K, Masai H (2014) Temporal and spatial regulation of eukaryotic DNA replication: From regulated initiation to genome-scale timing program. Semin Cell Dev Biol 30C:110–120, Epub 2014/04/15. doi: S1084-9521(14)00085-8 [pii]10.1016/j.semcdb.2014.04.014
11. Lanzuolo C, Lo Sardo F, Diamantini A, Orlando V (2011) PcG complexes set the stage for epigenetic inheritance of gene silencing in early S phase before replication. PLoS Genet 7, e1002370, Epub 2011/11/11. doi: 10.1371/journal.pgen.1002370PGENETICS-D-11-00832 [pii]. PubMed PMID: 22072989; PubMed Central PMCID: PMC3207895
12. Lanzuolo C, Lo Sardo F, Orlando V (2012) Concerted epigenetic signatures inheritance at PcG targets through replication. Cell Cycle 11(7):1296–1300, Epub 2012/03/17. doi: 19710 [pii]. PubMed PMID: 22421150
13. Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol 9:1167–1174, Epub 2007/09/11. doi: ncb1637 [pii]10.1038/ncb1637. PubMed PMID: 17828248
14. Lo Sardo F, Lanzuolo C, Comoglio F, De Bardi M, Paro R et al (2013) PcG-mediated higher-order chromatin structures modulate replication programs at the Drosophila BX-C. PLoS Genet 9, e1003283, Epub 2013/02/26. doi: 10.1371/journal.pgen.1003283PGENETICS-D-12-01096 [pii]. PubMed Central PMCID: PMCPMC3578750
15. Gilbert DM (1986) Temporal order of replication of Xenopus laevis 5S ribosomal RNA genes in somatic cells. Proc Natl Acad Sci U S A 83:2924–2928, Epub 1986/05/01. PubMed Central PMCID: PMCPMC323419
16. Azuara V (2006) Profiling of DNA replication timing in unsynchronized cell populations. Nat Protoc 1:2171–2177, Epub 2007/05/10. doi: nprot.2006.353 [pii]10.1038/nprot.2006.353
17. Mariani BD, Schimke RT (1984) Gene amplification in a single cell cycle in Chinese hamster ovary cells. J Biol Chem 259:1901–1910, Epub 1984/02/10

18. Braunstein JD, Schulze D, DelGiudice T, Furst A, Schildkraut CL (1982) The temporal order of replication of murine immunoglobulin heavy chain constant region sequences corresponds to their linear order in the genome. Nucleic Acids Res 10:6887–6902, Epub 1982/11/11. PubMed Central PMCID: PMCPMC326972
19. Marchionni MA, Roufa DJ (1981) Replication of viral DNA sequences integrated within the chromatin of SV40-transformed Chinese hamster lung cells. Cell 26:245–258, Epub 1981/10/01. doi: 0092-8674(81)90307-X [pii]
20. Schubeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine M (2002) Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. Nat Genet 32(3):438–442. doi:10.1038/ng1005ng1005, Epub 2002/10/02[pii]. PubMed PMID: 12355067

Part II

Chapter 6

Noncoding RNA Interplay with the Genome

Davide Gabellini*

Abstract

The majority of our genome is transcribed to produce RNA molecules that are mostly noncoding. Among them, long noncoding RNAs (lncRNAs) are the most numerous and functionally versatile class.

LncRNAs have emerged as key regulators of gene expression at multiple levels.

This section describes bioinformatics aspects important for lncRNA discovery and molecular approaches to perform structure-function characterization of this exciting class of regulatory molecules.

Key words LncRNA, RNA-seq, Bioinformatics, RIP, RNA pull-down, ChIRP, Chart, RAP

Some four billion years ago, RNA was responsible for processing the information and the metabolic transformations needed for life to emerge from chemistry [1]. Long from being a molecular fossil, RNA continues to evolve purposes in parallel to its ubiquitous roles in protein translation. Indeed, the majority of our genome is transcribed to produce RNA molecules that, for the most part, display little or no protein coding potential [2, 3]. Noncoding transcription scales with organismal complexity [4] and noncoding RNAs (ncRNAs) participate to an extensive repertoire of processes to influence cellular behavior.

An expanding class of ncRNAs is made of long ncRNAs (lncRNA), operationally defined as transcripts longer than 200 nt and lacking the ability to be translated to a functional protein. Like protein-coding RNAs, most lncRNAs are transcribed by RNA polymerase II and thus display features such as a 5′-cap, polyadenylation, splicing, and H3K4me3 and H3K36me3 enrichment over the promoter and gene body, respectively [5]. According to the position and orientation of transcription, lncRNAs can be defined as intergenic, bidirectional, overlapping, antisense, or intronic [6]. LncRNAs can also be classified according to their location as cytoplasmic, nucleoplasmic, or chromatin associated.

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_6,

The lncRNA transcriptome is at least as large and complex as the coding one [7]. While lncRNAs are usually lowly expressed, their transcription is often tissue- and cell-type specific, much more than protein-coding RNAs. Moreover, lncRNA expression is highly dynamic, being context and time dependent and suggesting that lncRNAs are intimately linked with cell phenotypes. Accordingly, lncRNAs are taking a center stage in understanding higher eukaryote complexity [8].

The recent advent of relatively low-cost, next-generation RNA sequencing (RNA-Seq) technologies is providing an ever-increasing list of lncRNAs. However, annotating a locus as a bona fide lncRNA gene is not an easy task and current lists of lncRNAs likely include both properly and improperly annotated RNAs. As described in the chapter by Arrigoni et al., there are several critical aspects to consider in transcript assembly, especially using bulk short-read RNA-seq data. Sequencing depth is a major issue, as the availability of more reads leads to an increasing transcript variety and complexity. Another important aspect is the bioinformatics pipeline. For example, the combination of different transcripts reconstruction approaches allows for higher sensitivity and specificity. The goal is to use a logical and flexible pipeline for cataloging newly identified transcripts in order to facilitate their functional and mechanistic exploration.

LncRNAs are able to regulate gene expression at the transcriptional, posttranscriptional, or posttranslational level. LncRNAs perform their activities through RNA-DNA, RNA-RNA, or RNA-protein interactions. Similarly to proteins, a single lncRNA can contain several modular domains able to bind nucleic acids via base pairing or proteins by RNA structures, allowing to coordinate signals between different types of macromolecules. Nevertheless, general functional rules have not emerged yet, since structure-function information is available only for a few lncRNAs.

Given the fact that the majority of lncRNAs appear to function as part of a ribonuclear protein complex, the characterization of the lncRNA-bound proteome is crucial for understanding lncRNA functions. Protein- and RNA-centric technologies have been developed to identify lncRNA biochemical partners.

As discussed by Matarazzo, RNA immunoprecipitation (RIP) is a major protein-centric approach to identify RNAs associated to a specific protein. In RIP, antibodies specific for a protein of interest are used to enrich for RNA associated to the selected protein. There are several modifications of RIP based on the use or type of RNA-protein cross-linking agent and the restricted or genome-wide analysis of the immunopurified material. While RIP is a

powerful technique allowing to precisely map the binding site of the protein of interest to the RNA molecule, its major limitation is the availability of very specific antibodies.

One of the conventional RNA-centric methods to identify proteins associated to a given lncRNA is the in vitro RNA pull-down assay described by Barnes et al. The approach uses in vitro-transcribed and biotinylated RNA as a bait to fish out interacting proteins from a cell extract, followed by mass spectrometry analysis. Although this method has yielded important insights for many functional RNA molecules, its major limitations are that the RNA bait might not be present at physiological concentration or could not be correctly structured, thus leading to artifactual protein interactions.

Techniques allowing for a direct capture of endogenous RNPs using antisense oligonucleotide probes could avoid the complications associated to in vitro pull-down studies. There are three major direct RNP purification methods: chromatin isolation by RNA purification (ChIRP) described by Chu et al., capture hybridization analysis of RNA targets (Chart) described by Sexton et al., and RNA antisense purification (RAP) [9]. While all approaches use antisense DNA oligonucleotides to capture and purify specific lncRNA-chromatin complexes from cross-linked cells, they differ for the oligonucleotide probe design and for the method of cross-linking. In ChIRP and RAP, probes are simply tiled across the entire target RNA to cover all potential hybridization spots and ensure the capture of all fragments of the RNA. Instead, Chart uses RNase H to identify accessible regions on the target RNA and guide probe selection. While more time consuming, Chart could produce lower background due to the reduced number of probes used. Chart and ChIRP use formaldehyde or glutaraldehyde, which leads to protein-nucleic acid and protein-protein cross-linking. RAP uses ultraviolet cross-linking to create covalent bonds between directly interacting RNA and proteins. Hence, RAP should enrich for direct RNA-protein interactions.

While available tools for direct capture of endogenous RNPs allow identifying the genomic loci, proteins, and RNAs that interact with a given lncRNA, their current limitation is that so far they have been showed to work only on abundant lncRNAs. It will thus be important to test this approaches on lowly expressed lncRNAs, which represent the majority of lncRNAs.

The development of robust and high-throughput techniques for RNA structure-function characterization will lead to a new and exciting phase of lncRNA research.

References

1. Higgs PG, Lehman N (2014) The RNA World: molecular cooperation at the origins of life. Nat Rev Genet 16:7–17
2. Kapranov P et al (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res 15:987–997
3. Carninci P et al (2005) The transcriptional landscape of the mammalian genome. Science 309:1559–1563
4. Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29:288–299
5. Mallory AC, Shkumatava A (2015) LncRNAs in vertebrates: advances and challenges. Biochimie 117:3–14
6. Mattick JS, Rinn JL (2015) Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol 22:5–7
7. Clark MB et al (2015) Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. Nat Methods 12: 339–342
8. Gloss BS, Dinger ME (2015) The specificity of long noncoding RNA expression. Biochim Biophys Acta. doi:10.1016/j.bbagrm.2015.08.005
9. Chu C, Spitale RC, Chang HY (2015) Technologies to probe functions and mechanisms of long noncoding RNAs. Nat Struct Mol Biol 22:29–35

Chapter 7

RIP: RNA Immunoprecipitation

Miriam Gagliardi and Maria R. Matarazzo*

Abstract

The relevance of RNA-protein interactions in modulating mRNA and noncoding RNA function is increasingly appreciated and several methods have been recently developed to map them. The RNA immunoprecipitation (RIP) is a powerful method to study the physical association between individual proteins and RNA molecules in vivo. The basic principles of RIP are very similar to those of chromatin immunoprecipitation (ChIP), a largely used tool in the epigenetic field, but with some important caveats. The approach is based on the use of a specific antibody raised against the protein of interest to pull down the RNA-binding protein (RBP) and target-RNA complexes. Any RNA that is associated with this protein complex will also be isolated and can be further analyzed by polymerase chain reaction-based methods, hybridization, or sequencing.

Several variants of this technique exist and can be divided into two main classes: native and cross-linked RNA immunoprecipitation. The native RIP allows to reveal the identity of RNAs directly bound by the protein and their abundance in the immunoprecipitated sample, while cross-linked RIP leads to precisely map the direct and indirect binding site of the RBP of interest to the RNA molecule.

In this chapter both the protocols applied to mammalian cells are described taking into account the caveats and considerations required for designing, performing, and interpreting the results of these experiments.

Key words RNA immunoprecipitation, Native RIP, Cross-linked RIP, Protein-RNA interaction

1 Introduction

The interest in the interaction between proteins and RNAs as key aspect of gene regulation has increased over the last decade [1]. The growing expansion in sequencing technologies has facilitated the investigation of the transcriptome at unprecedented depth [2]. Therefore, the importance of messenger RNA (mRNA) processing, including alternative splicing, nuclear export, subcellular localization, and editing in producing diverse isoforms and in controlling the stability and translation of mRNAs, has largely reported [3–5]. Moreover, the identification of diverse classes of noncoding RNAs (ncRNAs), including many thousands of long noncoding RNAs (lncRNAs), has increasingly emerged [6, 7].

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_7,

All aspects of controlling gene expression, either by small regulatory RNAs, like microRNAs, and lncRNAs involve RNA-protein interactions. Indeed, RNA molecules can interact with proteins through their secondary or tertiary structure to create ribonucleoprotein complexes (RNPs). The central function of RNPs in mRNA processing and ncRNA function are well-established concepts. Thus, the main challenge for understanding RNA-mediated biological processes is identifying the RNAs associated with RNA-binding proteins (RBPs) in a cellular context.

Historically, individual mRNA targets have been identified using in vitro techniques such as cross-linking with ultraviolet light, nitrocellulose filter binding, and RNA electromobility shift assays (REMSAs; [8]). Although these methods have provided ample biochemical information, they are inadequate to identify unknown RNA targets when starting with an RBP. Furthermore, bioinformatic algorithms have been developed to search for novel mRNA targets of particular RBPs, but the efficacy of such approaches is not complete because they identify RNA-binding sites of few nucleotides which therefore appear more frequently among mRNAs than expected.

More recently, the predominant methods for exploring RNA-protein interactions are based on protein immunoprecipitation [9]. These methods require knowledge of the protein; therefore they are not useful for identifying the proteins that interact with a given RNA transcript. RNA immunoprecipitation (RIP) is an antibody-based technique used to identify RNA-protein interactions in vivo [10, 11]. Specific ribonucleoprotein complexes can be immunoprecipitated from a cellular lysate with an antibody raised against the protein of interest. Every RNA interacting with this protein complex can also be isolated and further examined by PCR-based methods, hybridization, or massive sequencing [11–13]. Bioinformatic tools have been developed to map reads to their transcripts of origin and to identify protein-binding sites, in case of cross-linked-based methods.

Binding maps of several RNA-binding proteins across the transcriptome have been created by using these techniques, thus providing key insights into how mRNA processing is regulated in the cell [14]. Also, early insights into the proteins interacting with lncRNAs, such as the protein of the Polycomb-repressive complex 2 (PRC2), have been gained with these approaches [15–17].

There are several variants of these methods, which can be divided into two main groups: native [12, 13, 17, 18] and cross-linked RNA immunoprecipitation [19–22]. Native methods detect RNA-protein complexes in physiological conditions. Although these approaches are valuable because preserve the native complexes existing in the cell, they nevertheless have several shortcomings. The first and well described is that ribonucleoproteins can re-associate after cell lysis, and therefore not accurately reproduce the interactions that occur in vivo [23].

Moreover, the presence of abundant transcripts with nonspecific interactions may lead to the underestimation of specific RNA-protein interactions. Indeed, ribosomal RNAs are often the largest contaminating RNA species in protein purifications [24]. Because of these concerns, the nature of the interactions detected by these methods has been quite debated. For instance, many lncRNAs, as well as mRNAs, were identified as interacting with PRC2 by using native protocols [13]. However, a recent study has claimed that virtually all transcripts may interact with PRC2 in the cell [25]. Thus, the biological significance of identified lncRNA-PRC2 interactions is currently subject of discussion, with scientists arguing that they are merely nonspecific interactions [26]. Yet, it is undoubted that some of these lncRNAs-PRC2 interactions have been confirmed and exhibit well-defined functional roles [15, 16, 27]. Given that the extent to which nonspecific RNA-protein associations are detected by the native approaches is not clearly quantifiable, the interactions identified with these methods often require further experimental validation, such as through the integration of multiple distinct experimental approaches [16, 28].

To prevent the re-association of proteins and RNA after cell lysis and to "freeze" their interactions in the cell, cross-linking agents can be used to fix all the interactions. UV cross-linking may be used to identify direct interactions between RNA and proteins with the limitation that UV-cross-linking is not reversible [19]. Cross-linking agents that are reversible may be more beneficial for subsequent characterization of the associated molecules [21]. One of the reversible cross-linking agents is formaldehyde, which is able to rapidly preserve cellular complexes in their native state, and to rapidly penetrate the cell membrane. These are the qualities that have led to its broad application in methods such as RNA immunoprecipitation [22]. An additional general weakness shared by both the approaches, the native and cross-linked RNA immunoprecipitation, concerns the quality of the antibody raised against the protein of interest. Indeed, the efficacy of the result is highly dependent on the antibody used and the abundance of the target ribonucleoprotein.

As shown in Fig. 1, either cross-linked or non-cross-linked RNA-protein complexes from living cells are involved in an RIP assay. These complexes are isolated by immunoprecipitation using a specific antibody towards the protein of interest. After reversing the cross-links, the interacting RNA can be analyzed by reverse transcription, followed by a polymerase chain reaction.

Immunoprecipitation of RNPs may be followed by genomic analysis using microarrays, known as RIP-Chip, or more recently using next-generation sequencing methods known as RIP-Seq. These are powerful high-throughput techniques for in vivo identifying RNA targets associated to specific proteins in cellular context. A considerable amount of bioinformatic analysis is necessary

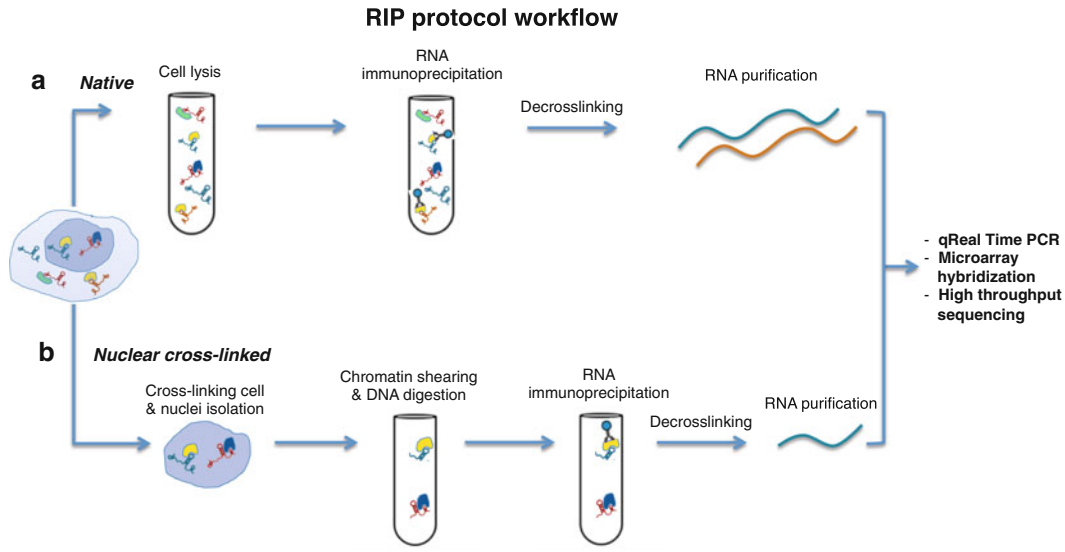


Fig. 1 Schematic representation of native and cross-linked RIP protocols. (**a**) For the native RIP, the harvested cells are directly lysed. The immunoprecipitation is performed using an antibody raised against the protein of interest. The RNA species are purified and analyzed by qPCR, microarray hybridization, and/or next-generation sequencing. (**b**) In the cross-linked protocol, live cells are treated by formaldehyde or another cross-linking agent (e.g., UV light) to fix the RNA/protein interactions. After the nuclear extraction, the chromatin is sheared and the DNA is degraded by DNase treatment. The immunoprecipitation is performed using specific antibody against the protein of interest. Following the immunoprecipitation and the reverse cross-linking to release the immunoprecipitated RNA, the RNA is extracted subsequently analyzed by qPCR, microarray hybridization, and/or next-generation sequencing

for processing raw array data to create a list of target gene mRNAs (or transcriptional fragments, in the case of RNA sequencing) as well as for statistical interpretation and analysis of the data. Primary data analysis often imposes further computational and experimental validation of putative identified associated RNA and potential RBP-binding sites [29, 30].

Here, we describe the detailed native and cross-linked RNA immunoprecipitation protocols allowing to select multiple RNA molecules, expressed in a specific cellular context, which are directly and/or indirectly interacting with the protein of interest.

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents.

1. Phosphate-buffered saline (PBS) pH 7.4 (stable at room temperature).
2. Polysome lysis buffer (10× PLB): 1000 mM KCl, 50 mM $MgCl_2$, 100 mM HEPES-NaOH pH 7, 5% Nonidet P-40

(NP-40). Prepare 10× stock buffer and store it at room temperature. Before using it, prepare PLB dilution to 1× and add 1 mM dithiothreitol (DTT), 200 units/ml RNase OUT, and EDTA-free Protease Inhibitor Cocktail.

3. NT-2 buffer (5×): 250 mM Tris-HCl pH 7.4, 750 mM NaCl, 5 mM $MgCl_2$, 0.25% NP-40. Store the stock buffer at 4 °C. Before use prepare NT-2 buffer 1× dilution.
4. NET-2 buffer: 1× NT-2 buffer supplemented with 20 mM EDTA pH 8.0, 1 mM DTT, 200 units/ml RNase OUT.
5. Formaldehyde solution: 50 mM HEPES-KOH, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 11% formaldehyde.
6. Cell lysis buffer: 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 0.5% NP-40. Before use add 1 mM DTT, 200 units/ml RNase OUT, and EDTA-free Protease Inhibitor Cocktail.
7. Nuclei resuspension buffer: 50 mM HEPES-NaOH pH 7, 10 mM $MgCl_2$. Before use add 1 mM DTT, 200 units/ml RNase OUT, and EDTA-free Protease Inhibitor Cocktail.
8. Immunoprecipitation buffer: 150 mM NaCl, 10 mM Tris–HCl pH 7.4, 1 mM EDTA, 1 mM EGTA pH 8, 1% Triton X-100, 0.5% NP-40. Before use add 1 mM DTT, 200 units/ml RNase OUT, and EDTA-free Protease Inhibitor Cocktail.
9. Proteinase K digestion buffer: 1× NT-2 buffer supplemented with 1% sodium dodecyl sulfate (SDS), 1.2 mg/ml Proteinase K.

3 Methods

3.1 Lysate Preparation (Native RIP)

1. Grow cells in an appropriate culture medium, stimulate or treat them, if necessary, and collect them when they are at ~80–90% of confluence (*see* **Note 1**).
2. Count cells using a hemacytometer. Consider to use ~2–5 mg of total protein extract for each immunoprecipitation which corresponds to ~$5–20 \times 10^6$ mammalian cells.
3. Collect cells by centrifugation at $1000 \times g$ for 5 min at 4 °C and discard the supernatant.
4. Wash cells twice with 1× ice-cold PBS. Collect cells by centrifugation at $1000 \times g$ for 5′ at 4 °C and discard the supernatant.
5. Resuspend cells in equal pellet volume of polysome lysis buffer. Pipette up and down to break clumps of cells (*see* **Note 2**).
6. Incubate on ice for 5 min.
7. Store at −80 °C to promote the cell lysis. The lysate may be stored for several months to −80 °C.

3.2 Preparation of Magnetic Beads and Immobilization of Antibodies (Native RIP)

1. Completely resuspend protein-A- or protein-G-coated magnetic beads before taking magnetic beads by pipetting or end-over-end rotation (*see* **Note 3**).
2. Add 75 µl protein-A or -G magnetic beads into a 1.5 ml tube and wash twice with 0.5 ml of NT-2 buffer.
3. Resuspend the beads in 100 µl of NT-2 and add 5 µg of the antibody of your interest and the negative control to the tubes.
4. Incubate with rotation for 1 h at room temperature to allow the binding of antibody to coated beads.
5. Centrifuge the tubes at 5000 × *g* for 15 s, place them on the magnetic rack, and remove the supernatant using a vacuum aspirator (*see* **Note 4**).
6. Remove the tubes from the rack, add 1 ml of NT-2, mix them by pipetting, spin at 5000 × *g* for 15 s, place the tube in the magnetic support, and remove the supernatant using a vacuum aspirator. Repeat this step another five times for a total amount of six washes.
7. After the sixth wash, resuspend the beads with 900 µl of NET-2 buffer and keep them in the ice (*see* **Note 5**).

3.3 Immunoprecipitation of Protein-RNA Complexes (Native RIP)

1. Centrifuge the cell lysate at 20,000 × *g* for 10 min at 4 °C, remove 100 µl of supernatant, and add it to each antibody-bead reaction. The final volume of the immunoprecipitation reaction will be 1 ml.
2. Take 10 µl (10%) of the cell lysate supernatant and place it in a new tube labeled "Input" (*see* **Note 6**). Store it at −80 °C until starting RNA immunoprecipitation.
3. Incubate all tubes on rotating wheel for 3 h up to overnight at 4 °C.
4. After the overnight incubation, spin down the tubes, place them on the magnetic support in the ice, incubate for 1 min, and discard the supernatant (*see* **Note 7**).
5. Remove the tubes from the magnet, add 1 ml of ice-cold NT-2 to each tubes, and vortex the samples vigorously.
6. Spin down the tubes, place them on the ice-cold magnetic separator, incubate for 1 min, and discard the supernatant.
7. Repeat **steps 5** and **6** another five times with 1 ml of ice-cold NT-2.

3.4 Lysate Preparation (Cross-Linked RIP)

1. Grow cells in an appropriate culture medium, stimulate or treat them, if necessary, and collect them when they are at ~80–90% of confluence (*see* **Note 1**).
2. Count cells using a hemacytometer. Consider to use ~2–5 mg of total protein extract for each immunoprecipitation, which corresponds to ~5–20 × 10^6 mammalian cells.

3. Add to cell suspension the necessary volume of formaldehyde solution to have 1 % final concentration. Incubate for 10 min at room temperature (*see* **Note 8**).
4. Stop the cross-linking reaction adding one-tenth the volume of 2.66 M glycine, and incubate for 5 min at room temperature and then 10 min on ice (*see* **Note 9**).
5. Wash cells twice with 1× ice-cold PBS. Collect cells by centrifugation at 1000 × *g* for 5′ at 4 °C and discard the supernatant.
6. Resuspend cells in 4 ml cell lysis buffer and incubate for 10–15 min in ice.
7. Homogenize by Dounce with 10 strokes with pestle A and 40 strokes with pestle B to allow the release nuclei.
8. Recover nuclei by centrifugation at 1000 × *g* for 10 min at 4 °C.
9. Resuspend nuclei with 3 ml of nuclei resuspension buffer and sonicate the nuclei to obtain DNA fragments in a range between 1000 and 200 bp.
10. After the sonication, add 250 unites/ml of DNase to the chromatin and incubate for 30 min at 37 °C.
11. Stop the DNase reaction adding EDTA to a final concentration of 20 mM.

    Adjust the sample with 1 % Triton X-100, 0.1 % sodium deoxycholate, 0.01 % SDS, and 140 mM NaCl.

3.5 Preparation of Magnetic Beads and Immobilization of Antibodies (Cross-Linked RIP)

1. Completely resuspend protein-A- or protein-G-coated magnetic beads before taking magnetic beads by pipetting or end-over-end rotation (*see* **Note 3**).
2. Add 75 µl protein-A or -G magnetic beads into a 1.5 ml tube and wash twice with 0.5 ml of nuclei resuspension buffer supplemented with 1 % Triton X-100, 0.1 % sodium deoxycholate, 0.01 % SDS, and 140 mM NaCl (*see* **Note 4**).
3. Resuspend the beads in 100 µl of complete nuclei resuspension buffer and add 5 µg of the antibody of your interest and the negative control to the tubes.
4. Incubate with rotation for 1 h at room temperature to allow the binding of antibody to coated beads.
5. Centrifuge the tubes at 5000 × *g* for 15 s, place them on the magnetic rack, and remove the supernatant using a vacuum aspirator.
6. Remove the tubes from the rack, add 1 ml of complete nuclei resuspension buffer, mix them by pipetting, spin at 5000 × *g* for 15 s, place the tube in the magnetic support, and remove the supernatant using a vacuum aspirator. Repeat this step another five times for a total amount of six washes.

7. After the sixth wash, resuspend the beads with 75 µl of complete nuclei resuspension buffer and keep them in the ice.

3.6 Immunoprecipitation of Protein-RNA Complexes (Cross-Linked RIP)

1. Centrifuge the cell lysate at 20,000 × *g* for 10 min at 4 °C, remove 975 µl of supernatant, and add it to each antibody-bead reaction. The final volume of the immunoprecipitation reaction will be 1 ml.
2. Take 9.75 µl (1%) of the cell lysate supernatant and place it in a new tube labeled "Input" (*see* **Note 6**). Store it at −80 °C until starting RNA immunoprecipitation.
3. Incubate all tubes on rotating wheel for 3 h up to overnight at 4 °C.
4. After the overnight incubation, spin down the tubes, place them on the magnetic support in the ice, incubate for 1 min, and discard the supernatant.
5. Remove the tubes from the magnet, add 1 ml of ice-cold Immunoprecipitation buffer to each tubes, and vortex the samples vigorously.
6. Spin down the tubes, place them on the ice-cold magnetic separator, incubate for 1 min, and discard the supernatant.
7. Repeat **steps 5** and **6** another five times with 1 ml of ice-cold immunoprecipitation buffer.

3.7 RNA Purification (Native and Cross-Linked RIP)

1. Resuspend each immunoprecipitate in 150 µl of Proteinase K buffer. To each "Input" add 107 µl of NT-2, 15 µl SDS 10%, and 18 µl of Proteinase K, to reach a total volume of 150 µl. Incubate all tubes at 55 °C for 30 min with shaking to digest the proteins.
2. After 30 min of incubation, spin down all tubes, place them in the magnetic rack, transfer the supernatants in a new tubes, and add to each of them 250 µl of NT-2.
3. Add 400 µl of phenol:chloroform:isoamyl alcohol (125:24:1) to each tube and vortex vigorously for 15 s. Centrifuge the tubes at 20,000 × *g* for 10 min at room temperature to separate the phases.
4. Carefully remove 350 µl of the aqueous phase without disturbing the protein interface. Place it in a new tube and add 400 µl of chloroform. Vortex the tubes for 15 s and centrifuge them at 20,000 × *g* for 10 min at room temperature to separate the phases.
5. Gently take 300 µl of the aqueous phase and place it in a new tube. To each tube add 50 µl of ammonium acetate 5 M, 15 µl of LiCl 7.5 M, 5 µl glycogen (5 mg/ml), and 850 µl of absolute ethanol.

6. Keep at −80 °C for 1 h to overnight to allow the RNA precipitation, then centrifuge at 20,000 × *g* for 30 min at 4 °C, and discard the supernatant gently.
7. Wash the pellet with 500 µl of cold 80 % ethanol.
8. Centrifuge at 20,000 × *g* for 15 min at 4 °C. Discard the supernatant carefully and air-dry the pellets.
9. Resuspend the pellets in 10–20 µl of RNase-free water and place the tube on ice.
10. Treat all the volume of each sample with DNase to remove residual contaminant DNA in the further analysis.

3.8 Gene-Specific Studies

1. The reverse transcription of the immunoprecipitated and input RNAs may be carried out with any commercially available reverse transcription enzyme and kit that use random examers as primers. Consider preparing two RT+ and one RT− for each sample, starting from the same volume of DNase-treated RNA.
2. After the cDNA synthesis of the positive and negative IPs and input, perform the real-time PCR preparing the dilution 1:8 with H_2O of all cDNAs [e.g., to have 16 µl of cDNA dilution add 2 µl cDNA stock + 14 µl pure H_2O].
3. For the protocol of the single reaction follow the datasheet of chosen supermix. To robust results it is important to perform triplicates for each samples.

3.9 qPCR Analysis

1. Use the software of the real-time instrument to monitoring the amplification reaction. At the end of the run, in the log-scale view, the slopes of the amplification curves for all the assays should be parallel to each other to be comparable.
2. In the log-scale view of the amplification curve, manually position the threshold near the mid-point of the linear range. The threshold value should be the same for all the triplicate reactions in the same gene study.
3. Check the dissociation curve to confirm that each reaction produces a single specific product. In this case the chart should appear a single peak at a melting temperature (Tm) greater than 75 °C.
4. Export all Ct with appropriate labels in an Excel spreadsheet.
5. Calculate the average Ct between replicates.
6. Normalize each IP fractions' Ct average to the input fraction Ct average for the same qPCR Assay (ΔCt) to account for sample preparation differences:

$$\mathbf{\Delta Ct[normalized\ RIP]} = (\text{Average Ct[RIP]} - (\text{Average Ct[Input]} - \log_2(\text{Input Dilution Factor})))$$

where Input Dilution Factor = (fraction of the input RNA saved)$^{-1}$

7. Calculate the % input for each RIP fraction (linear conversion of the normalized RIP ΔCt):

   $$\textbf{\% Input} = 2^{(-\Delta Ct[\text{normalized RIP}])}$$

8. It is possible to adjust the normalized RIP fraction Ct value for the normalized background (negative control) fraction Ct value (first ΔΔCt):

   $$\textbf{\Delta\Delta Ct[negative control]} = \Delta Ct[\text{normalized ChIP}] - \Delta Ct[\text{negative control}]$$

9. Calculate RIP fold enrichment above the sample-specific background (linear conversion of the first ΔΔCt):

   $$\textbf{Fold Enrichment} = 2^{(-\Delta\Delta Ct\,[ChIP/NIS])}$$

3.10 Microarray Processing

Several platforms are available to perform the microarray hybridization assay after the purifying the immunoprecipitated RNAs. They share the same limit that is based on the supervised concept of technique. In fact, the microarray method allows investigating the presence of annotated genes in the immunoprecipitated material. Bioinformatic analysis of the data coming from these experiments is commonly performed with the "Human Gene 1.0 ST Array" from Affimetrix and the Agilent's "GeneSpring GX10" software. A typical analysis is carried out on triplicates of each sample (INPUT RNA and positive and negative IPs). The replicates in each set are subject to the basic quality control including the correlation study and the principal component analysis (PCA). PLIER16 algorithm, an iterative pipeline that is followed by the filtering for the 20th–100th percentile, is often used to obtain the gene abundance estimation in each set. Using the same algorithm it is possible to extract a list of genes that show a minimum twofold increase of measured abundance in the treatment versus input. A one-way analysis of variance (ANOVA) may be performed on this RNA subset.

3.11 Sequencing

1. The analysis of whole transcriptome interacting with a specific protein (RIP-Seq) may be performed using the next-generation sequencing. This technology is potentially unbiased if the choice of the library preparation is correct. For instance, the poli(A) separation will reduce much the RNA population if most of bound RNAs are noncoding RNAs (ncRNAs) or the introduction of the size selection step will discard the small ncRNAs.
2. Another bias that must be considered is that considering the less abundant RNAs. For this reason a depth of sequencing minimum of ~30 million of reads is suggested [29].
3. After the sequencing, the produced reads can be aligned to the reference genome with either Tophat or Botwie tools.
4. For the native RIP-seq, the estimation of the abundance of each RNA molecules in the samples can be analyzed using

programs to assemble transcriptomes from RNA-Seq data and quantify their expression [e.g., Cufflinks]. The estimated abundance of each type of RNA in the RIP sample can be normalized against the proper input and then compared with the values in the negative control, after the normalization.

5. In case of cross-linked RIP-seq, the analyses after mapping the reads are different from the native procedure because only RNA fragments selected for the interaction with the protein of interest should be enriched in the immunoprecipitated sample. The position of the protein-binding site on the RNA transcript is mapped by using the peak caller algorithms as used for the ChIP [e.g., MACS1.4]. After detecting those positions, it is also possible to evaluate the presence in the peak of a frequent motif with appropriate motif discovery tools [e.g., MEME suite]. These type of studies allow understanding if the binding of protein is mediated by a sequence recognition.

4 Notes

1. The generic precaution on working with RNA is to use instruments, tips, and tubes DNase and RNase free. Gloves, benches, and pipettes may be accurately cleaned before the use.
2. During cell lysate preparations occasionally vortex the cell pellet-PLB mixture to promote thawing. Once the cells are fully thawed, vortex vigorously to allow cell lysis. Poor vortexing may result in a low amount of protein-RNA available for the immunoprecipitation.
3. The type of beads utilized for the immunoprecipitation depends on the immunoglobulin isotype and species. It is useful to check the bead manufacturer's binding chart to determine the best choice of beads. In most cases, mouse monoclonal antibodies have stronger affinity for protein G and rabbit polyclonal antibodies have stronger affinity for both protein G and protein A. To solve this problem it is possible to prepare a 1:1 stock mixture of protein A- and protein G-coated magnetic beads, washed with the proper buffer, and stored at 4 °C with 0.02 % sodium azide.
4. To remove the supernatant, place the tubes in the magnetic support and wait for the complete settling of beads on the tube site that interact with the magnet. While aspirating the supernatant, be sure to change tips between the samples.
5. The presence of EDTA in the NET-2 buffer avoids the immunoprecipitation of ribosomal RNA disrupting the interaction between these molecules and the ribosomal proteins.
6. The input will be used to generate the standard curve in the further real-time PCR analysis. It is essential to compare the

sample to a negative control since detecting the enrichment from a specific RNA alone may not indicate an existent interaction. One control is to normalize the level of an RNA observed after purification to its abundance in total lysate (the input sample in RIP assay). Moreover, interactions can also occur due to unspecific associations with the purification resin or other reagents of the procedure. To measure these artifactual associations, other proteins can be used as negative controls. However, the negative control should be carefully selected, as a non-RNA-binding protein is likely to have lower nonspecific RNA binding.

7. The immunoprecipitation efficiency of the chosen antibody should be verified by Western blot in SDS-PAGE. Therefore, it is recommended to store additional 10 μl of the input material and compare it with the abundance of the protein of interest in 100 μl of the first unbound and 100 μl of the bead suspension after the sixth wash.
8. To efficiently cross-link adherent cells, after harvesting and counting them, it is necessary to wash them with ice-cold PBS and perform the cross-linking in cold PBS. As in the ChIP assay, the time of the cell exposure to the formaldehyde is important to obtain optimal results. A too short time of incubation with formaldehyde leads to the underestimation of RNA-protein interaction due to the incomplete cell fixation. A too long time of cross-linking increases the number of protein fixed together, thereby reducing the shearing efficiency and increasing nonspecific signals.
9. To perform an efficient sonication is a crucial step to obtain a good quality of immunoprecipitation. The sonication is strictly dependent on the cell type, and the density and amount of treated cells. It is also important to preserve protein degradation incubating the sample at 4 °C and using a medium power intensity with the alternation of the pulse step with an OFF step to avoid the increase of the temperature. The sonication efficiency could be analyzed on an aliquot of sheared chromatin which has to be de-cross-linked at 56 °C for 4 h. After the reverse cross-linking the DNA is purified with a common protocol using phenol:chlorofom:isolamyl alcohol (24:24:1) or with specific kits to DNA extraction. The sheared and purified DNA is loaded on agarose gel 1.5 % in TAE and for 30 min at 100 V. After the running, the gel is incubated for 5 min in EtBr solution (40 mg in 100 ml H_2O) and then in pure H_2O for 10 min. The fragment length is checked by using UV lamp. A good sonication should show an important enrichment of fragment between 1000 and 200 bps. If the DNA is not sheared enough repeat the sonication steps.

References

1. Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 10(11):741–754. doi:10.1038/nrm2777
2. Blencowe BJ, Ahmad S, Lee LJ (2009) Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. Genes Dev 23(12):1379–1386. doi:10.1101/gad.1788009
3. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321(5891):956–960. doi:10.1126/science.1160342
4. Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X, Guo J, Dong Z, Liang Y, Bao L, Wang J (2012) Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol 30(3):253–260. doi:10.1038/nbt.2122
5. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11(1):75–87. doi:10.1038/nrg2673
6. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458(7235):223–227. doi:10.1038/nature07672
7. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25(18):1915–1927. doi:10.1101/gad.17446611
8. Thomson AM, Rogers JT, Walker CE, Staton JM, Leedman PJ (1999) Optimized RNA gel-shift and UV cross-linking assays for characterization of cytoplasmic RNA-protein interactions. Biotechniques 27(5):1032–1039, 1042
9. McHugh CA, Russell P, Guttman M (2014) Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol 15(1):203. doi:10.1186/gb4152
10. Gilbert C, Svejstrup JQ (2006) RNA immunoprecipitation for determining RNA-protein associations in vivo. Curr Protoc Mol Biol Chapter 27:Unit 27 24. doi:10.1002/0471142727.mb2704s75
11. Selth LA, Gilbert C, Svejstrup JQ (2009) RNA immunoprecipitation to determine RNA-protein associations in vivo. Cold Spring Harb Protoc 2009(6):pdb prot5234. doi:10.1101/pdb.prot5234
12. Keene JD, Komisarow JM, Friedersdorf MB (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nat Protoc 1(1):302–307. doi:10.1038/nprot.2006.47
13. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40(6):939–953. doi:10.1016/j.molcel.2010.12.011
14. Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S, Ares M Jr, Yeo GW (2012) Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep 1(2):167–178. doi:10.1016/j.celrep.2012.02.001
15. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7):1311–1323. doi:10.1016/j.cell.2007.05.022
16. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322(5902):750–756. doi:10.1126/science.1163045
17. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106(28):11667–11672. doi:10.1073/pnas.0904715106
18. Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol 6(10):e255. doi:10.1371/journal.pbio.0060255
19. Ule J, Jensen K, Mele A, Darnell RB (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods

37(4):376–386. doi:10.1016/j.ymeth.2005.07.018

20. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141(1):129–141. doi:10.1016/j.cell.2010.03.009
21. Niranjanakumari S, Lasda E, Brazas R, Garcia-Blanco MA (2002) Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. Methods 26(2):182–190. doi:10.1016/S1046-2023(02)00021-X
22. Klockenbusch C, O'Hara JE, Kast J (2012) Advancing formaldehyde cross-linking towards quantitative proteomic applications. Anal Bioanal Chem 404(4):1057–1067. doi:10.1007/s00216-012-6065-9
23. Mili S, Steitz JA (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA 10(11):1692–1694. doi:10.1261/rna.7151404
24. Darnell RB (2010) HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdiscip Rev RNA 1(2):266–286. doi:10.1002/wrna.31
25. Davidovich C, Zheng L, Goodrich KJ, Cech TR (2013) Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol 20(11):1250–1257. doi:10.1038/nsmb.2679
26. Brockdorff N (2013) Noncoding RNA and Polycomb recruitment. RNA 19(4):429–442. doi:10.1261/rna.037598.112
27. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341(6147):1237973. doi:10.1126/science.1237973
28. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 142(3):409–419. doi:10.1016/j.cell.2010.06.040
29. Jayaseelan S, Doyle F, Tenenbaum SA (2014) Profiling post-transcriptionally networked mRNA subsets using RIP-Chip and RIP-Seq. Methods 67(1):13–19. doi:10.1016/j.ymeth.2013.11.001
30. Milek M, Wyler E, Landthaler M (2012) Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Semin Cell Dev Biol 23(2):206–212. doi:10.1016/j.semcdb.2011.12.001

Chapter 8

Capture Hybridization Analysis of DNA Targets

Alec N. Sexton, Martin Machyna, and Matthew D. Simon*

Abstract

There are numerous recent cases where chromatin modifying complexes associate with long noncoding RNA (lncRNA), stoking interest in lncRNA genomic localization and associated proteins. Capture Hybridization Analysis of RNA Targets (CHART) uses complementary oligonucleotides to purify an RNA with its associated genomic DNA or proteins from formaldehyde cross-linked chromatin. Deep sequencing of the purified DNA fragments gives a comprehensive profile of the potential lncRNA biological targets in vivo. The combined identification of the genomic localization of RNA and its protein partners can directly inform hypotheses about RNA function, including recruitment of chromatin modifying complexes. Here, we provide a detailed protocol on how to design antisense capture oligos and perform CHART in tissue culture cells.

Key words CHART, lncRNA, lincRNA, RNA, Genomic DNA, Chromatin, High-throughput sequencing

1 Introduction

The recent expansion of annotated noncoding RNAs calls into question what functions this vast pool of RNA has on the genome [1]. Many have been shown or suggested to function in cis or trans on chromatin (e.g., Xist), recruiting or activating chromatin modifying complexes that act on specific genomic loci [2]. Identification of genomic sequence that these RNAs associate with in trans is informative both of mechanism of action and biological function. While proteins have long benefited from specific antibodies that can be used to purify and identify specific associations with both RNA and DNA, similar analysis for RNA has developed more slowly. Recently, tools have been developed to aid in the isolation and identification of targets (including DNA, RNA and proteins) natively associated with specific RNAs with greater stringency and specificity. Here, we describe Capture Hybridization Analysis of RNA Targets (CHART), which uses complementary biotinylated oligonucleotides to specifically hybridize with and purify cross-linked RNAs out of cell extract

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_8,

[3]. The associated biomolecules can provide clues to function. For example, the enriched DNA sequences can illuminate the landscape of specific RNA contacts in trans with genomic DNA.

CHART, followed by several other protocols with common characteristics have emerged using complementary oligonucleotides to purify target RNAs [4–6]. In CHART, formaldehyde cross-linked and sonicated extract is incubated with biotinylated DNA oligonucleotides against an RNA of interest (Fig. 1). The resulting RNA–DNA complexes are bound to streptavidin-conjugated resin, washed under denaturing conditions, and the resulting purified RNA and DNA can be purified and analyzed by qPCR. Enrichment of the

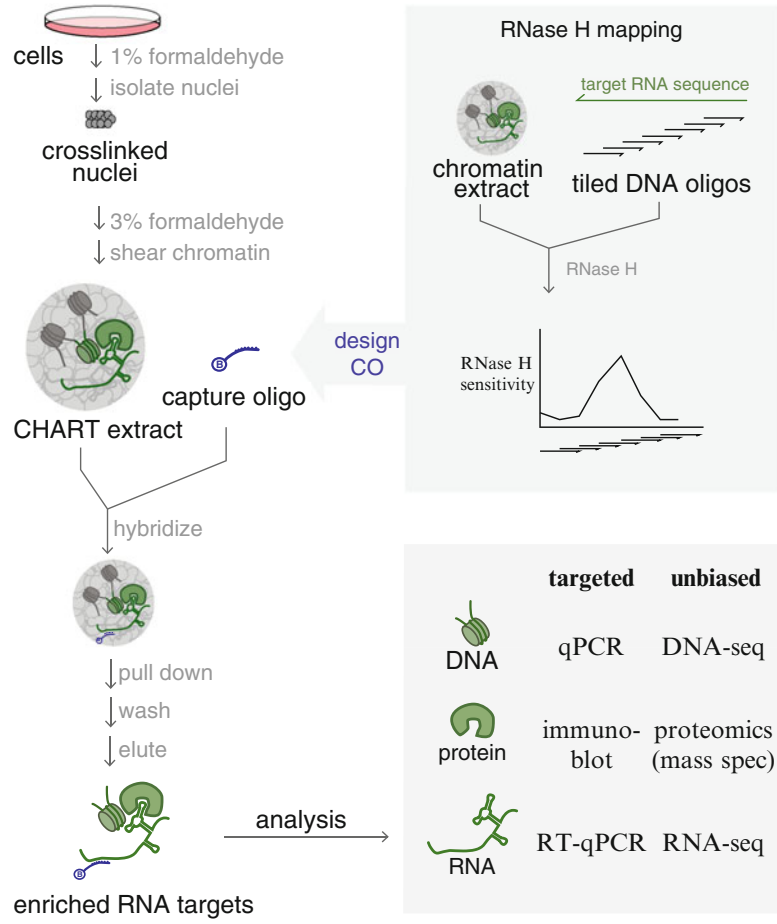


Fig. 1 Schematic of CHART protocol. Determination of accessibility of capture oligonucleotides (CO) is determined by RNase H mapping of DNA oligonucleotides in chromatin extract. Cells are cross-linked with formaldehyde in two steps, incubated with a biotinylated CO cocktail, and immobilized to streptavidin beads before washing. Enriched and eluted RNA targets can be analyzed for associated DNA and protein in directed and unbiased assays

target RNA over scrambled or sense oligo control purifications at specific genomic loci is reflective of biological function. Furthermore, temporal and spatial correlation of genomic localization of RNA with protein localization derived from ChIP datasets can be highly suggestive of biological function.

2 Materials

2.1 Reagents and Equipment

1. 150 × 25 mm tissue culture dishes.
2. DMEM.
3. FBS.
4. Penicillin–streptomycin.
5. PBS.
6. Trypsin, 0.25%.
7. 16% formaldehyde solution, methanol-free, 10 × 10 mL ampules.
8. Liquid N_2.
9. Glass Dounce homogenizer with tight pestle (7 mL).
10. Conical tubes, 15 and 50 mL.
11. Refrigerated centrifuge.
12. End-over-end rotator.
13. Shaking incubator.
14. NanoDrop spectrophotometer.
15. Sonicator.
16. Covaris AFA 12 × 12 mm glass vessels with fiber and cap.
17. Superscript VILO cDNA Synthesis Kit.
18. iTaq UniverSYBR Green.
19. RQ1 DNase.
20. PCR purification kit.
21. RNA purification kit.
22. SUPERase IN RNAse inhibitor 20 U/μL.
23. Spermidine.
24. Spermine.
25. Dynabeads MyOne C1 Streptavidin magnetic beads.
26. Magnetic rack for 1.5 mL microtubes.
27. 37% *N*-lauroylsarcosine sodium salt solution.
28. RNase H.
29. 50× Denhardt's solution.
30. Complete EDTA-free Protease Inhibitor.
31. Proteinase K.

2.2 Solutions and Buffers

1. Sucrose buffer: 0.3 M sucrose, 1% Triton X-100, 10 mM HEPES, 100 mM KOAc, 0.1 mM EGTA, with fresh supplements 0.5 mM spermidine, 0.15 mM spermine, 1 mM DTT, 20 U/mL SUPERase IN, 1× protease inhibitor cocktail.
2. Glycerol buffer: 25% glycerol, 10 mM HEPES, 100 mM KOAc, 0.1 mM EGTA, 1 mM EDTA, with fresh supplements 0.5 mM spermidine, 0.15 mM spermine, 1 mM DTT, 20 U/μL SUPERase IN.
3. Nuclei rinse buffer: 50 mM HEPES pH 7.5, 75 mM NaCl, 0.1 mM EGTA, with freshly added 1 U/mL SUPERase IN, 1 mM DTT, 0.5× Complete EDTA-free Protease Inhibitor.
4. Quenching buffer: 0.125 M EDTA, 250 mM Tris–HCl, pH 7.2, 2.5% SDS, 5 mg/mL proteinase K, this is conveniently made by diluting equal parts standard stocks of all components.
5. Denaturant buffer: 8 M urea, 200 mM NaCl, 100 mM HEPES pH 7.5, 2% SDS.
6. Hybridization buffer: 1.5 M NaCl, 2 M urea, 7.5× Denhardt's buffer, 10 mM EDTA pH 8.0, with fresh supplements 0.5 μL SUPERase IN, 1.2 μL 0.1 M DTT, 5 μL 25× protease inhibitors.
7. Sonication buffer: 50 mM HEPES pH 7.5, 75 mM NaCl, 0.1 mM EGTA, 0.5% *N*-lauroylsarcosine, 0.1% sodium deoxycholate, with fresh supplements 10 U/μL SUPERase IN, 5 mM DTT.
8. Beads buffer: 250 μL denaturant buffer, 100 μL 5 M NaCl, 650 μL dH_2O.
9. WBI: 1 part denaturant buffer, 3 parts WB250.
10. WBII: 1 part denaturant buffer, 1 part hybridization buffer.
11. WB250: 100 mM NaCl, 10 mM HEPES pH 7.5, 2 mM EDTA, 1 mM EGTA, 0.2% SDS, 0.1% *N*-lauroylsarcosine, with 0.1 mM PMSF added fresh.
12. WB12: 10 mM HEPES pH 7.5, 6 M urea.
13. RNase H buffer: 50 mM HEPES pH 7.5, 2 mM $MgCl_2$, 75 mM NaCl, 0.125% *N*-lauroylsarcosine, 0.5% Triton X-100.
14. Reverse cross-link buffer, 5× (RCL): 100 mM Tris–HCl pH 7.5, 5% SDS, 10 mM EDTA.
15. PBST: 1× PBS, 0.1% Triton X-100.

3 Chart Purification

3.1 Preparing Cross-Linked Nuclei

1. Grow eight plates of mammalian cells to 90% confluency (~2×10^7 per plate, 1.6×10^8 cells total; scale up or down as needed) in DMEM supplemented with 10% FBS, penicillin, and streptomycin.

2. Wash cells with PBS.
3. Treat the cells with 0.25% trypsin to release them from the plate.
4. Add the cells to media in 50 mL conical tubes to inactivate the trypsin.
5. Spin the tubes 5 min at 200 × *g* in a swinging rotor tabletop centrifuge.
6. Remove the supernatant and resuspend in 5 mL PBS.
7. Spin cells for 5 min at 200 × *g* using a tabletop centrifuge.
8. Remove the supernatant, resuspend the cells in 20 mL 1% formaldehyde PBS, let sit at RT for 10 min with occasional mixing to keep cells suspended. Cells can also be fixed in solution after trypsinization (*see* **Note 1**).
9. Spin the tube for 5 min at 200 × *g* using a tabletop centrifuge and remove the supernatant. Then resuspend the pellet in 5 mL PBS.
10. Spin the tube for 5 min at 200 × *g* on tabletop centrifuge and remove the supernatant. This sample can be flash-frozen in liquid N_2 and store at −80 °C as cross-linked pellet. Alternatively, the sample can be used immediately for the next step.
11. Resuspend the cross-linked pellet in 7 mL sucrose buffer and transfer the sample to a prechilled Dounce homogenizer.
12. Slowly Dounce 20× with the tight pestle, minimizing bubble formation. Let the mixture sit on ice for 10 min.
13. Pipette 7 mL of glycerol buffer into a fresh 15 mL tube and add the nuclei-containing sucrose buffer very slowly on top of the glycerol buffer layer. Be careful not to mix the two layers.
14. Spin the tube at 1000 × *g* at 4 °C for 10 min to pellet the nuclei.
15. Remove supernatant gently by aspiration, changing tips at the transition between sucrose and glucose buffer so as not to transfer any cytoplasmic extract to the nuclei pellet.
16. **Steps 12–17** may be repeated one more time to further enrich the nuclei. Pelleted nuclei can be used for RNase H mapping in order to identify suitable complementary oligos for purification of the RNA of interest (*see* Subheading 3.3). Otherwise continue with next step to prepare for CHART enrichment.
17. Resuspend the nuclei in 50 mL of 3% formaldehyde in PBST and mix by rotating the tube for 30 min at RT.
18. Spin the tube at 1000 × *g*, 5 min, 4 °C.
19. Discard supernatant and resuspend nuclei pellet in 5 mL ice-cold PBST. Dounce with the tight pestle about five times until the nuclei are resuspended.
20. Spin the resuspended nuclei at 1000 × *g* for 5 min at 4 °C.

21. Repeat **steps 21–22** once.
22. Resuspend the nuclei in ice-cold sonication buffer with fresh supplements. For every 100 μL of nuclei, use 1 mL of sonication buffer. Keep on ice for 5 min.
23. Centrifuge the tube at 400 × *g* for 5 min at 4 °C.
24. Remove the supernatant and resuspend the sample in 0.9 mL ice-cold sonication buffer. The pellet may be frozen in liquid N_2 and stored at −80 °C or used immediately for sonication.

3.2 Fragmenting Chromatin by Sonication

1. Using a Covaris E210, sonicate the sample in a microtube for 8 min, with intensity 175, 20% duty cycle, and 200 bursts/cycle.
2. If the chromatin shear size has been determined for a previous sample with the same cell type and density, proceed to **step 7** in Subheading 3.2. Otherwise empirically determine the optimal sonication time by taking 20 μL from sonication solution mix every 2 min for up to 12 min.
3. Reverse the formaldehyde by following the CHART enrichment protocol, Subheading 3.7, **steps 1–2**.
4. Purify the DNA using a Qiagen PCR purification kit as per the manufacturer's instructions.
5. Analyze the DNA using a 1% agarose gel that does not contain ethidium bromide.
6. Post stain gel with ethidium bromide for 15 min to equally stain the DNA fragments of all sizes. Ideal sonication conditions will produce DNA with a median size between 500 and 3000 bp.
7. Spin the sonicated chromatin at ca. 20,000 × *g* for 20 min at 4 °C. Save the supernatant as CHART extract and use immediately or snap-freeze 40 μL aliquots (each good for one CHART experiment) and store at −80 °C. A NanoDrop analysis of the chromatin extract should be between 300 and 3000 ng/μL DNA.

3.3 Design of Capture Oligonucleotides That Target Accessible Regions of the RNA

This section describes the use of RNase H to determine RNA sites that are available for hybridization in a cross-linked chromatin extract. To accomplish this, DNA oligonucleotides are tested for their ability to hybridize to the RNA, and hybridization can be indirectly observed using RNase H which digests RNA at the site of RNA–DNA hybrids, and analysis of RNA cleavage by RT-qPCR. Candidate DNA sequences should provide maximal coverage of the RNA of interest. For a small RNA, this can be achieved by tiling candidate 20 nt DNA oligomers with 10 bp of overlap each, to effectively achieve 2× coverage of potential capture oligonucleotide target sites. When the RNA of interest is long

enough to make this tiling impractical, the capture oligonucleotide targets should be chosen based on other criteria such as evolutionarily conservation, regions of an RNA that have known protein interactions, and regions that have low repeat density.

1. Resuspend the 1% formaldehyde cross-linked nuclear pellet from **step 16** in Subheading 3.1 in 500 μL of ice-cold nuclei wash buffer.
2. Centrifuge the tube at 1000 × *g* for 10 min at 4 °C. Discard the supernatant.
3. Repeat **steps 1–2** in Subheading 3.3 one additional time.
4. Resuspend the pellet in 300 μL of sonication buffer. Then centrifuge at 1000 × *g* for 10 min at 4 °C and discard the supernatant.
5. Resuspend the nuclear pellet in sonication buffer (add ~150 μL) to a final volume of 300 μL.
6. Sonicate the sample using a Covaris E210 instrument as described in Subheading 3.2, **step 1**.
7. Spin the sonicated sample for 10 min, at 16,100 × *g* at RT.
8. This extract can be used immediately or flash-frozen in liquid N_2 and stored at −80 °C.
9. For each DNA oligo to be tested, make the following master mix: 10 μL cleared extract, 0.03 μL 1 M $MgCl_2$, 0.1 μL 1 M DTT, 1 μL 5 U/μL RNase H, 0.5 μL 20 U/μL SUPERasIN. Account for two additional control reactions where dH_2O is added in place of the DNA oligo.
10. Aliquot 10 μL of master mix into PCR tubes (eight-strips are convenient) and add 1 μL of 100 μM stock of a candidate DNA oligomer per tube (every tube includes a different DNA sequence), with only water added to two tubes for controls. Mix by inversion and spin down the tubes to consolidate the liquid in the bottom of the tube.
11. Incubate the PCR strips in a thermal cycler for 30 min at 30 °C.
12. Spin these tubes briefly to collect any condensate and to each reaction add 1 μL RQ1 DNase and 0.1 μL 60 mM $CaCl_2$ (can dilute 6 μL of 1 M stock into 94 μL dH_2O).
13. Incubate these reactions in a thermal cycler for 10 min at 30 °C.
14. Spin the tubes and add 2 μL quenching buffer (it is convenient to add quenching buffer to the lids), mix the tubes by flicking, and spin down the reaction.
15. Incubate the PCR tubes in thermal cycler 60 min, 55 °C, then 30 min, 65 °C.

16. Purify the RNA from these reactions using an RNA purification kit (e.g., RNeasy, Qiagen).
17. The RNA concentration can be determined here by measuring on a NanoDrop. Successful purification will usually yield ~100–200 ng/μL and all samples should contain similar amounts of RNA.
18. Set up reverse transcription reaction (RT) with 1 μL 5× VILO master mix, 7 μL purified RNA from **step 17**, 0.5 μL VILO RT enzyme. Save one tube for a no-enzyme control.
19. Incubate the reaction in a thermocycler for 10 min at 25 °C; 60 min at 42 °C; 5 min at 85 °C; then hold the reaction at 4 °C.
20. Dilute each reverse transcription reaction with 90 μL dH_2O.
21. Prepare iTaq UniverSYBR Green reactions with 10 μL Supermix; 5 μL primer mix (6 pmol of each primer, final); 5 μL RT reaction.
22. Perform qPCR with the following cycles:

1 cycle:	5 min	94 °C
40 cycles:	30 s	94 °C
	30 s	52 °C
	1 min	72 °C

23. Analyze qPCR results with the following formula: sensitivity = $(C_{T,\ oligo} - C_{T,\ no\ oligo}$ Target primers$) / (C_{T,\ oligo} - C_{T,\ no\ oligo}$ Control primers).
24. Use the RNase H sensitivity profiles to determine which regions are available for hybridization. Using these data, determine capture oligonucleotide sequences to use for biotinylated oligo synthesis guided by the following criteria: (1) focus on regions where two or more DNA oligos cause RNase H sensitivity; (2) sequences should have high specificity in the genome as determined by BLAST; (3) melting temperature of oligos should ideally be between 58 and 65 °C. Generally successful capture oligonucleotides are between 21 and 26 nt. Note that some RNAs have long stretches of highly accessible RNA (RNase H sensitivity >10). It is convenient to order chemically biotinylated oligonucleotides from commercial suppliers (e.g., IDT). A convenient working dilution of biotinylated capture oligonucleotides is 300 pmol/μL (300 μM).

3.4 Performing CHART Enrichment

1. If starting this protocol from frozen aliquots of CHART extract (from Subheading 3.2, **step 7**), quickly thaw the aliquots and remove insoluble material by spinning at >12,000 × *g* for 20 min at 4 °C. Proceed using the cleared extract, taking care to avoid resuspending any solid material.

2. Add 80 μL hybridization buffer to the cleared extract, along with 0.5 μL SUPERasIN, 1.2 μL 0.1 M DTT, 5 μL 25× protease inhibitors, mix. Let sit at RT for 10 min.
3. Save 5 μL from **step 2** in Subheading 3.4 as input. To the remaining material, add 3 μL capture oligonucleotide mix (20 μM total for all DNA oligos combined) to each sample.
4. Rotate the hybridization reactions overnight at RT. It is sometimes helpful to Parafilm the caps to ensure that the mixture does not leak.
5. Remove insoluble material by spinning the tubes at >12,000 × *g* for 20 min at RT. Transfer the supernatant to new tube.
6. During the previous step (Subheading 3.4, **step 5**), prepare MyOne streptavidin C1 beads by taking 60 μL beads per reaction. Capture the beads on a magnetic stand. Wash the beads 1× with ddH_2O (750 μL). Between rinses, capture beads in magnetic rack and remove supernatant. Resuspend the C1 beads in 40 μL per tube of beads buffer, and add 40 μL beads to each cleared hybridization reaction from **step 5**.
7. Incubate the beads with the extract at RT for 1–3 h with rotation (here, can also prepare denaturant buffer; leave in bath sonicator to solubilize the urea or prepare the day before with rotation at room temperature).
8. After incubation, add beads and extract to 0.9 mL WBI in a separate tube. Capture beads and aspirate supernatant. For this step and for all future washes, resusupend the beads completely by inversion and let the tube sit for 1 min. Then, spin the tube quickly to remove buffer from the cap and capture the beads on magnetic rack for 1–2 min before aspiration. Do not to let the beads dry out in between rinses.
9. Wash the beads one time with 1 mL WBI.
10. Wash the beads one time with 1 mL WBII.
11. Wash the beads two times with 1 mL WB250.
12. Wash the beads one time with 1 mL WB12.
13. Capture the beads and proceed immediately to either RNase H elution (Subheading 3.5) or proteinase K elution (Subheading 3.6).

3.5 Elution with RNase H

1. Wash the beads one time with 1 mL of RNase H buffer.
2. Add 100 μL RNase H buffer to beads, then add 3 μL RNase H and resuspend by pipetting up and down several times gently, trying not to introduce bubbles into the solution.
3. Incubate the tubes at 25 °C for 10 min with gentle shaking.
4. Capture the beads and save the supernatant as the RNase H eluate.

5. For nucleic acid analyses, add 20 μL of 5× RCL buffer and 5 μL of 20 mg/mL proteinase K to the eluant and proceed to cross-link reversal (Subheading 3.7).

3.6 Elution with Proteinase K

1. As an alternative to RNase H elution, the bound material can be eluted with proteinase K. This is useful because the yields of recovery are higher, although the elution is less specific. Add 100 μL 1× RCL buffer to beads from **step 12** in Subheading 3.4.
2. Add 5 μL of 20 mg/mL proteinase K and mix by pipetting up and down gently, trying not to introduce bubbles into the solution.
3. Incubate at 55 °C for 1 h with shaking.
4. Capture the beads and transfer the supernatant to a fresh tube.

3.7 Cross-Link Reversal

1. Dilute the input sample to 100 μL with RNase H buffer. Add 5 μL of 20 mg/mL proteinase K and mix by pipetting up and down gently, trying not to introduce bubbles into the solution.
2. Incubate the CHART enriched material (from either Subheading 3.4 or 3.5) and the input sample at 55 °C for 1 h, then at 65 °C for 1 h.
3. From here, the CHART enriched material can be analyzed in numerous ways. For analysis of genomic DNA, purify the DNA using a Qiagen PCR purification kit. For RNA, purify the sample with a Qiagen RNeasy Mini Kit. During optimization it is often helpful to split the samples and analyze both the DNA and the RNA (*see* **Note 2**).
4. If preparing a DNA library for deep sequencing, treat CHART-enriched DNA with RNase cocktail and further shear eluted DNA below 500 base pairs (Covaris E210 sonicator, 5% duty, 200 bursts/cycle, intensity 5, 4 min total). Then follow conventional ChIP library preparation methods.

4 Notes

1. Note that first fixation step can alternatively be performed directly on plates without trypsinization, and with subsequent removal of cells with cell scraper. To do this, wash cells on plate with PBS, then remove PBS and add 20 mL 1% formaldehyde in PBS, let sit at RT for 10 min, add 2 mL 0.125 M glycine, then remove 1% formaldehyde media, wash plate once with PBS, then add 8 mL PBS with 0.1 mM PMSF and scrape cells off plate before proceeding to **step 10**.
2. It is expected that the RNase H eluate will have lower, but cleaner (higher signal to background) yields of target RNA-associated genomic loci. When just beginning, it is recommended to take both eluates.

References

1. Ulitsky I, Bartel DP (2013) lincRNAs: genomics, evolution, and mechanisms. Cell 154:26–46. doi:10.1016/j.cell.2013.06.020
2. Brockdorff N (2013) Noncoding RNA and polycomb recruitment. RNA 19:429–442. doi:10.1261/rna.037598.112
3. Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT (2013) High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature 504:465–469. doi:10.1038/nature12719
4. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE (2011) The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A 108:20497–20502. doi:10.1073/pnas.1113536108
5. McHugh CA, Chen C-K, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521:232–236. doi:10.1038/nature14443
6. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44:667–678. doi:10.1016/j.molcel.2011.08.027

Chapter 9

Identification of RNA–Protein Interactions Through In Vitro RNA Pull-Down Assays

Claire Barnes and Aditi Kanhere*

Abstract

Recent advances in next-generation sequencing have revealed that majority of the human genome is transcribed into long and short RNA (ncRNA) transcripts. Many ncRNAs function by interacting with proteins and forming regulatory complexes. RNA–protein interactions are vital in controlling core cellular processes like transcription and translation. Therefore identifying proteins that interact with ncRNAs is central to deciphering ncRNA functions. Here we describe an RNA–protein pull-down assay, which enables the identification of proteins that interact with an RNA under study. As an example we describe pull-down of proteins interacting with ncRNA XIST, which assists in the recruitment of the polycomb-repressive complex-2 (PRC2) and drives X-chromosomal inactivation.

Key words Noncoding RNA, RNA–protein interactions, Biotin, In vitro transcription, XIST, Pull-down

1 Introduction

Majority of the human genome is transcribed into ncRNA transcripts but functions of only few ncRNAs have been characterized [1]. These few examples support the idea that many ncRNAs function through their interactions with protein partners [2–4]. Therefore, identification of novel RNA–protein interactions could be highly important in determining unknown functions of ncRNAs. Majority of molecular biology tools are protein centric which is one of the reasons behind our lack of understanding of functional roles of many ncRNAs. Fewer RNA-centric methods are available for biologists that allow identification of the whole repertoire of RNA-interacting proteins. One such method is RNA–protein pull-down assay where one or multiple proteins binding to an RNA can be identified.

This method (Fig. 1) utilizes in vitro transcription to produce a RNA or a part of the RNA of interest which is then end-labeled using biotin and then proteins interacting with it are pulled down from whole-cell or protein extracts using streptavidin beads. These

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_9,

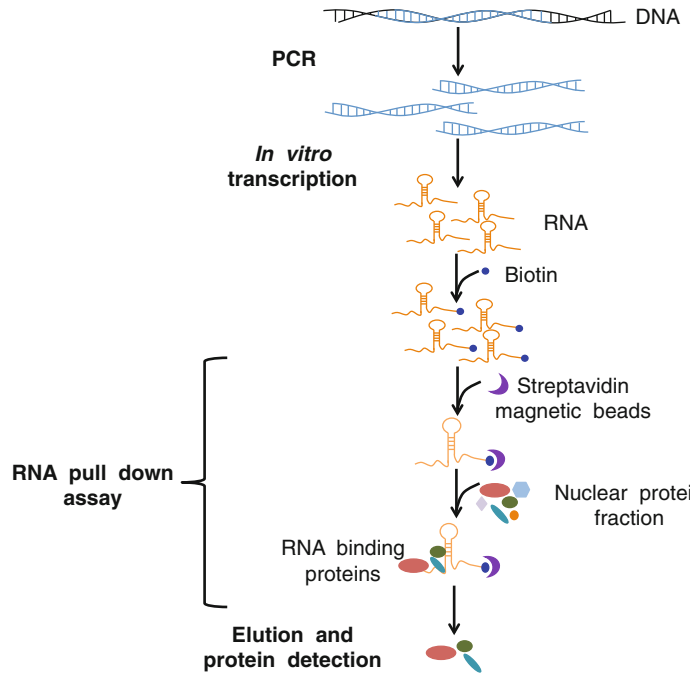


Fig. 1 Schematic representation of in vitro transcription and RNA–protein pull-down method. The DNA region of interest (shown in *blue*) was amplified, in vitro transcription performed to produce RNA for biotin labeling. Biotin-labeled RNA was captured by streptavidin magnetic beads, incubated with nuclear fraction extract to allow RNA–protein interactions to form, and the RNA-bound protein was eluted for detection by western blotting, silver staining, or mass spectrometry

proteins can be then identified using mass spectrometry or western blot. In the first step a DNA template corresponding to the RNA sequence is produced with an added 5′ T7 RNA polymerase promoter sequence (5′-TAATACGACTCACTATA-3′) [5]. The T7 RNA polymerase promoter sequence is introduced in the DNA template using a polymerase chain reaction (PCR) step, which uses specially designed primers as shown in Fig. 2. The T7 promoter is used to in vitro transcribe and produce copies of RNA. The biggest advantage of using in vitro transcription is that it allows synthesis of different parts and different length of RNAs with specified sequence.

We used this pull-down assay to identify proteins, which interact with 672 bp transcript corresponding to 5′ end of XIST ncRNA (XIST-5′). This region contains the RepA repeats, which have been previously shown to associate with polycomb proteins such as EZH2 and aid in its role in X-chromosomal interaction [2, 6]. Recently this RNA was shown to bind a number of other proteins which probably assist in the recruitment of polycomb proteins or to tether XIST to X-chromosome during its inactivation [7–9]. The proteins identified in our in vitro pull-down assay showed good overlap with previously identified XIST-interacting proteins in in vivo assay [7].

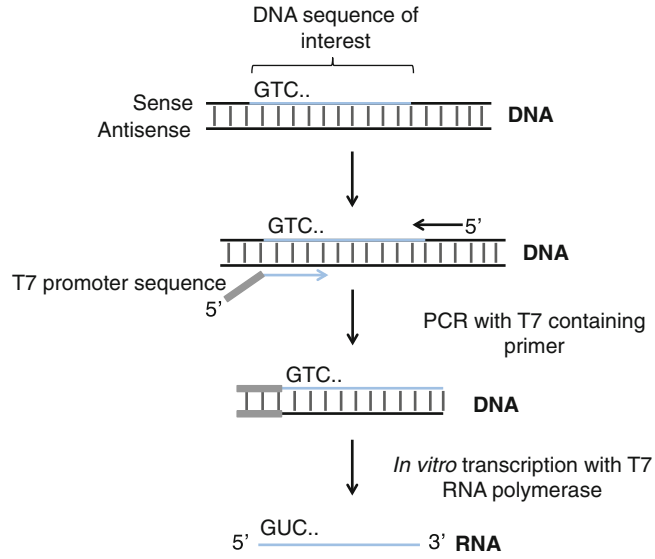


Fig. 2 Production of the DNA template with T7 promoter sequence. The T7 promoter sequence is introduced through its addition to the 5′ end of the primer. PCR introduces the T7 promoter sequence 5′ of the DNA region of interest (shown in *blue*); this DNA is then in vitro transcribed with T7 RNA polymerase to produce RNA. The RNA sequence is produced from the first G after T7 promoter

2 Materials

Maintain RNase-free environment and keratin-free environment (*see* **Note 1**). All solutions should be made in DEPC-treated RNase- and DNase-free water.

2.1 General Reagents

1. RNase-free H_2O.
2. 100% Ethanol.
3. Isopropanol.
4. Phenol-chloroform-isoamyl alcohol (25:24:1).

2.2 DNA Template Preparation Components

1. Genomic or cDNA (100 ng/μL) from K562 cells.
2. 5 μM Stock of primers for PCR: Primers for XIST-5′ (forward: 5′-GAAAT**TAATACGACTCACTATA**GGAATCATTTTTGGTTGACATCT-3′, reverse: 5′-CGAGTTATGCGGCAAGTCTA-3′) were ordered from Sigma. But other oligonucleotide synthesizing companies can also be used (for primer design *see* Fig. 2; *see* **Note 2**).
3. MyTaq Redmix (Bioline) or a suitable alternative ready-to-use PCR mix.
4. SYBR Safe DNA gel stain (Invitrogen) or a suitable alternative.

5. TBE buffer: For 1 L of a 10× stock, mix 108 g Tris base, 55 g boric acid, 80 mL 0.5 M EDTA, and 700 mL of dH_2O. Adjust pH to 8.0 and bring to 1 L volume with dH_2O.
6. Agarose gel: For 100 mL 1% gel, 1 g agarose in 100 mL 1× TBE buffer.
7. DNA ladders (1 Kb and 100 bp).
8. QIAquick gel extraction kit (Qiagen) or a suitable alternative.

2.3 T7 In Vitro Transcription Components

1. T7 RNA polymerase (Thermo Fisher Scientific).
2. 5× Transcription buffer supplied with T7 RNA polymerase (Thermo Fisher Scientific).
3. dNTP mix (10 mM solution).
4. RNase inhibitor (for example, RiboLock RNase Inhibitor from Thermo Fisher Scientific).
5. DNase I (AMPD1; Sigma Aldrich).
6. 10× DNase I reaction buffer (supplied with DNase I).
7. DNase I stop solution (supplied with DNase I).
8. RNA 3′ End Desthiobiotinylation Kit (Thermo Fisher Scientific) supplied with magnetic RNA–protein pull-down kit.

2.4 Protein Extraction Components

1. Approximately 5×10^7 K562 cells.
2. RLN buffer: 50 mM Tris–Cl (pH 8.0), 1.5 mM $MgCl_2$, 140 mM NaCl, 0.5% Nonident P-40 (store at 4 °C).
3. Protein lysis buffer: 50 mM Tri–HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 5 mM EGTA, 1% NP-40, protease inhibitor (store at 4 °C).
4. Phosphate-buffered saline, pH 7.0 (store at 4 °C).

2.5 RNA–Protein Pull-Down Components

1. Pierce magnetic RNA–protein pull-down kit (Catalog number: 20164, Thermo Fisher Scientific).
2. 3′ Untranslated-region androgen receptor (AR) RNA as a positive control (Thermo Fisher Scientific, supplied with the RNA pull-down kit).
3. Poly$(A)_{25}$ RNA as a negative control (Thermo Fisher Scientific, supplied with RNA pull-down kit).
4. 0.1 M NaOH, 50 mM NaCl.
5. 100 mM NaCl.
6. 5× Loading buffer: 0.25% (w/v) bromophenol blue, 0.5 M dithiothreitol, 50% (w/v) glycerol, 10% (w/v) SDS, 0.25 M pH 6.8 Tris–Cl.

2.6 Polyacrylamide Gel Electrophoresis and Western Blot Components

1. PageRuler Protein Ladder (or a suitable alternative).
2. PVDF membrane (0.45 μm).
3. Filter paper.
4. 10× SDS running buffer: 0.25 M Tris, 1.92 M glycine.
5. 1× SDS running buffer: 100 mL 10× Polyacrylamide running buffer, 10 mL 10% (w/v) SDS and bring to 1 L volume with dH_2O.
6. 2× Transfer buffer: Mix 5.82 g Tris and 2.93 g Glycine per 500 mL dH_2O.
7. 1× Transfer buffer: 250 mL 2× Transfer buffer, 5 mL 10% (w/v) SDS, 25 mL methanol and bring to 500 mL volume with dH_2O.
8. TBS-T solution: 40 mL 1 M Tris, 60 mL 5 M NaCl, 1 mL Tween-20, bring to 2 L with dH_2O.
9. TBS: 15 mL 5 M NaCl, 10 mL 1 M Tris (pH 7.5), bring to 500 mL with dH_2O.
10. 5% (w/v) milk blocking solution: Dissolve 2.5 g milk powder in 50 mL TBS.
11. 4% Stacking gel: 425 μL 30% Polyacrylamide, 625 μL 0.5 M Tris (pH 6.8), 25 μL 10% (w/v) SDS, 1.34 mL dH_2O, 25 μL 10% (w/v) APS, 5 μL TEMED.
12. 10% Polyacrylamide gel (1.5 mm) with 4% stacking gel: 3.33 mL 30% Polyacrylamide, 3 mL 1.5 M Tris (pH 8.8), 125 μL 10% (w/v) SDS, 4.96 mL dH_2O, 125 μL 10% (w/v) APS, 10 μL TEMED.
13. 12.5% Polyacrylamide gel (1.5 mm) with 4% stacking gel: 4.17 mL 30% Polyacrylamide, 3.13 mL 1.5 M Tris (pH 8.8), 125 μL 10% (w/v) SDS, 4.96 mL dH_2O, 125 μL 10% (w/v) APS, 10 μL TEMED.
14. Odyssey blocking buffer (LI-COR)
15. Primary and secondary antibodies: For example, HuR monoclonal mouse antibody (supplied with RNA–protein pull-down kit Thermo Fisher Scientific).

2.7 Silver Staining Components

1. ProteoSilver stain kit (Sigma Aldrich or a suitable alternative).
2. Ultrapure H_2O.
3. Fixing solution: 25 mL Ethanol, 5 mL acetic acid, 20 mL ultrapure H_2O.
4. 30% Ethanol: 15 mL 100% ethanol, 35 mL ultrapure H_2O.
5. Sensitizer solution: 250 μL sensitizer solution (Sigma Aldrich), 24.75 mL ultrapure H_2O.

6. Silver solution: 250 μL Silver solution (Sigma Aldrich), 24.75 mL ultrapure H_2O.
7. Developer solution: 1.25 mL Developer 1 solution (Sigma Aldrich), 50 μL developer 2 solution (Sigma Aldrich), 23.5 mL ultrapure H_2O.

2.8 Equipment

1. RNase free: Eppendorfs, pipette tips, PCR tubes, Falcon tubes.
2. Needle syringes.
3. Thermocycler.
4. Horizontal and vertical gel electrophoresis equipments.
5. UV transilluminator.
6. NanoDrop spectrophotometer.
7. Centrifuge able to cool to 4 °C.
8. Water or dry bath capable of 16–85 °C.
9. Water bath sonicator.
10. Magnetic stand or a magnet.
11. Rotator for 50 mL Falcon tubes.
12. Plastic or glass tray or container.
13. Odyssey scanner or imager to scan western blots.

3 Methods

All methods should be carried out under RNase-free conditions (*see* **Note 1**).

3.1 DNA Template Preparation

1. Order the forward and reverse deoxyribo-oligonucleotide primers for PCR from companies such as Sigma. The primer with same sequence as RNA strand should include T7 RNA polymerase promoter sequence (5′-TAATACGACTCACTATAG-3′) at its 5′ end enabling its incorporation at the 5′ of the RNA sequence of interest. As shown in Fig. 2 underlined G is the first nucleotide incorporated in transcribed RNA. T7 RNA polymerase has increased efficiency when the T7 promoter sequence is preceded by several nucleotides and immediately followed by GG (*see* **Note 2**).
2. Prepare 2× 50 μL PCR reactions for each primer pair as follows: 4 μL each of forward and reverse primers (5 μM stock), 25 μL of MyTaq Redmix or other suitable PCR mix, RNase-free H_2O (to make a final volume of 50 μL), 100 ng of cDNA or genomic DNA (*see* **Note 3**).
3. Amplify he DNA through PCR as follows (or in case of using alternative PCR mix, according to the vendor's instructions):

(a) Denaturation: 94 °C, 1 min.

(b) Amplification (45 cycles): to 94 °C, 15 s; annealing temperature (*see* **Note 4**), 15 s; 72 °C, 40 s.

(c) Final extension and storage: 72 °C, 5 min; 4 °C ∞.

4. Check the size and purity of the PCR product on a 1% agarose gel containing 0.3 μg/mL SYBR safe DNA gel stain (load 5–10 μL of PCR product). Use a standard UV illuminator to view the gel. If only one band of desired size is observed then proceed to Subheading 3.2.
5. If multiple products are observed in **step 4**, you will require purification from a gel (*see* **Note 5**); run remaining PCR product on an agarose gel and cut the desired band with a clean scalpel (use suitable UV protection).

3.2 PCR Product Cleanup Using Qiagen QIAquick Gel Extraction Kit

1. Add 3 volumes of QG buffer (Qiagen) and 1 volume of isopropanol and mix thoroughly using a vortex mixer; for example, if you have a 50 μL reaction, add 150 μL of QG buffer and 50 μL of isopropanol.
2. Load 750 μL of above mix onto a spin column provided with the kit and centrifuge at 17,900 × *g* for 1 min. Discard the flow-through, which is collected in the bottom tube.
3. Add 500 μL of QG buffer to the column and centrifuge at 17,900 × *g* for 1 min. Discard the flow-through, which is collected in the bottom tube.
4. Add 750 μL PE buffer (Qiagen) to the column and centrifuge at 17,900 × *g* for 1 min. Discard the flow-through, which is collected in the bottom tube.
5. Centrifuge the spin column (empty) at 17,900 × *g* for 1 min. This is essential to get a good-quality DNA sample.
6. Place spin column in a clean Eppendorf and elute DNA by adding 30 μL of RNase-free H_2O (37 °C) to column and centrifuge at 17,900 × *g* for 1 min.
7. Determine the DNA concentration using Nanodrop or an alternative spectrophotometric method (*see* **Note 6**).

3.3 Ethanol Precipitation (See Note 7)

1. Ethanol precipitate the nucleic acids by adding 0.1 volumes of 3 M sodium acetate (pH 5.2), 2.2 volumes of 100% ethanol (ice cold), and 1 μL glycogen, mix thoroughly, and store at −20 °C overnight (*see* **Note 8**).
2. On the next day vortex samples and centrifuge at 4 °C, 10,600 × *g*, for 20 min. Keeping an eye on the pellet, carefully discard the supernatant.
3. Add 500 μL 70% ethanol to the pellet and vortex.

4. Centrifuge at 4 °C, 10,600 × *g*, for 20 min and carefully discard the supernatant.
5. Centrifuge at 4 °C, 10,600 × *g*, for 1 min and discard any remaining supernatant.
6. Allow to air-dry for 5 min prior to resuspension in 30 μL RNase-free H_2O (at 37 °C).
7. Determine the DNA concentration using Nanodrop or an alternative spectrophotometric method.

3.4 T7 RNA Polymerase In Vitro Transcription

1. Set up a 50 μL T7 polymerase reaction by incubating 2 μL of T7 RNA polymerase, 10 μL of 5× reaction buffer (supplied with T7 RNA polymerase by the manufacturer), 10 μL of 10 mM NTP mix, 1 μg of purified DNA (containing T7 promoter sequence), and RNase-free H_2O up to 50 μL (*see* **Note 9**).
2. Incubate the reaction for up to 72 h at 4–37 °C (*see* **Note 10**).
3. DNase I treatment of samples: Incubate 48 μL of T7 RNA polymerase reaction mix with 6 μL DNase I and 6 μL of 10× reaction buffer (supplied by the manufacturer) at 25 °C for 15 min, add 6 μL of stop solution (manufacturer supplied), and heat at 70 °C for 10 min before immediately chilling on ice.
4. Check for completeness of DNA digestion using agarose gel electrophoresis: run 1 μL of DNase I-treated and untreated samples on a 1% agarose containing 0.3 μg/mL SYBR safe DNA gel stain. If successful, after DNase I treatment band corresponding to DNA should disappear.

3.5 (Optional) Phenol-Chloroform RNA Extraction (See Note 11)

1. Add 140 μL of RNase-free H_2O to the ~60 μL DNase I-treated RNA and mix well.
2. Add 200 μL of phenol-chloroform-isoamyl alcohol (mix well before adding to ensure that the two phases have not separated).
3. Vortex vigorously and immediately centrifuge at 10,600 × *g*, 25 °C, for 10 min.
4. Carefully remove the aqueous, RNA-containing phase (top layer) and place in a clean Eppendorf (*see* **Note 12**).
5. Ethanol precipitate the RNA following the protocol described in Subheading 3.3; except in the last step use 6–16 μL RNase-free H_2O for resuspension (*see* **Note 13**).
6. Determine the RNA concentration using Nanodrop or an alternative spectrophotometric method.

3.6 Biotin Labeling Using 3′ End Desthiobiotinylation Kit

1. Carry out 3′ end labeling of RNA with the 3′ end desthiobiotinylation labeling kit from Thermo Fisher Scientific (or a suitable alternative) as described below.
2. Incubate 50–100 picomoles RNA with 25% DMSO at 85 °C for 5 min, and immediately chill on ice (*see* **Note 14**).

3. Prepare the biotin labeling reaction mixture: 6 μL of 10× RNA ligase buffer, 2 μL RNase inhibitor, 50–100 pmol of RNA, 2 μL biotinylated cytidine bisphosphate, 4 μL T4 RNA ligase, and 30 μL PEG (30%) at 16 °C overnight (*see* **Note 15**).
4. Ethanol precipitate the RNA as described in Subheading 3.3 (*see* **Note 16**), and ensure to resuspend in 20 μL of RNase-free H_2O (37 °C).
5. Determine the RNA concentration using Nanodrop or an alternative spectrophotometric method.

3.7 Nuclear Protein Extraction

1. Using ~5×10^7 K562 cells.
2. Centrifuge at 240 × *g*, 4 °C, for 5 min and remove media.
3. Wash the cells twice with 3 mL PBS.
4. Remove all of the supernatant and resuspend in 1 mL of cold (4 °C) RLN buffer.
5. Incubate on ice for 5 min before splitting the sample across two Eppendorfs.
6. Centrifuge at 1450 × *g*, 4 °C, for 2 min and discard supernatant.
7. Wash with 100 μL cold RLN buffer (*see* **Note 17**).
8. Resuspend each pellet in 50 μL of cold protein lysis buffer (*see* **Note 18**).
9. Incubate on ice for 30 min.
10. Centrifuge at full speed, 4 °C, for 15 min.
11. Remove and keep supernatant.
12. Perform a Bradford assay or alternative spectrophotometric method to determine the protein concentration in the sample.

3.8 RNA–Protein Pull-Down Assay

For all steps involving the removal of supernatant, ensure to place a magnet against the side of the tube to collect the beads before removing the supernatant. For all wash steps resuspend beads by pipetting.

1. Mix the beads well prior to aliquoting 50 μL of beads per reaction.
2. Remove supernatant and wash beads twice with 100 μL of 0.1 M NaOH and 50 mM NaCl.
3. Wash once with 100 μL 100 mM NaCl.
4. Wash twice with 50 μL of 20 mM Tris (pH 7.5).
5. Add 50 μL RNA capture buffer and resuspend beads.
6. Add 50 pmol of labeled RNA and mix gently by pipetting.
7. Incubate at 25 °C for 30 min with agitation (*see* **Note 19**).

8. Remove supernatant and wash twice with 50 µL of 20 mM Tris (pH 7.5).
9. Wash once with 100 µL of 1× protein–RNA binding buffer (dilute 10× in RNase-free H_2O).
10. Prepare protein master mix on ice (100 µL per pull-down assay): 10 µL of 10x protein–RNA binding buffer, 30 µL of 50% glycerol, 100 µg of protein lysate, and RNase-free H_2O to bring to 100 µL volume.
11. Add 100 µL of protein master mix and incubate at 4 °C for 60 min with agitation.
12. Remove supernatant. Keep this at −20 °C for later analysis.
13. Wash twice with 100 µL of 1× wash buffer supplied. If desired save supernatants at −20 °C for later analysis.
14. Add 25 µL of elution buffer and vortex well (*see* **Note 20**).
15. Add 6.25 µL of 5× loading buffer, mix well, and heat to 70 °C for 10 min (*see* **Note 21**).
16. Keep cold on ice if proceeding to next step immediately. For later use store at −20 °C.

3.9 Western Blot

Used for the detection of proteins that are in high abundance in the pulled-down sample and are known candidates binding to the RNA in question.

1. Prepare a 1.5 mm 12.5% polyacrylamide gel with a 4% stacking gel.
2. Load protein ladder and 30 µL of each protein sample from RNA pull-down assay (positive and negative controls) onto 12.5% gel.
3. Run the gel at 100 V for approximately 1 h or until the blue front reaches the end of the gel. Remove the gel from the plates and put in a clean plastic container.
4. Soak gel in 1× transfer buffer for 10 min prior to performing semidry transfer with a PVDF membrane (*see* **Note 22**), by running at 22 V for 16 min (*see* **Note 23**).
5. Block membrane at room temperature for 1 h with 20–25 µL of Odyssey blocking solution used for HuR (*see* **Note 24**).
6. Wash membrane briefly with TBS-T.
7. Incubate overnight at 4 °C on a shaker, with the primary antibody diluted 1:1000 for HuR, in Odyssey blocking buffer.
8. Wash the membrane thrice with TBS-T (5 min).
9. Incubate for 1 h at room temperature with secondary antibody diluted (1:10,000–1:15,000) in 5% (w/v) milk blocking solution.
10. Wash the membrane twice with TBS-T (5 min).
11. Wash the membrane once with PBS (5 min).

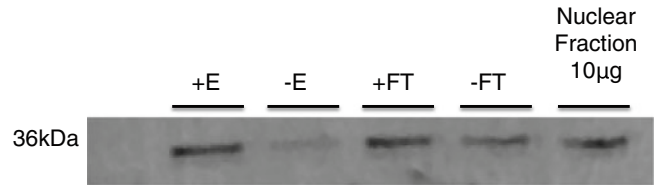


Fig. 3 HuR protein is pulled down using HuR-binding AR RNA. Western blot showing the results of the RNA–protein pull-down with AR RNA (+E) and a negative control Poly(A)$_{25}$ RNA (–E). HuR interacts strongly with AR RNA and weakly with Poly(A)$_{25}$ RNA. Unbound HuR protein in flow through from the assay with AR RNA (+FT) as well as Poly(A)$_{25}$ RNA (–FT) are also shown

12. Visualize using Odyssey Imager to determine if the control RNA transcripts have been pulled-down HuR protein in the assay (*see* Fig. 3 for the resulting western blot from our pull-down assay).

3.10 Silver Staining Using Sigma Aldrich ProteoSilver Silver Stain Kit

Used for the detection of proteins that are in low abundance in the pull-down elute.

1. Prepare a 1.5 mm 10% polyacrylamide gel with a 4% stacking gel.
2. Load 30 µL of pull-down protein samples and run at 150 v for ~60 min.
3. Ensure that the plastic or glass container is thoroughly cleaned and rinse with ultrapure H_2O.
4. Place membrane in clean container with 50 mL of fixing solution overnight (*see* **Note 25**).
5. Discard fixing solution and incubate with 50 mL 30% ethanol for 10 min.
6. Wash with 25 mL ultrapure H_2O for 10 min.
7. Incubate with 25 mL sensitizer solution for 10 min.
8. Wash twice with 25 mL ultrapure H_2O for 10 min.
9. Incubate with 25 mL silver solution for 10 min.
10. Wash with 25 mL ultrapure H_2O for 1 min.
11. Incubate with 25 mL developer solution and incubate for 3–10 min (*see* **Note 26**).
12. Add 1.25 mL stop solution (Sigma Aldrich) and incubate for 5 min (for the results from our silver staining *see* Fig. 4).
13. Wash with ultrapure H_2O for 15 min.
14. Store in fresh ultrapure H_2O.

For mass-spec analysis, cut the bands of interest using a brand-new scalpel and collect in a fresh Eppendorf with a clean forceps (*see* **Note 27**). This can be given for mass-spec identification to the local facility.

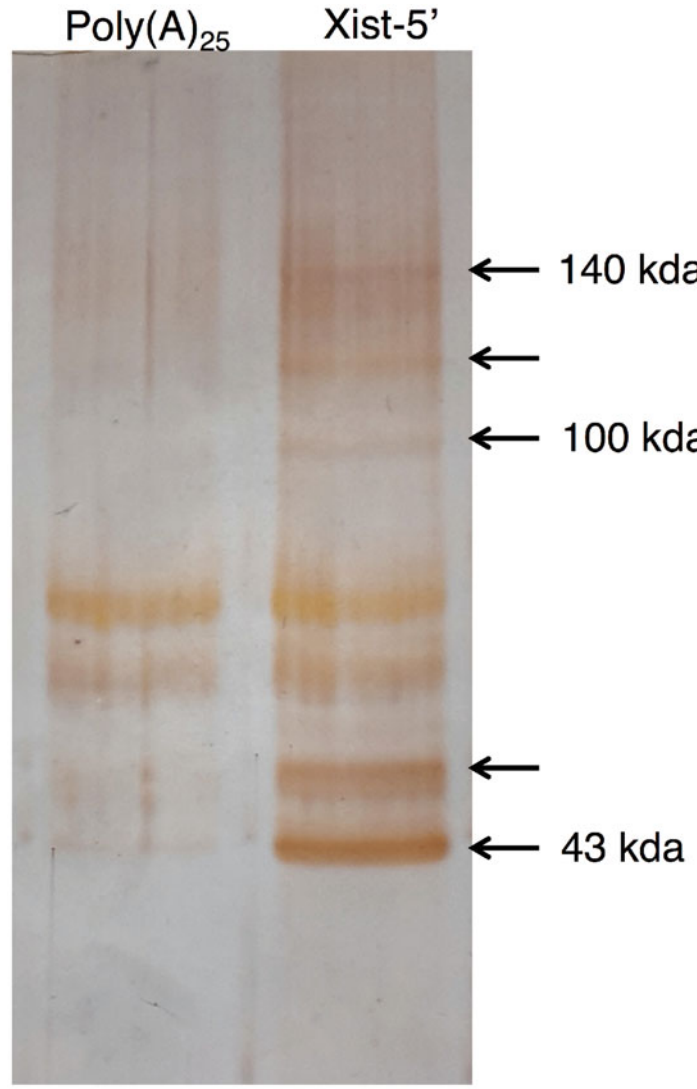


Fig. 4 Silver staining results for poly(A)$_{25}$ RNA and XIST-5′ pulled-down proteins. Despite some proteins being pulled down by both of the RNAs, there are several proteins specifically pulled down by XIST-5′ as compared to the negative control. Unique bands in XIST pull-down are shown using arrows. Some of these bands were cut and identified using mass spectrometry. A number of them such as HNRPU, SFPQ, DHX9, ILF3, and XRN2 matched with previous publication [7]

4 Notes

1. To maintain RNase- and keratin-free conditions, use disposable, sterile Eppendorfs, and pipette tips with filters, and ensure that the workspace is thoroughly clean (with RNase Zap and ethanol if possible) and ensure that disposable gloves and a lab

coat are worn at all times. We recommend changing gloves frequently.

2. When designing primers for the PCR, first design the oligonucleotide sequences (excluding the T7 promoter sequence) to have similar melting temperatures before adding T7 promoter sequence at one end.
3. cDNA should be used if primers are designed across different exons; however, if the primers are designed to amplify a region of RNA which is not spliced we have found that genomic DNA works sufficiently well. 2× PCR reactions containing 100 ng genomic DNA should be sufficient to produce 1 ug template DNA.
4. It is necessary to optimize the PCR annealing temperature for all primer pairs. We used a temperature range between 48 and 58 °C, in 2 °C increments, to determine the optimum temperature. At optimum temperature only one bright band corresponding to PCR product of desired length should be observed upon gel electrophoresis. Furthermore, we found it useful to increase the number of PCR cycles to 45 to increase the yield.
5. In some cases it might not be possible to produce a singular, clear band after PCR. In such a scenario the correct size band must be extracted from the gel. For those that do produce a unique band, direct purification of PCR product can be used.
6. We advise sequencing of all PCR products to ensure that the correct DNA is being produced and used for subsequent reactions.
7. Following gel extraction or PCR cleanup using spin columns, salt contamination of the sample is often high. It is important that the product is further purified prior to in vitro transcription using ethanol precipitation; otherwise it can inhibit the T7 RNA polymerase.
8. We recommend overnight incubation of ethanol precipitation; however incubation times >2 h can be sufficient.
9. It is not necessary to have 1 ug of DNA; we have used as little as 500 ng of DNA successfully; however, the amount of RNA product produced may be reduced.
10. For RNA with a strong secondary structure, lowering the reaction temperature to 16 °C or 4 °C and incubating for longer (24–72 h) improve the reaction yield. We found that for the 672 nucleotide fragment of XIST (which contains multiple hairpins) no product was detected following incubation at 16 °C for 16 h, but following 72-h incubation at 4 °C RNA product could be detected on agarose gel and ~40 μg of RNA was produced.

11. If performing RNA extraction we strongly recommend phenol-chloroform-isoamyl alcohol extraction after T7 RNA polymerase reaction, as RNA spin column extraction led to the significant loss of RNA product.
12. When removing the aqueous layer, ensure that none of the organic phase is taken, as this will lead to bad-quality RNA sample.
13. 50–100 pmol of RNA is required in 6–16 μL for biotin labeling; therefore it is important to ensure that the RNA is concentrated following ethanol precipitation. Although the addition of DMSO and heating to 85 °C are not required, we recommend the inclusion of this step, especially for long RNAs or RNAs with a strong secondary structure.
14. We advise that 100 pmol of RNA is biotin labeled, as some will be lost during ethanol precipitation.
15. Overnight incubation in not required; we have used 6-h incubation successfully.
16. Do not perform phenol-chloroform extraction, as it will significantly reduce the amount of RNA retrieved. When performing ethanol precipitation after biotin labeling, increase the volume of the reaction to ~300 μL to dilute the 30% PEG and aid pellet formation during centrifugation.
17. The pellet may not resuspend completely at this stage.
18. It is important that the entire pellet is resuspended, which can be difficult. We found that sonication of the sample in a cold water bath sonicator, for 4–6 10-s bursts with 10-s intervals, followed by vortexing and syringing ~30 times, was sufficient to resuspend the pellet and produce 1.45 mg of protein from 4.5×10^7 K562 cells.
19. To maximize the amount of RNA that binds to the beads, we recommend rotation rather than agitation to be used.
20. We found it more efficient to elute in a smaller volume to concentrate the protein samples.
21. To maximize elution, it is preferable to heat the streptavidin beads with loading buffer than to elute by incubating at 37 °C with elution buffer for 30 min.
22. Membranes were placed in methanol for 1 min and washed with 1x transfer buffer prior to transfer.
23. We have optimized transfer for the following protein sizes, gels, and BIO-RAD Trans-Blot® SD semidry transfer machine : 16 min used to transfer EZH2 (100 kDa) from a 10% (1.5 mm) polyacrylamide gel and HuR (36 kDa) from a 12.5% (1.5 mm) polyacrylamide gel. For alternative proteins or transfer system transfer times may need to be reoptimized.

24. Odyssey blocking solution was used for antibodies that recommend BSA blocking (as it enables better detection when using the Odyssey imaging system). We used fluorescently labeled secondary antibodies and used Odyssey imager to scan the blots but **steps 9–12** can be replaced with any standard western blot techniques such as chemiluminescence method.
25. A minimum of 40-min fixation is required.
26. Longer development times will enable the detection of proteins with lower concentrations; however the amount of background staining also increases.
27. It is very important to maintain keratin-free environment if the procedure is to be followed by mass spectrometry. Besides keeping workspace dust free, use fresh reagents and tubes. If possible work in a laminar flow hood.

Acknowledgements

This work was supported by University of Birmingham and BBSRC MIBTP program.

References

1. Carninci P, Kasukawa T, Katayama S et al (2005) The transcriptional landscape of the mammalian genome. Science 309: 1559–1563
2. Zhao J, Sun BK, Erwin J et al (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X-chromosome. Science 322:750–756
3. Rinn JL, Kertesz M, Wang JK et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323
4. Hung T, Wang Y, Lin MF, Koegel AK et al (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet 43:621–629
5. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15:8783–8798
6. Zhao J, Ohsumi TK, Johnny T et al (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40: 939–953
7. Chu C, Zhang QC, da Rocha ST et al (2015) Systematic discovery of Xist RNA binding proteins. Cell 161:404–416
8. McHugh CA, Chen CK, Chow A et al (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521:232–236
9. Minajigi A, Froberg JE, Wei C et al (2015) A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science 349 aab2276

Chapter 10

Understanding RNA-Chromatin Interactions Using *Ch*romatin *I*solation by *R*NA *P*urification (ChIRP)

Ci Chu and Howard Y. Chang*

Abstract

ChIRP is a novel and easy-to-use technique for studying long noncoding RNA (lncRNA)-chromatin interactions. RNA and chromatin are cross-linked in vivo using formaldehyde or glutaraldehyde, and purified using biotinylated antisense oligonucleotides that hybridize to the target RNA. Co-precipitated DNA is then purified and analyzed by quantitative PCR (qPCR) or high-throughput sequencing.

Key words lncRNA, Chromatin, High-throughput sequencing, Cross-linking

1 Introduction

Many lncRNAs are found to have gene regulatory functions [1], but how lncRNAs control gene expression has largely remained a mystery. This is in part due to a lack of technologies that probe genomic binding sites of lncRNAs. Here we describe ChIRP, the first method to isolate chromatin bound by lncRNAs in vivo [2]. Much like steps taken in chromatin immunoprecipitation (ChIP), we cross-link lncRNA-chromatin interactions in vivo, solubilize chromatin by sonication, use affordable biotinylated antisense DNA oligonucleotides to capture the target lncRNA, and isolate pure lncRNA and associated proteins or genomic DNA after stringent denaturing washes (Fig. 1). Since the original publication of the method and a video demonstration [2, 3], we and many labs around the world have ChIRPed many lncRNAs successfully to map not only their genomic binding sites, but also interacting proteins and RNAs (reviewed in ref. 4).

Alternative methods have since been reported (e.g., CHART, RAP) [5, 6]. These protocols are largely similar to ChIRP, and differ mainly in ways of cross-linking, chromatin fragmentation, and probe design (reviewed in ref. 4). We have also expanded on

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_10,

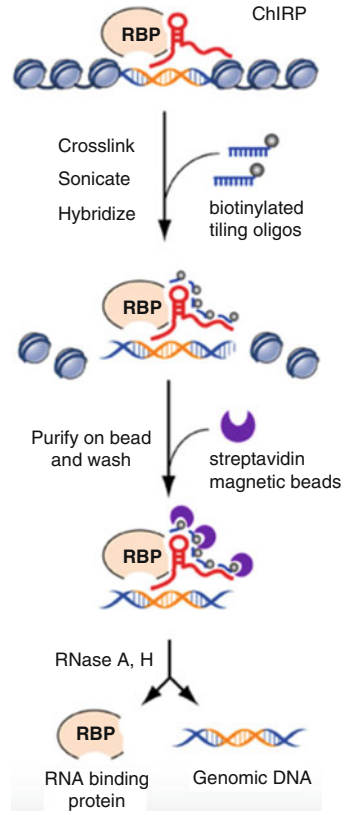


Fig. 1 Flow chart of ChIRP. Previously published in ref. 2

ChIRP and invented a related technique that interrogates functional subunits of an lncRNA (dChIRP) [7, 8]. In addition, many technical improvements have been made over the original ChIRP protocol [3]. Here we describe the most up-to-date working procedures of the standard ChIRP-seq method.

2 Materials

Prepare all solutions using ultrapure nuclease- and proteases-free water and reagents. Prepare and store all solutions at room temperature. Use a fume hood when cross-linking cells or tissues with formaldehyde or glutaraldehyde, or using organic solvents (e.g., trizol). Follow proper waste disposal procedures when disposing harmful reagents.

2.1 ChIRP Buffers

1. Cell lysis buffer: 50 mM Tris–Cl pH 7.0, 10 mM EDTA, 1% SDS. Add 1 mM PMSF (Sigma-Aldrich, St. Louis, MO, USA), protease inhibitors (P.I., GE Healthcare, Piscataway, NJ, USA),

and Superase-in (Life Technologies, Carlsbad, CA, USA) fresh before use (*see* **Note 1**).

2. Hybridization buffer (always make fresh): 750 mM NaCl, 1% SDS, 50 mM Tris–Cl pH 7.0, 1 mM EDTA, 15% formamide (store in parafilm-sealed bottle in the dark at 4 °C, Life Technologies) (*see* **Note 2**). Add 1 mM PMSF, P.I., and Superase-in fresh before use.
3. Wash buffer: 2× SSC (diluted from 20× SSC, Life Technologies), 0.5% SDS, add 1 mM PMSF fresh before use.
4. Proteinase K (pK) buffer: 100 mM NaCl, 10 mM Tris–Cl pH 7.0, 1 mM EDTA, 0.5% SDS, add 5% v/v pK (Life Technologies) fresh before use.
5. DNA elution buffer: 50 mM $NaHCO_3$, 1% SDS.
6. RNase A: 10 mg/ml solution in water (Sigma).
7. RNase H (Epicentre, Madison, WI, USA).
8. Glutaraldehyde (EM grade) (Sigma) (*see* **Note 3**).
9. Formaldehyde (methanol free) (Pierce, Rockford, IL, USA).

2.2 ChIRP Components

1. Bioruptor (Diagenode, Denville, NJ, USA).
2. Covaris Ultrasonicator (Covaris, Woburn, MA, USA).
3. Falcon tubes for sonication (Corning, Corning, NY, USA) (*see* **Note 4**).
4. C-1 magnetic beads (Life Technologies).
5. Dynamag-15/Dynamag-2 magnet (Life Technologies).
6. Hybridization oven.

2.3 ChIRP Probes

3′-Biotin-TEG-20-mer DNA oligonucleotides (*see* **Methods**).

2.4 Molecular Biology Reagents

1. PCR purification kit (Qiagen, Venlo, Limburg, The Netherlands).
2. Rneasy mini kit (Qiagen).
3. Trizol (Life Technologies).
4. Chloroform.
5. Phenol:chloroform:isoamyl alcohol.
6. Glycoblue (Life Technologies).
7. Glycine.
8. PBS pH 7.4.
9. RNase-free buffer kit (Life Technologies).
10. Phase lock gel heavy (5 Prime, Gaithersburg, MD, USA).

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 Probe Design

1. Design 20-mer DNA probes with single-molecule FISH online designer (https://www.biosearchtech.com/stellarisdesigner/). Use maximum masking for nonspecific binding (level 5) and adjust to achieve even and sparse tiling (1 probe per 200–400 nt of RNA) across the target RNA. Manually check that probe sequences are not mapping to repeats or other homologous transcripts.
2. Label probes based on their positions along the RNA, and separate them into "even" and "odd" pools, to create two independent probe sets. Double-check that there is no cross homology between the two sets (*see* **Note 5**).
3. Synthesize probes from Biosearch Technologies (Petaluma, CA, USA) or any local equivalent vendor with 3′-biotin-TEG modification.
4. Pool probes into "even" and "odd" sets, dilute to 100 μM (all probes, not 100 μM each probe), and store at −20 °C (*see* **Note 6**).

3.2 Cell Harvesting

1. Wash confluent cell culture in plates with PBS once, trypisinize, pellet at 800RCF for 3 min. Wash cell once again with PBS, pellet again, aspirate all PBS carefully.
2. Resuspend cells in 1% glutaraldehyde, make sure that cell clumps are broken up, and cross-link for 10 min with mixing. Alternatively, cross-link cells in 3% formaldehyde for 30 min (*see* **Note 7**).
3. Quench glutaraldehyde or formaldehyde with 0.125 mM glycine for 5 min. Pellet cells at 2000 rcf (*see* **Note 8**), rinse again with PBS, pellet again, and carefully aspirate all liquid. Weigh cell pellets and record weight. Snap freeze cell pellets in liquid nitrogen and store at −80 °C indefinitely.

3.3 Sonication

1. Use 20 million cells typically for one ChIRP experiment.
2. Dissolve cell pellets in cell lysis buffer (100 mg cell pellet/1 ml buffer). Resuspend well.
3. Sonicate cell lysate in either a bioruptor or a Covaris ultrasonicator, until lysate is clear.
4. Take a 5 μl aliquot of lysate, dilute with 95 μl pK buffer, and incubate at 50 °C for 45 min. Extract DNA using PCR purification kit as per the manufacturer's instructions. Run out DNA on a 1% agarose gel and make sure that the bulk is below

500 bp. If DNA is still >500 bp, repeat sonication till the size is correct.

5. Spin down lysate at 16,000 rcf for 10 min at 4 °C. Transfer supernatant to fresh tubes and either immediately proceed to ChIRP or store at −80 °C indefinitely.

3.4 ChIRP

1. Design the experiment with the appropriate controls (*see* **Note 9**).
2. Save some lysate for "input" samples (typically 10 μl for RNA and DNA each). Leave input samples in the 37 °C during the hybridization and wash steps. Aliquot equal amounts of lysate for experimental samples and controls.
3. For RNase treatment control, add RNase A to the lysate (10 μg/ml final concentration) (*see* **Note 10**). Incubate together with non-treated samples at 37 °C for 30 min with end-to-end rotation. For alternative negative controls, skip this step.
4. To all samples, add 2× volume hybridization buffer, and ChIRP probes (1 μl of 100 μM probes per 1 ml of lysate). Incubate at 37 °C with end-to-end rotation for 4–16 h in a hybridization oven (*see* **Note 11**).
5. Prior to completion of hybridization, wash C-1 beads three times on a magnet stand with cell lysis buffer (use 100 μl beads per 1 μl of 100 μM probes). Remove all buffer after last wash.
6. When hybridization is complete, briefly spin down the tubes, and use 1 ml of sample to transfer beads into the reaction. Continue to mix samples at 37 °C for 30 min.
7. Briefly spin down contents, place tubes on a Dynamagnet for 1–2 min or until beads clear up. Decant and use a pipette tip to remove residual liquid.
8. Wash with 1–5× wash buffer (v/v to original beads volume). Do each wash at 37 °C with constant mixing for 5 min (*see* **Note 12**). Perform five washes in total.
9. At last wash, transfer 10% of sample to a fresh tube. This is for RNA analysis. Use the bulk material (90%) for DNA extraction and analysis.
10. Remove residual buffer after last wash. Proceed to DNA or RNA extraction.

3.5 RNA Extraction

1. Resuspend RNA samples (on beads) in 100 μl pK buffer. Top up RNA input to 100 μl with pK buffer also.
2. Incubate at 50 °C for 45 min with constant mixing to reverse cross-links.
3. Heat samples at 95 °C for 10 min (*see* **Note 13**).

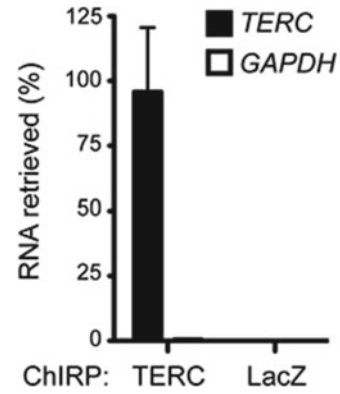


Fig. 2 TERC-ChIRP pulls down most of the RNA from the cell, while LacZ-ChIRP (non-targeting control) does not. GAPDH (negative control) is not enriched by either ChIRP experiment. Previously published in ref. 2

4. Add to each RNA samples 500 µl Trizol. Vortex to mix thoroughly, and incubate for 5 min. Add 100 µl chloroform, and vortex to mix again. Incubate for another 5 min.
5. Spin down all samples at 4 °C for 15 min. Transfer top aqueous layer to a fresh tube. Add half volume of 100% ethanol, mix well, and pass through RNeasy mini columns.
6. Follow the manufacturer's protocol to clean up RNA on column and elute in 30 µl nuclease-free water.
7. Use 1 µl RNA eluent per well for quantitative reverse transcription PCR (qRT-PCR) to confirm RNA enrichment. An abundant housekeeping transcript can be used as negative control (e.g., 18S rRNA, GAPDH mRNA). Typically 250- to 1000-fold enrichment is achieved over negative control and majority of input target RNA is captured (Fig. 2).

3.6 DNA Extraction

1. Resuspend DNA samples (on beads) in 150 µl DNA elution buffer. Top up DNA input to 150 µl with elution buffer.
2. Add 10 µl RNase A (10 mg/ml) and 10 µl RNase H (10 U/µl) per 1 ml of DNA elution buffer. Vortex to mix.
3. Incubate at 37 °C for 30 min with shaking.
4. Place samples on magnet for 1 min or until beads clear up, and transfer supernatant to a fresh tube.
5. Repeat RNase extraction and pool second eluant with the first batch. Total volume is 300 µl per sample.
6. Add 15 µl pK per sample. Vortex to mix. Incubate at 50 °C for 45 min with mixing.
7. Spin down phase-lock gel tubes. Transfer pK-treated DNA samples to each gel tube, and add 300 µl phenol:chloroform:isoamyl

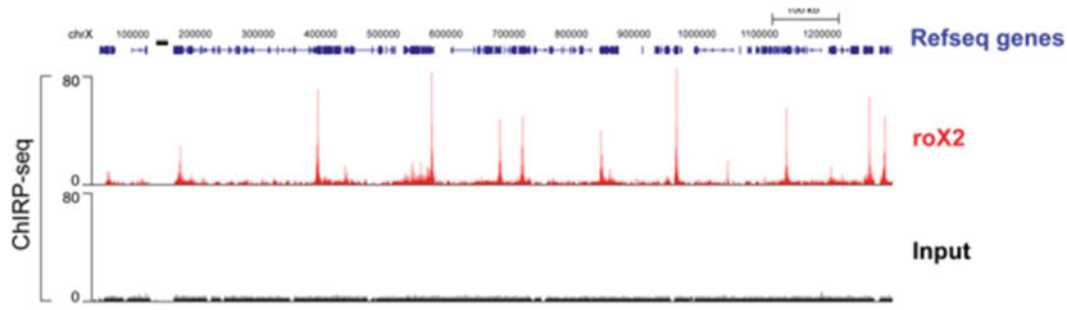


Fig. 3 ChIRP-seq data of roX2 RNA in SI2 *Drosophila* cells. "Even" and "odd" data were merged to produce the signal overlap. Previously published in ref. 2

alcohol per sample. Vigorously mix by hand for 10 s, and spin down at 16,000 rcf for 5 min at 4 °C. Transfer aqueous layer above the gel to a fresh tube.

8. To each sample, add 3 μl glycoblue (1% v/v), 30 μl 3 M NaOAc (10% v/v), and 900 μl 100% ethanol (3× v/v). Mix well and store at −20 °C overnight.
9. Next day, spin down DNA precipitate at 16,000 rcf at 4 °C for 30 min. Carefully decant supernatant without disturbing the pellet. Add 1 ml 70% ethanol and vortex to mix. Spin at 16,000 rcf at 4 °C for 5 min. Remove supernatant. Briefly spin again and remove residual liquid. Air-dry for 1 min, and resuspend DNA pellets in 30 μl EB provided in PCR purification kit.
10. DNA samples are ready for analysis by qPCR or high-throughput sequencing. Post-sequencing analysis is needed to eliminate probe-specific noise and isolate signal overlap between "even" and "odd" pools (Fig. 3).

4 Notes

1. Use P.I. at dilution factor recommended by the manufacturer, but use Superase-in at 200×.
2. Do not leave this buffer (or cell lysate) on ice as precipitation will occur. Brief 4 °C storage is usually fine.
3. Use only fresh glutaraldehyde (and formaldehyde).
4. Do not use polypropylene tubes for sonication; use polystyrene or polyethylene instead.
5. It is critical that no significant homology exists between the even probes and the odd probes, to ensure that the two sets are truly independent sets and to prevent false positive hits in their common signal.

6. Make several aliquots to prevent excessive freeze and thaw cycles.
7. Exact cross-linking condition needs to be optimized for each lncRNA. We have found that for some lncRNAs both glutaraldehyde and formaldehyde work well (e.g., roX2) [2, 7], while for some others only glutaraldehyde works (e.g., TERC binding to telomere) [2]. It is possible that for certain lncRNAs other cross-linkers may be needed to capture their chromatin interactions.
8. Sometimes after cross-linking, cells do not pellet well at 800 rcf. When this happens, a faster centrifugation speed at 2000 rcf may be necessary.
9. Perform control experiment on two levels: (1) ChIRP with "even" and "odd" probes, and take only their overlapped signal, in order to eliminate false-positive hits; (2) ChIRP with "even" and "odd" in RNase-treated chromatin, or a control cell line that does not express the target RNA (e.g., knockdown, inducible promoter, very low endogenous expression), to obtain RNA-independent background noise. In general, genetic controls mentioned above are preferred over using a non-targeting probe, which does not eliminate probe-specific noise.
10. Be very careful not to contaminate experimental samples with RNase A, as it is a very potent nuclease.
11. For convenience, start hybridization late in the day and let it run overnight. If the reaction is <1.5 ml, use Eppendorf tubes. If reaction is >2 ml, use 15 ml Falcon tubes. Use size-appropriate magnets later.
12. For best results, perform all hybridization and wash steps inside a temperature-controlled hybridization oven. Pre-warm all buffers prior to use. If hybridization is performed in a 15 ml Falcon tube, it may be more convenient to wash in a reduced volume in an Eppendorf tube.
13. Samples are boiled to complete the cross-link reversal and denature streptavidin beads. The brief boiling step will do minimal harm to the RNA samples at pH 7.0.

Acknowledgement

This work was supported by the Singapore Agency for Science, Technology, and Research (A*STAR to C.C.) and the US National Institutes of Healthy (H.Y.C.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

References

1. Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166
2. Chu C et al (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44(4):667–678
3. Chu C, Quinn J, Chang HY (2012) Chromatin isolation by RNA purification (ChIRP). J Vis Exp 61:e3912
4. Chu C, Spitale RC, Chang HY (2015) Technologies to probe functions and mechanisms of long noncoding RNAs. Nat Struct Mol Biol 22(1):29–35
5. Engreitz JM et al (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341(6147):1237973
6. Simon MD et al (2011) The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A 108(51):20497–20502
7. Quinn JJ et al (2014) Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat Biotechnol 32:933–940
8. Quinn JJ, Chang HY (2015) In Situ Dissection of RNA Functional Subunits by Domain-Specific Chromatin Isolation by RNA Purification (dChIRP). Methods Mol Biol 1262:199–213

Chapter 11

Analysis RNA-seq and Noncoding RNA

Alberto Arrigoni, Valeria Ranzani, Grazisa Rossetti, Ilaria Panzeri, Sergio Abrignani, Raoul J.P. Bonnal*, and Massimiliano Pagani*

Abstract

RNA-Seq is an approach to transcriptome profiling that uses deep-sequencing technologies to detect and accurately quantify RNA molecules originating from a genome at a given moment in time. In recent years, the advent of RNA-Seq has facilitated genome-wide expression profiling, including the identification of novel and rare transcripts like noncoding RNAs and novel alternative splicing isoforms.

Here, we describe the analytical steps required for the identification and characterization of noncoding RNAs starting from RNA-Seq raw samples, with a particular emphasis on long noncoding RNAs (lncRNAs).

Key words RNA-seq, lncRNAs, Bioinformatics

1 Introduction

In recent years, advances in transcriptome reconstruction technologies have made possible the identification and the characterization of thousands of novel long noncoding RNAs (lncRNAs) from short read RNA-seq data [1–3].

The rapid increase of sequencing depth and read length has considerably improved the accuracy of transcript reconstruction and offers the unprecedented possibility to characterize lncRNAs on a global scale.

LncRNAs are defined as transcripts of length >200 nucleotides that are characterized by a low coding potential [4]. The choice of this length threshold is somewhat arbitrary, but it is instrumental in order to separate lncRNAs from other noncoding RNA classes, such as microRNAs (miRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), and other short RNAs [4].

LncRNAs are broadly classified according to the genomic context in which they are located (Fig. 1)—antisense lncRNAs are transcripts that span at least one exon of a nearby protein coding,

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_11,

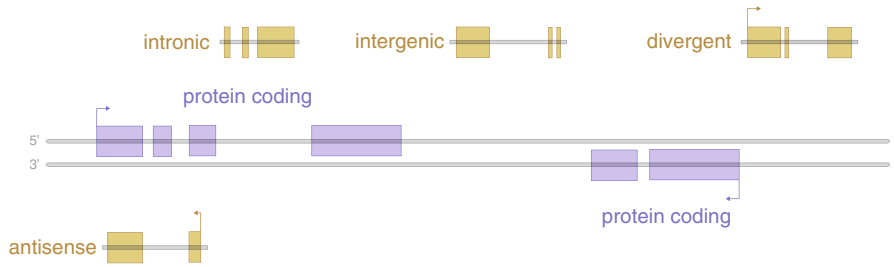


Fig. 1 Long noncoding RNAs are classified according to the genomic context in which they are located [5]

and are transcribed in the opposite direction—intronic lncRNAs originate from intronic regions, and they do not overlap any annotated exon—bidirectional lncRNAs are transcripts that initiate in a divergent fashion from the promoter of a protein-coding gene—intergenic lncRNAs are lncRNAs with separate transcriptional units from protein-coding genes [5].

In contrast to what has been reported for other noncoding RNA classes [6], long ncRNAs lack strong inter-species conservation [7]. Moreover, the evolutionary linkage between lncRNAs in different species is difficult to infer since the majority of approaches for conservation studies are based on primary sequence analysis. More comprehensive studies that integrate primary sequence analysis, structure, and functional role of lncRNAs are therefore needed to assess their conservation on a global scale [7].

This aspect is also relevant in light of the correlation that exists between lncRNA structure and function, as it has been demonstrated that they physically associate with chromatin modifiers to alter the epigenetic state of target regions, whether in cis or in trans [5].

LncRNAs are specifically expressed in different tissues, as demonstrated by integrative studies that used RNA-seq to accurately detect and quantify them [1]. This contributed to develop the notion that lncRNAs have been used by evolution as molecular switches whose activity influences the onset and the maintenance of differentiative states of several different tissues and cellular populations [4].

The de novo identification of transcripts from RNA-seq data is performed using algorithms that follow two slightly different approaches: *mapping-first* algorithms like Cufflinks [8] and Scripture [9] and *assembly-first* methods like Trinity [10], SOAPdenovo [11], and Oases [12] (*see* **Note 3**).

Then, some peculiar characteristics of noncoding transcripts are leveraged in order to isolate lncRNAs from protein-coding transcripts in datasets that are tipically constituted by thousands of previously unidentified transcripts. The analysis of evolutionary patterns across different species (PhyloCSF [13]) and (more recently) classifiers trained on linguistic features (iSeeRNA [14], CPAT [15], PLEK [16]) are used for this purpose.

Moreover, the use of chromatin immunoprecipitation (ChIP) coupled with NGS (ChIP-seq) provides important and complementary information that fosters the identification process of novel lncRNAs [17]. These methodologies rely on the generation of chromatin maps based on the presence of epigenetics marks (H3K4me3, H3K36me3) and Pol II occupancy to define novel transcriptional units.

Here, we present an analytical protocol (Fig. 2) for the characterization of lncRNAs starting from raw RNA-seq samples obtained from sequencing Poly-A$^+$ fractions using paired-end Illumina reads.

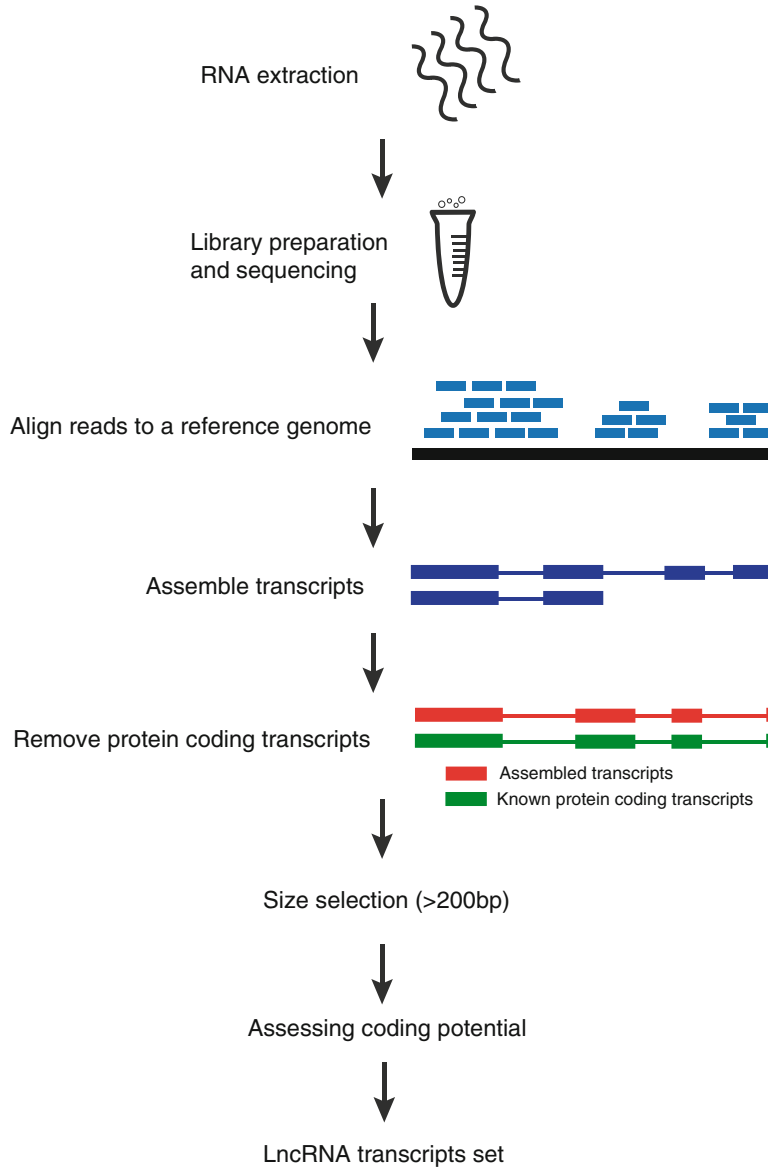


Fig. 2 Overview of the pipeline for the identification of lncRNAs. Schematics of the workflow described in Subheading 3

2 Materials

2.1 Description of a Standard Bioinformatics Architecture for NGS Data Analysis and Software Requirements

1. Any UNIX-based operating system (Linux, BSD, Solaris, Mac OSX) could be used to perform the analyses described in Subheading 3, although the majority of tools for NGS analyses have been developed with Linux as first choice. The methodologies described in the following sections have been tested on Ubuntu 12.04.5 LTS.
2. Since the pipeline involves the use of the STAR mapper, we recommend the use of a workstation with at least 30 GB of RAM.
3. Experimental settings: The pipeline described in Subheading 3 was designed and tested for paired-end, poli-A^+ Illumina libraries. Read length is ≥100 pb.

2.2 Annotation Sources and Integration

1. Reference annotation .gtf files and .fasta genomic sequence files (hg19/hg38) can be obtained at http://www.ensembl.org/info/data/ftp/index.html (Ensembl FTP data repository) or ftp://hgdownload.cse.ucsc.edu/goldenPath/ (UCSC FTP data repository).
2. Integration of available transcript annotations not included in UCSC/Ensembl databases can be merged to the official annotation release by using Cufflinks' utility Cuffcompare [8].

2.3 Software Versions Used for this Protocol

1. cufflinks/2.2.1, fastqc/0.11.3, samtools/1.2, STAR/2.4.1c, trimmomatic/0.33, cutadapt/1.8, R/3.2.0 (DESeq2), HTSeq/0.6.1, CPAT v1.2.

3 Methods

3.1 Preprocessing

1. *Quality assessment* (pre-trimming): (*See* **Note 1**) Raw .fastq files resulting from a paired-end sequencing experiment are analyzed in order to assess overall quality. The initial inspection is carried out using FastQC v0.11.3. This analysis step is important, as it may raise attention regarding library preparation problems that may have occurred (the analysis is carried out both on "forward" and "reverse" strand reads) [<fastqc> --outdir . <sample_path>/R1(R2).fastq.gz].
2. *Adapter removal*: Adapter sequences are removed using cutadapt [18]. This is usually necessary when the read length of the sequencing machine is longer than the molecule that is sequenced (for example when sequencing microRNAs). "Cutadapt" is run both for R1 and R2 indicating adapters' sequences [--anywhere <adapter1> --anywhere <adapter2> --overlap 10 --times 2 --mask-adapter].

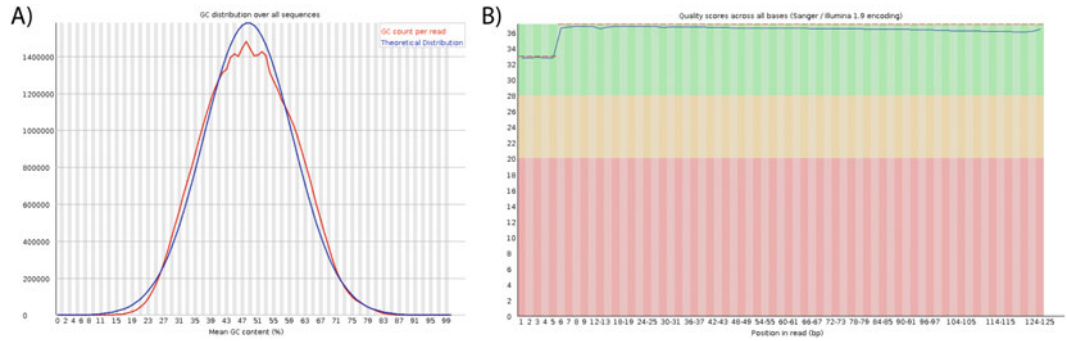


Fig. 3 Quality control after trimming using FastQC output (high-quality data): (**a**) Per-sequence GC content is shown as a density graph and compared to a random Gaussian function. (**b**) Boxplot graphs representing quality scores over the entire length of reads in the sample. After the trimming step only high-quality reads are expected to be found in output

3. *Trimming*: Trimmomatic [19] is used with a sliding window approach to remove lower quality bases at the end. Standard parameters used for phred33 encoding: ILLUMINACLIP (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15), MINLEN parameter is set to 50.
4. *Quality assessment* (post-trimming): The output of Trimmomatic is used as input to FastQC in order to inspect the results of the pre-processing steps (adapter removal and trimming) (results for high-quality Illumina data are reported in Fig. 3).

3.2 Alignment to the Reference Genome

1. *Reference indexing*: In order to map the short reads to a reference genome, a .fasta reference index should be built for the STAR aligner [20].
2. *Mapping*: (*See* **Note 2**) Paired-end reads are mapped to the reference genome using the STAR aligner [STAR --genomeDir <index_star> --runThreadN <cpu_number> --readFilesIn <trimmed>_R1.fastq.gz <trimmed> _R2_P.fastq.gz --readFilesCommand zcat]. The .sam output of STAR is then converted to its compressed format .bam [samtools view -bS aln.sam > aln.bam], and it is indexed for further analyses (e.g., for use with genome browser and further quality inspection of mapping) using command "samtools index" (https://github.com/samtools).
3. *Alignment statistics*: Statistics are calculated on output bam files [samtools flagstat bam_name>].

3.3 Data Exploration

1. *BAM sorting*: In order to perform data exploration using DESeq2 .bam files are sorted by gene identifier, using command [samtools sort -n].
2. *Read counts*: The overlap of reads with annotation features found in the reference .gtf is calculated using HT-seq [21].

The output computed for each sample (raw read counts) is used as input for Bioconductor's DESeq2.

3. *Principal component analyses of RNA-seq samples*: (Fig. 4a) Raw counts produced with HT-seq are parsed using DESeq2 (v 1.8.1) [22] for each sample and the biological labels (e.g., cellular population identifiers) are provided in a tabular "samplesheet" along with the count files. Raw counts are normalized using DESeq2's function "rlog," which outputs "variance stabilized" values and transforms the original count data to the log scale. Normalized counts are used to calculate and plot principal component analysis (PCA) (using DESeq2's "plotPCA" function) (http://www.bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.pdf) (example PCA performed using 12 samples belonging to two different biological classes is depicted in Fig. 4a).
4. *Heatmap of sample-to-sample distances*: (Fig. 4b) Normalized counts used to produce the PCA plot of the previous step are also used to calculate sample-to-sample distances with DESeq2's functions "hclust" and "heatmap.2" (example hierarchical cluster analysis performed using 12 samples belonging to two different biological classes is reported in Fig. 4b).

3.4 De Novo Identification of Transcripts

1. *BAM sorting*: In order to perform de novo discovery using Cufflinks [8], the .bam files of every sample should be sorted by coordinate using samtools.

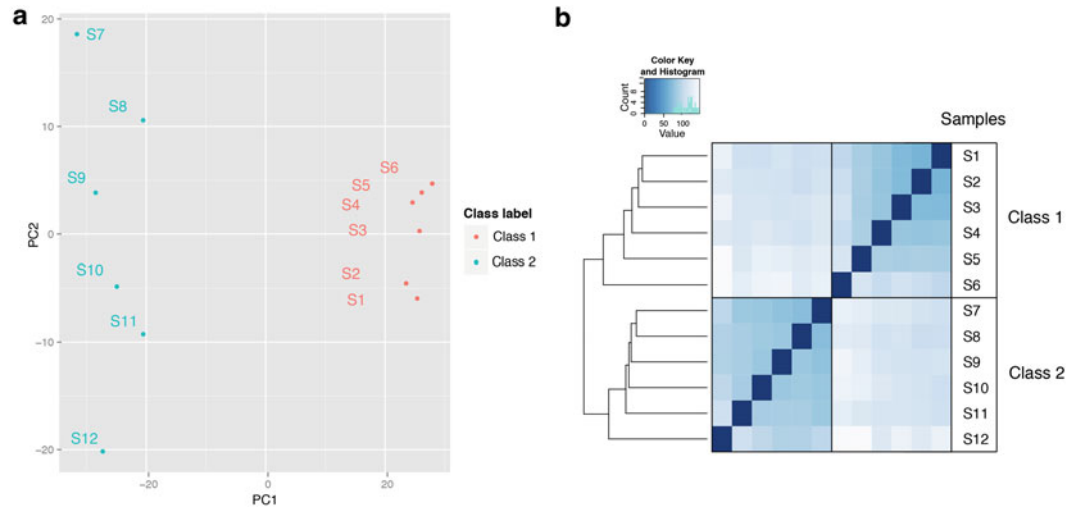


Fig. 4 Exploratory data analysis of raw RNA-seq samples performed using DESeq2. (**a**) Principal component analysis (PCA) performed on rlog-normalized expression data reveals a separation between samples belonging to different biological classes ("labels"). (**b**) Similar results are obtained by performing hierarchical clustering on rlog-normalized expression data

2. *Cufflinks de novo*: (*See* **Notes 3** and **4**) Transcript identification is performed on a per-sample basis using a tool from the Cufflinks suite ("RABT") [8]. For the software to discover novel transcripts, the reference .gtf file used in the initial mapping step must not be supplied. The resulting file "transcripts.gtf" found in the output folder contains the assembled transcripts identified by Cufflinks [cufflinks -o OutputDirectory/mappedReads.bam].
3. *Cuffmerge*: (*See* **Note 5**) Predicted transcripts (contained in one .gtf file for each sample) are merged using Cufflinks utility "cuffmerge." The .gtf reference file is supplied to "cuffmerge" so that newly discovered genes and transcripts are integrated in the original annotation. The resulting .gtf file is used as input for the differential expression analysis [cuffmerge <gtfs_list> -g <ref.gtf> -s <ref_fasta> -p <cpu_number>] (*see* **Notes 6** and **13**).
4. *Length filter*: The transcripts identified using the de novo approach are filtered by length, as lncRNAs are by definition longer than 200 pb. Sequence lengths can be evaluated after the results of "gtf_to_fasta" (included in the Cufflinks suite).

3.5 De Novo Identification of lncRNAs: Coding Potential Evaluation

1. The first step of the classification pipeline filters out the transcripts for which a PFAM match [23] is reported by HMMER's utility "hmmscan." The input for this analysis is a multi .fasta file containing the translated nucleotide sequences of each transcript (all six possible transcription frames are considered) (*see* **Note 10**).
2. The coding potential of the remaining transcripts is calculated using CPAT [15]. The threshold for the combined classifier's score is 0.364, where transcripts scoring <0.364 are considered putative "noncoding" (threshold calculated for the human transcriptome by the authors of CPAT) (*see* **Note 11**).

3.6 Differential Expression

1. *Cuffdiff*: (*See* **Note 7**) Gene/isoform differential expression is performed using Cuffdiff (included in the Cufflinks suite), having as input the .bam files produced by the quantification step. Input .bam files should be grouped in the command line (1.bam,2.bam 3.bam,4.bam) in order to describe different biological classes/cellular populations (class separator is the white space). Gene/transcript annotation used for the differential expression analysis must be the .gtf file produced in the previous step by cuffmerge [cuffdiff [options]* <transcripts.gtf> <sample1.bam>,<sample2.bam> <sample3.bam>,<sample4.bam>].
2. The selection of differentially expressed genes/isoforms from Cuffdiff results is carried out by extracting *P*-value statistics and log-fold change ratios for the classes identified in input. Among the several files produced by Cuffdiff, *gene_exp.diff* can be loaded into Excel/R and used for further analyses and inspection (*see* **Note 8**).

3. *Consistency filter*: Cuffdiff's output is used to assess the consistency of novel genes across different samples of the same class/cellular population (*see* **Note 9**).

3.7 Downstream Analyses and Visualization

1. Downstream analyses (*see* **Note 12**) are performed on normalized gene expression values obtained from Cuffdiff output. These can be loaded into external visualization software such as Mev (http://www.tm4.org/mev.html), Excel, or R. Cluster analysis is then performed using K-means algorithm (*see* **Note 14**) in order to identify the lncRNAs that are specifically expressed in a single cellular population/class [1].
2. To further investigate the specificity of expression values for lincRNAs the JS (Jensen Shannon) score is calculated on the vector of values returned by Cuffdiff using an appropriate model distribution [1, 24].

4 Notes

1. We recommend using compressed .fastq files in order to limit disk occupancy. Compressed files are directly parsed by software used in subsequent steps of the pipeline, i.e., mapping step (STAR).
2. If several samples are to be analyzed consequently using STAR, the option "--genomeLoad LoadAndKeep" should be added to the command line. STAR will then load the genome index into shared memory so that it can use it the next time a calculation is run.
3. Cufflinks uses a *mapping-first* approach to attempt transcript reconstruction by leveraging genome sequence information to reduce the computational complexity of calculations. Other algorithms are based on an *assembly-first* approach, and though being more computationally expensive can reach higher sensitivity (more transcripts are identified). It is often a good idea to combine these two different approaches in order to obtain higher yields. For human data, it is advisable to start from more than 50 million read pairs to balance specificity and sensitivity of detection. (This estimate is based on the Trinity publication [10], where 52.6 million 76 bp read pairs were used for the reconstruction.)
4. If multiple de novo strategies (e.g., Cufflinks, Scripture [9], Trinity [10]) are used annotation outputs (.gtf files) must be merged into a single coherent annotation (using "cuffmerge" utility from the Cufflinks suite).
5. The Cufflinks de novo procedure followed by Cuffmerge generates a new custom nomenclature for novel genes and iso-

forms, which are, respectively, named with a standard code followed by a progressive integer "id" (XLOC_id and TCONS_id). These new codes can be conveniently renamed when releasing a new catalogue not to overlap any existing external annotation (*see* **Note 13** for further reference).

6. De novo transcript reconstruction is strictly biased by the number and the quality of raw reads that are used for the discovery. For this reason, it may be necessary to pull all available information in a single Cufflinks run in order to increase detection sensitivity. Otherwise, when attempting to characterize transcript isoforms, data should be kept separate.
7. Different normalization options are available for Cuffdiff: "classic-fpkm," "geometric," and "quartile." The default for this parameter is "geometric," so that FPKMs and fragment counts are scaled via the median of the geometric means of fragment counts across all libraries, as described by Anders and Huber in [25]. We recommend using the default option to obtain expression values that are comparable to those obtained with DESeq2.
8. Performing differential analyses using Cuffdiff can be time consuming, and this may impact significantly on the execution average time of the pipeline. For this reason, DESeq2 can be used as an alternative. It should be noted that the differential analysis performed by DESeq2 is at gene level, while for isoform expression characterization (based on exons occupancy evaluation) DEXSeq is available on Bioconductor. Alternative approaches also present in Bioconductor are "edgeR" (http://bioconductor.org/packages/release/bioc/html/edgeR.html) and "bayseq" (http://www.bioconductor.org/packages/release/bioc/html/baySeq.html).
9. For newly identified genes/isoforms, it is important to consider expression consistency across different biological replicates. When analyzing data from different individuals, a discrepancy may indicate biological differences that are peculiar to the set of donors chosen for the experiment. A high consistency ensures that the observed results are in fact real and not technical/biological artifacts.
10. It has been demonstrated [2] that the filtering approach based on PFAM and the subsequent CPAT analysis are consistent with each other, though the combined use of these different methodologies increases specificity of the final results [1].
11. PhyloCSF has been widely used for the prediction of novel lncRNAs in many works. It has been recently demonstrated though [14–16] that classifiers based on linguistic features only (or the integration of different variables like conservation and ORF length prediction) are more accurate and consider-

ably faster (even by orders of magnitude). For these reasons, we suggest to rely on CPAT (or IseeRNA) for the classification step of the pipeline.

12. Downstream analyses can alternatively be performed on normalized counts data (rlogs) produced during the previous "data exploration" step. Love et al. demonstrate in a recent paper [22] that variance-stabilized "rlogs" produced using DESeq2 can be used to perform robust hierarchical cluster analyses.
13. At the time of writing, lncRNA transcript nomenclature is still a source for debate, as only some general guidelines have been proposed by the Human Genome Nomenclature Committee (HGNC). Thus, we advise the readers to follow the suggestions proposed in a recent review by Mattick and Rinn [26]. Antisense lncRNAs are annotated according to the genomic context and take their names after the overlapping genes, while intergenic lncRNAs should be named as LINC-X, where X is a numerical unique identifier.
14. In downstream analyses, K-means clustering is used to identify lncRNAs that are specifically expressed in a single cellular subset or condition. In order to calculate the ideal number of clusters (K) a "silhouette" measure can be used (available in the R package "cluster"). Alternatively, a plot of the within-groups sum of squares by number of clusters extracted can be used (http://www.statmethods.net/advstats/cluster.html).

References

1. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A et al (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25:1915–1927
2. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y et al (2015) The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 47(3):199–208
3. Ranzani V, Rossetti G, Panzeri I, Arrigoni A, Bonnal RJ, Curti S et al (2015) The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. Nat Immunol 16:318–325
4. Pagani M, Rossetti G, Panzeri I, Candia P, Bonnal RJ, Rossi RL et al (2013) Role of microRNAs and long non coding RNAs in CD4(+) T cell differentiation. Immunol Rev 253:82–96
5. Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166
6. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O et al (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37:766–770
7. Diederichs S (2014) The four dimensions of noncoding RNA conservation. Trends Genet 30:121–123
8. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31:46–53
9. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X et al (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28:503–510
10. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512

11. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666
12. Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092
13. Lin MF, Jungreis I, Kellis M (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 27:i275–i282
14. Sun K, Chen X, Jiang P, Song X, Wang H, Sun H (2013) iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. BMC Genomics 14:S7
15. Wang L, Park HJ, Dasari S, Wang S, Kocher J, Li W (2013) CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 41:e74
16. Li A, Zhang J, Zhou Z (2014) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinform 15:311
17. Barski A, Cuddapah S, Cui K, Roh T, Schones DE, Wang Z et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837
18. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12
19. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120, btu170
20. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21
21. Anders S, Pyl PT, Huber W (2014) HTSeq–A Python framework to work with high-throughput sequencing data. Bioinformatics 31(2):166–169, btu638
22. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550
23. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S et al (2004) The Pfam protein families database. Nucleic Acids Res 32:D138–D141
24. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515
25. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106
26. Mattick JS, Rinn JL (2015) Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol 22:5–7

Part III

Chapter 12

The Dynamics of Polycomb Complexes

Daniela Palacios*

Abstract

Polycomb complexes are essential regulators of embryonic and adult stem cells, highly conserved from flies to mammals. Traditionally, their study was based on biochemical and genetic approaches. More recently, the development of novel technologies and the improvement and standardization of existing ones has allowed to address previously unexplored aspects of Polycomb biology, such as dynamics and regulation. In this chapter, relevant researchers in the field discuss novel technologies aimed at dissecting the dynamics of Polycomb complexes in normal and pathological conditions.

Key words Polycomb complexes, Subcellular fractionation, Polycomb body, Fluorescence Resonance Energy Transfer, Automated imaging analysis

Since the discovery of Polycomb group (PcG) genes over 50 years ago [1], numerous approaches have been used to understand how the function of PcG proteins is regulated during development and disease. Initially discovered in Drosophila as negative regulators of Hox genes [1, 2], research in the Polycomb field exploded thanks to two early observations. First, PcG genes are not exclusive of Drosophila but are highly conserved from flies to humans [3–6]; second, their functions extend well beyond the regulation of Hox genes [7]. Nowadays, PcG proteins are amongst the most studied epigenetic regulators in flies and mammals and a myriad of methodological approaches have been developed to address their function.

The study of Polycomb complexes started with biochemical and genetic approaches. Protein purification, subcellular fractionation and imaging analysis are amongst the numerous techniques traditionally used to study PcG dynamics during development and cellular differentiation [8]. If initial biochemistry studies were fundamental to shed light into the mechanisms regulating the formation, distribution and function of Polycomb complexes, it is now clear that finer and integrated approaches are necessary to finely dissect the dynamics that regulate the function of these epigenetics regulators.

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_12,

The following articles cover some of the experimental tools commonly used to study the composition and distribution of PcG proteins and highlight the potential of novel technologies, such as live cell imaging, to study spatial dynamics. Altogether, they provide an exhaustive overview on the different technologies that are commonly used to understand the dynamics of Polycomb proteins.

In their article Marasca et al. use an elegant protocol of subcellular fractionation to investigate PcG proteins redistribution during somatic cells differentiation. Using a four-step **chromatin fractionation** protocol that sequentially extracts soluble, DNase-sensitive, DNase-resistant and insoluble, matrix-associated, proteins, the authors recently uncovered an unexpected association of PcG proteins to the nuclear matrix [9]. Such association had remained obscure for a long time, likely due to the fact that traditional fractionation protocols considered the insoluble pellet as the nuclear scaffold that remains after salt extraction and not as a pure cellular fraction *per se*.

If the subcellular localization of PcG proteins is tightly associated to their biological function, the same holds true for their ability to form functional protein complexes. One of the most representative characteristics of PcG proteins is that they rarely act alone but usually function as part of multi-protein complexes. Initial biochemical approaches such as co-segregation and co-immunoprecipitation experiments strongly suggested that PcG proteins are usually found in complexes of precise stoichiometry [10, 11]. These observations prompted researchers in the field to use finer biochemistry approaches, such as mass spectrometry analysis, to dissect PcG complexes composition.

Here, Morey and di Croce provide a step-by-step protocol to investigate **Polycomb complexes composition**. In particular, they focus their attention on Polycomb Repressive Complex 1 (PRC1). To address dynamic complex formation at different stages, the authors used traditional co-immunoprecipitation experiments followed by mass spectrometry analysis. However, differently from the classical co-immunoprecipitation protocols, in this case antibody is previously cross-linked at the beads to avoid immunoglobulin contamination in the mass spectrometry experiment. In addition, ethidium bromide can be added to the protein extracts prior to immunoprecipitation to investigate if DNA mediates the observed interaction. The described protocol has been recently applied to the analysis of the dynamical composition of different PRC1 complexes during differentiation of Embryonic Stem cells (ESCs) [12].

A second approach to investigate PcG complexes structure and dynamics is through the use of **Fluorescence Resonance Energy Transfer (FRET) technologies**. FRET uses energy transfer between two different fluorophore-tagged proteins to detect

protein–protein interactions both in fixed and live cells. In their article, Cherubini and Zippo provide a detailed protocol for FRET analysis of Polycomb complexes in fixed cells and discuss how FRET technologies can be also used to measure binding affinities in competition assays.

Finally, one of the most characteristic properties of PcG proteins is their precise nuclear localization within the three dimensional nuclear space. It is now clear that fundamental nuclear processes do not occur randomly within the cell nucleus but are compartmentalized at discrete nuclear regions [13]. One of such compartments is composed by multi-protein aggregations of several PcG members and is commonly referred to as Polycomb bodies (or Polycomb foci) [14]. Polycomb bodies are nuclear structures involved in coordinated transcriptional repression of target genes. Whereas the identity of genes recruited to such structures can be addressed by immunofluorescence coupled to fluorescence in situ hybridization (FISH) or by high-throughput technologies derived from the Chromatin Conformation Capture assay, the spatial localization of Polycomb proteins is usually studied by imaging analysis, such as confocal or electron microscopy. Such techniques have been fundamental for our understanding of PcG function. However, they are limited by the need of extensive manual analysis that is rather subjective.

In the last article from this chapter Antonelli and co-workers address this issue and describe a novel algorithm for **automated imaging analysis of Polycomb bodies** from high-resolution cell images stacks. Image segmentation is a key point to all automated imaging analysis. It allows the separation of cells or subcellular structures, such as Polycomb bodies, from background. Here, the authors describe a recently published two-step protocol for image segmentation that can overcome some of the limitations of commonly used methods [9]. First, the protocol uses a modified version of the local Chan–Vese functional algorithm to separate nuclei from background [15]. Then, Polycomb bodies are separated from nuclei regions by means of a modified ISODATA method [16]. The combined use of these algorithms produces segmentation results that are more accurate that those obtained with conventional methods. Such algorithm was successfully applied to investigate the size of Polycomb bodies during muscle differentiation [9] and holds the promise for numerous other applications.

In summary, even if the study of the dynamics of PcG proteins started over 50 years ago, the introduction of novel high-throughput technologies and the refining of existing ones still allow us to get in-depth insight into the mechanism of action of this exciting group of proteins.

References

1. Lewis EB (1978) A gene complex controlling segmentation in Drosophila. Nature 276(5688): 565–570
2. Sato T, Denell RE (1985) Homoeosis in Drosophila: anterior and posterior transformations of Polycomb lethal embryos. Dev Biol 110(1):53–64
3. Hobert O, Jallal B, Ullrich A (1996) Interaction of Vav with ENX-1, a putative transcriptional regulator of homeobox gene expression. Mol Cell Biol 16(6):3066–3073
4. Schumacher A, Faust C, Magnuson T (1996) Positional cloning of a global regulator of anterior-posterior patterning in mice. Nature 384(6610):648
5. Haupt Y et al (1991) Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. Cell 65(5):753–763
6. van Lohuizen M et al (1991) Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. Cell 65(5): 737–752
7. Bracken AP et al (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 20(9):1123–1136
8. Lanzuolo C, Orlando V (2012) Memories from the polycomb group proteins. Annu Rev Genet 46:561–589
9. Cesarini E et al (2015) Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. J Cell Biol 211(3):533–551
10. Strutt H, Paro R (1997) The polycomb group protein complex of Drosophila melanogaster has different compositions at different target genes. Mol Cell Biol 17(12): 6773–6783
11. Franke A et al (1992) Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. EMBO J 11(8):2941–2950
12. Morey L et al (2015) Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms. Cell Stem Cell 17(3):300–315
13. Misteli T (2007) Beyond the sequence: cellular organization of genome function. Cell 128(4):787–800
14. Delest A, Sexton T, Cavalli G (2012) Polycomb: a paradigm for genome organization from one to three dimensions. Curr Opin Cell Biol 24(3):405–414
15. Chan TF, Vese LA (2001) Active contours without edges. IEEE Trans Image Process 10(2):266–277
16. El-Zaart A (2010) Images thresholding using ISODATA technique with gamma distribution. Pattern Recogn Image Anal 20(1):20–41

Chapter 13

Fluorescence Resonance Energy Transfer Microscopy for Measuring Chromatin Complex Structure and Dynamics

Alessandro Cherubini and Alessio Zippo*

Abstract

The Polycomb group (PcG) proteins form regulatory complexes that modify the chromatin structure and silence their target genes. Recent works have found that the composition of Polycomb complexes is highly dynamic. Defining the different protein components of each complex is fundamental for better understanding their biological functions. Fluorescent resonance energy transfer (FRET) is a powerful tool to measure protein–protein interactions, in nanometer order and in their native cellular environment. Here we describe the preparation and execution of a typical FRET experiment using CFP-tagged protein as donor and YFP-tagged protein as acceptor. We further show that FRET can be used in a competition assay to measure binding affinities of different components of the same chromatin complex.

Key words Fluorescence resonance energy transfer (FRET), Acceptor photobleaching, CFP, YFP, Confocal microscopy, PcG

1 Introduction

The Polycomb Group (PcG) proteins form large multimeric complexes that directly interact with the chromatin at their target genes [1]. Previous studies based on biochemical or X-ray crystallography analysis have identified two main families of PcG complexes with a substantial variety in number and composition of PcG proteins [2–6] (Fig. 1). However these tools do not permit to analyze whether the different complex compositions are independent entities with separate function. In addition these biochemical approaches do not allow measuring those dynamic changes of PcG composition that may occur in response to different environmental stimuli.

A common technique used to monitor protein–protein interactions in live or fixed cells is fluorescence resonance energy transfer (FRET) [7]. FRET measurement is based on energy transfer from a donor fluorophore to a suitable acceptor fluorophore through a long-range dipole–dipole coupling mechanism. FRET occurs only when the distance separating the two fluorophores is less than

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_13,

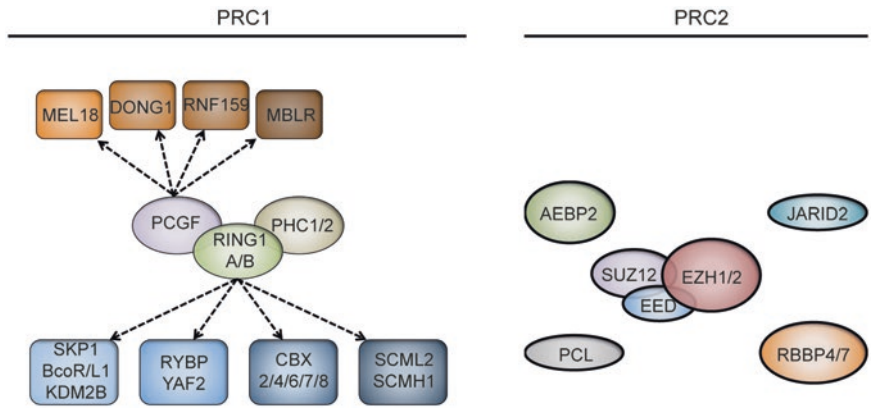


Fig. 1 Composition of the main Polycomb complexes. A schematic representation of the core components of PRC1 and PRC2 are shown. The diversity of Polycomb complexes is shown through the incorporation of homologous proteins. In the PRC1 complex, the core subunits include RING1A/B and a member of both CBX and HPH families. The core subunits of PRC2 are EZH1/2, EED, and SUZ12 proteins. Depending on different environmental stimuli or cell type, additional protein components can associate with the PRC1 and PRC2 complexes, respectively

approximately 10 nm because the efficiency of this energy transfer decreases proportionally to the sixth power of the distance separating the two fluorophores (Fig. 2a, b) [8, 9]. In addition, FRET takes place only if the donor emission spectra overlap substantially with the acceptor absorption spectra (Fig. 2c) [10]. These features, combined with recent technological advances in light microscopy imaging and an increasing number of different fluorescent proteins, make FRET a sensitive tool necessary to obtain spatial and temporal distribution of protein association within the cellular context.

In addition FRET assay can be used to study competition between two proteins for binding to the same protein. This FRET-based competition assay has been previously used to characterize interaction of inhibitors with kinases [11, 12]. In this method, an unlabeled competitor may perturb the dynamic interaction frequencies between donor-labeled and acceptor-labeled proteins, reducing the FRET signals. This is because the competitor displaces the acceptor-labeled protein, separating the two fluorophores, which can no longer transfer energy to each other (Fig. 3). This FRET-based competition assay has several advantages including high sensitivity and reproducibility, and it is compatible with high-throughput studies, thereby making it a potential tool to screen compounds against therapeutic targets [13–15].

Herein, we give a detailed description of FRET microscopy imaging and data analysis. The first part of this protocol describes the sample preparation for FRET microscopy experiments of fixed NIH-3T3 cells, which can be similarly used for analyses on living cells. In the second part we describe the experimental setup to optimize the image acquisition. This is followed by a description of

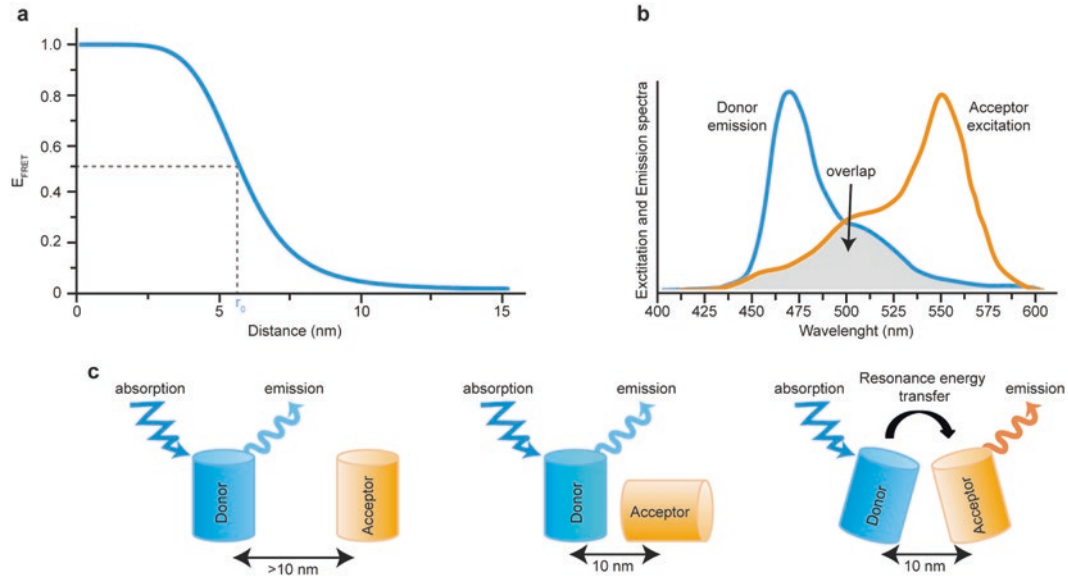


Fig. 2 Principles of fluorescence resonance energy transfer (FRET). (**a**) Diagram of the FRET efficiency as a function of the distance between a donor and an acceptor fluorophore (R_0). (**b**) FRET occurs only when the donor emission spectrum overlaps with the acceptor excitation spectrum. The *grey area* corresponds to the overlapped region. (**c**) Schematic representation of FRET. When the FRET donor and acceptor are more than 10 nm of one another (*left*) or are not correctly oriented (*middle*), then no FRET occurs and the donor emits fluorescence. If the donor and acceptor are within 2–10 nm of one another and correct oriented (*right*), excited donor transfer its energy by a nonradioactive process to the acceptor, causing it to emit fluorescence

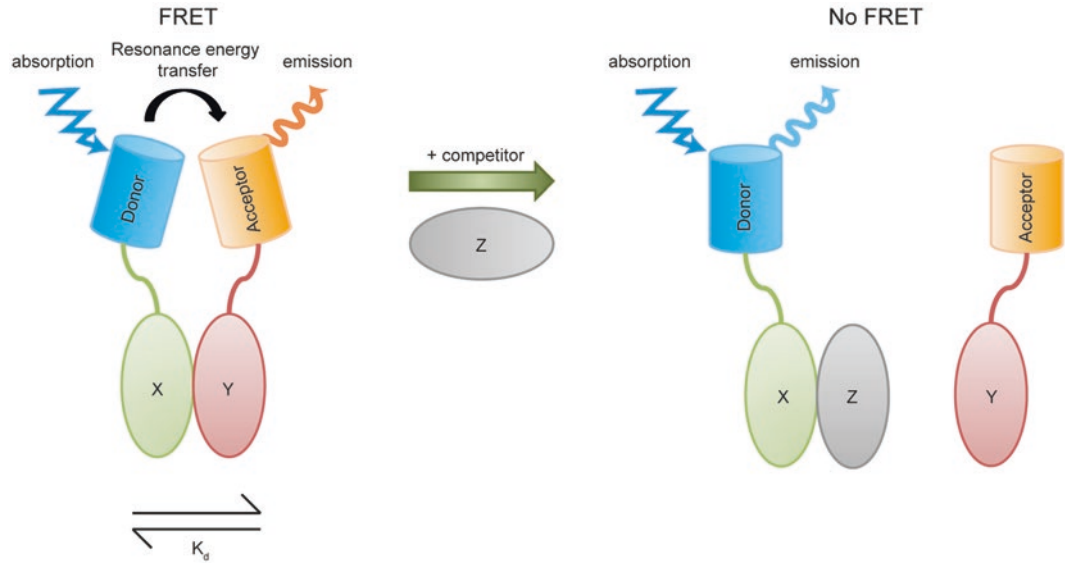


Fig. 3 Schematic representation of FRET-based competition assay. The scheme represents the basic principle of FRET competition assay. When an unlabeled target protein (Z) competes with the acceptor-labeled protein (Y) for binding with donor-labeled protein (X), the FRET efficiency may decrease or not occur

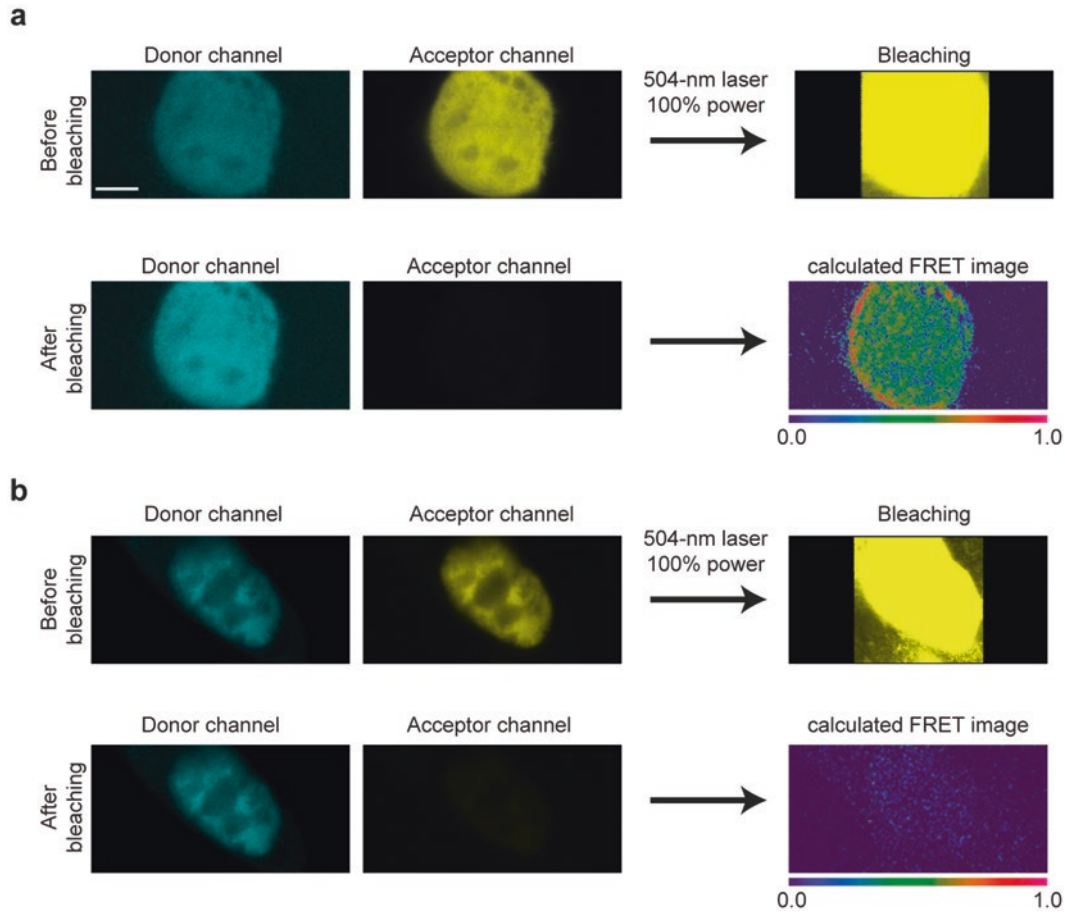


Fig. 4 An example of FRET AB microscopy images. This figure shows donor (CFP) and acceptor (YFP) fluorescent intensity images (before and after bleaching). Bleaching is performed using 514-nm laser line with the 100% of power for four bleach iterations. FRET efficiency is calculated according to equation described in Subheading 3.3, **step 7** and is presented as a color-coded map of FRET intensity. (**a**) FRET analysis of CFP fused to YFP (positive control) transiently expressed in NIH/3T3 cells. (**b**) FRET analysis of CFP-EED mixed with unconjugated YFP (negative control) transiently expressed in NIH/3T3 cells. Scale bar equals 5 μm

data analysis procedures for the determination of FRET efficiency (Fig. 4). In the last part, we describe methodologies to perform FRET-based competition analyses.

2 Materials

1. Tissue culture dishes, six-well.
2. Cover glass: 25 mm diameter.
3. 0.1% v/v gelatin solution in PBS. Keep at 4 °C.
4. NIH/3T3 cells.

5. Growth medium: DMEM supplemented with 10% v/v FBS, 1× L-glutamine, 1× nonessential amino acids, and 1× penicillin/streptomycin. The medium can be stored for up to 1 month at 4 °C.
6. Plasmid vectors: CFP-protein#1 and YFP-protein#2 (to test for FRET between protein#1 and protein#2), CFP–YFP fusion, unconjugated CFP and unconjugated YFP.
7. Lipofectamine 2000.
8. 4% paraformaldehyde solution. Filter the solution through a 0.45 μm filter prior to use.
9. Forceps.
10. Leica TCS SP5 confocal microscope.
11. 63×/1.4 NA oil immersion objective.
12. Coverslip holder: Autofluor cell chamber.
13. Leica SP5 software.

3 Methods

3.1 Cell Culture Transfection

1. Plate NIH-3T3 cells 24 h prior to transfection: remove growth medium from a 100 mm cell culture dish and rinse cells once with 5 ml of PBS, swirling gently (*see* **Note 1**).
2. Add 2 ml of 0.05% trypsin–EDTA to the cells and incubate at 37 °C for 2–3 min or until cells are detached from the culture dish, then add 4 ml of growth medium to neutralize trypsin digestion.
3. Count and seed 600,000 cells into a single well of a six-well plate containing a gelatin-coated coverslip, for each transfection condition.
4. Prepare the transfection mix using a ratio of 2.5 μl of Lipofectamine 2000/1 μg of total DNA. Transfect the cells with 4 μg of total DNA with the following constructs: (a) CFP fused to YFP (positive control, *see* **Note 2**); (b) CFP-protein#1; (c) YFP-protein#2; (d) CFP-protein#1, mixed with unconjugated YFP, and YFP-protein#2, mixed with unconjugated CFP (negative control, *see* **Note 3**); (e) CFP-protein#1 mixed with YFP-protein#2 (*see* **Notes 4**–**7**). For each transfection, the DNA and Lipofectamine 2000 are separately diluted into 500 μl of Opti-MEM in a 1.5 ml plastic micro tube, then combine Lipofectamine 2000 with DNA and mix it by vortexing.
5. Incubate the mixture for 20 min at room temperature.
6. Remove the growth medium from each well and wash gently with Opti-MEM medium. Add gently DNA–Lipofectamine 2000 mixture to each well (*see* **Note 8**). Add 1 ml of growth medium after 4 h and incubate the cells at 37 °C for 24 h.

7. Fix the cells by adding 4% paraformaldehyde and incubate for 20 min at room temperature. Wash the coverslips three times with PBS. The samples can be stored at 4 °C in the dark for 2 weeks in PBS before continuing (*see* **Note 9**).

3.2 Experimental Conditions Setting

1. Prepare the TCS SP5 confocal microscope (or comparable system) with a 40-mW argon laser set at 30% efficiency and tuned to lines at 458 and 514 nm.
2. Select FRET AB tool of SP5 software (Leica) and configure the acquisition parameters to 8-bit images, 512×512 pixels such that the region for FRET measurement occupies ~70% of the image, sequential scanning and pinhole to 1 Airy unit.
3. Remove the coverslip from the wells using forceps and place it into coverslip holder. Fill the holder with PBS.
4. Select a high-resolution immersion objective, such as 63×/1.4 NA, and apply immersion medium.
5. Start by acquiring cells containing only CFP-tagged protein. Find some fluorescent cells using microscope ocular and center them in the middle of the field of view.
6. Switch to the confocal mode, then adjust the laser transmission at 30% power for 458-nm laser and collect CFP emission from 465 to 495 nm. Set the image acquisition parameters to a maximum offset of 20, detector gain of 750, digital gain of 1 and line averaging of 4. Acquire an image of the CFP emission from a single cell and save the imaging parameters (*see* **Note 10**).
7. Repeat Subheading 3.2, **steps 5 and 6** with cells containing only YFP-tagged protein. Set the laser transmission at 30% power for 514-nm laser line and collect YFP emission from 555 to 630 nm (*see* **Note 11**).
8. Save the adjusted imaging parameters and do not change them in subsequent analyses.

3.3 Perform the Acceptor Photobleaching FRET Analysis

FRET Acceptor photobleaching involves measuring of the variation of the donor fluorescence in the presence of an acceptor. If FRET occurs, you will measure an increase of donor fluorescent intensity after photobleaching of the acceptor, as shown in Fig. 4.

1. Open the Edit Bleach window on the TCS SP5 to define the bleaching conditions.
2. Set up bleaching with a 514-nm laser line with the 100% of power (*see* **Note 12**).
3. Define the region of interest (ROI) and plot the ROI into the cell image where the photobleaching should occur (*see* **Note 13**).
4. To enable sufficient bleaching of the YFP-tagged protein, set the number of bleaching iteration at 4.

5. Run the bleaching experiment. The software acquires five images in a precise sequence: images of donor and acceptor before bleaching, followed by the bleaching of the acceptor and images of donor and acceptor after bleaching (Fig. 4).
6. The result in the ROI is automatically displayed and corresponds to your entire bleached area (Fig. 4). If necessary, it is possible to select additional ROI for better interpretation of results.
7. Apparent energy transfer efficiency is calculated as follow:

$$\mathrm{FRET}_{\mathrm{eff}} = \frac{D_{\mathrm{post}} - D_{\mathrm{pre}}}{D_{\mathrm{post}}}$$

   Here, D_{post} and D_{pre} refer to the intensity of the donor in a region of interest before and after the selective photobleaching of the acceptor. The $\mathrm{FRET}_{\mathrm{eff}}$ is considered positive when $D_{\mathrm{post}} > D_{\mathrm{pre}}$.
8. Repeat Subheading 3.3, **steps** from **3** to **6** for at least 30 individual cells.

3.4 Perform the Acceptor Photobleaching FRET Competition Analysis

In order to analyze whether two proteins or two isoforms of the same protein may compete for interaction with the protein of interest, a FRET competition analysis can be performed. In this method, FRET donor is bound to FRET acceptor and if the two fluorophores are in proximity with the proper orientation, FRET will occur between them. Displacement of acceptor with an unlabeled competitor can reduce FRET signal, indicating that the two proteins compete for the binding with FRET donor. In order to evaluate the FRET competition results, is appropriate to use different molar ratio between acceptor-labeled protein and the unlabeled competitor (e.g., 1:1, 1:4, and 1:8). In the case in which the investigated protein compete for their association with the protein of interest, the FRET efficiency should decrease proportionally to the increment of the acceptor–competitor ratio.

1. Transfect the cells with 4 μg of total DNA with the following constructs: (a) CFP fused to YFP (positive control); (b) CFP-protein#1, mixed with unconjugated YFP, and YFP-protein#2, mixed with unconjugated CFP (negative control); (c) CFP-protein#1 mixed with YFP-protein#2 and unconjugated protein#3 using different ratios; (d) CFP-protein#1 mixed with YFP-protein#3 and unconjugated protein#2 using different ratios (*see* **Note 14**).
2. Repeat steps from Subheading 3.2 to Subheading 3.3, to measure FRET efficiency. In case of competition between protein#2 and protein#3 for interaction with protein#1, reduction of FRET efficiency respect to the steady state will be measured.

4 Notes

1. This protocol is optimized for fixed NIH/3T3 cells. With appropriate modifications (e.g., coating coverslips, cell number, and quantity of transfected DNA), it may be used for other cell lines or using living cells. In the latter case, it would be necessary to keep cells within an humidified chamber with controlled temperature, thus avoiding perturbation of the cellular conditions.
2. It is appropriate to examine the setting and the maximum FRET efficiency using the positive control containing the two fluorophores connected with a short peptide linker (CFP–YFP).
3. It is appropriate to analyze the negative control: CFP-protein#1, mixed with unconjugated YFP, and YFP-protein#2, mixed with unconjugated CFP, which should not give FRET. This excludes that the fluorescent proteins affect the interaction.
4. Donor–acceptor choice is crucial for FRET. In fact, the FRET AB requires that the donor is a stable fluorescent protein, while the acceptor can be easily bleached during experimental conditions. Moreover, FRET efficiency dependent on both the distance between donor and acceptor molecules, and also on the overlaps between the donor molecule emission and acceptor molecule excitation spectra. The described protocol is optimized for CFP and YFP fluorescent proteins; however, other donor–acceptor pairs follow the same criteria (e.g., Cy3-Cy5 or BFP-GFP).
5. Before proceeding with FRET analysis it is appropriate to find the correct conditions of transfection since the ratio between the donor and acceptor strongly influences the outcome of FRET assays. Define experimental conditions in which the CFP and YFP-fusion protein are expressed at approximately equal amounts by performing immunoblot analyses of the transfected cells. Once you have optimized your experimental conditions, perform FRET assay putting particular attention on analyzing those cells which are expressing similar level of the two fluorescent protein, by comparing their relative intensity to the emission of the CFP–YFP positive control.
6. Certain proteins might cluster into aggregates making fluorescent dots that strongly affect the FRET analyses. In these cases, it would be necessary to tune the level of expression of the protein of interest and/or to co-transfect non-conjugated components of the investigated complex, thereby facilitating the proper organization of soluble form of the chromatin complex.

7. The effect of the N- or C-terminal fusion of the fluorescent protein to the protein of interest in terms of its localization and FRET efficiency should be carefully evaluated. Specifically, the tagged fluorescent proteins may cause steric effects and/or changing protein complex organization.
8. It is appropriate to work gently in order to avoid detachment of the cells from coverslips.
9. Even if the samples can be stored at 4 °C in the dark for 2 weeks, it is necessary to perform the analyses as soon as possible in order to avoid a reduction in the fluorescence intensity. Alternatively, after cell fixation the coverslips can be mount on a microscope slides and stored at −20 °C.
10. To exclude that CFP excitation can interfere with the excitement of YFP, excite the CFP and acquire an image in both the CFP and YFP detection channels. In case of high background signal in YFP image, adjust the setting for the CFP excitation.
11. To exclude that YFP excitation can interfere with the excitement of CFP, excite the YFP and acquire an image in both the CFP and YFP detection channels. In case of high background signal in CFP image, adjust the setting for the YFP excitation.
12. In case the donor will result bleached, decrease the power of the 514-nm laser reducing the strength of the photobleaching. Otherwise, decrease the scan speed or reduce the number of bleaching iterations.
13. If FRET takes place in a precise region of the cell, the ROI can be drawn on that area. However, in order to obtain reliable measures of intensity another ROI should be drawn in another compartment of the analyzed cell and the retrieved intensity should be use to determine the relative background signal.
14. Even for this experimental condition, test the proper ratio between the donor, acceptor, and competitor. In order to maintain the ratio of transfection, replace unconjugated competitor with empty vector in steady state condition.

References

1. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli D (2007) Genome regulation by polycomb and trithorax proteins. Cell 128(4):735–745
2. Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE (2002) The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. Mol Cell Biol 22(17): 6070–6078
3. Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V (2002) Drosophila enhancer of Zeste (ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111(2): 185–196
4. Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E (2012) Molecular architecture of human polycomb repressive complex 2. Elife 1:e00005

5. Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW, Koseki H, Brockdorff N, Ponting CP, Kessler BM, Klose RJ (2012) KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife 1:e00205
6. Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell 45(3):344–356
7. Sun Y, Wallrabe H, Seo SA, Periasamy A (2011) FRET microscopy in 2010: the legacy of Theodor Förster on the 100th anniversary of his birth. ChemPhysChem 12(3):462–474
8. Bastiaens PI, Pepperkok R (2000) Observing proteins in their natural habitat: the living cell. Trends Biochem Sci 25(12):631–637
9. Selvin PR (2000) The renaissance of fluorescence resonance energy transfer. Nat Struct Biol 7(9):730–734
10. Chen Y, Mills JD, Periasamy A (2003) Protein localization in living cells and tissues using FRET and FLIM. Differentiation 71(9–10): 528–541
11. Zhang WX, Wang R, Wisniewski D, Marcy AI, LoGrasso P, Lisnock JM, Cummings RT, Thompson JE (2005) Time-resolved Forster resonance energy transfer assays for the binding of nucleotide and protein substrates to p38alpha protein kinase. Anal Biochem 343(1):76–83
12. Tirat A, Freuler F, Stettler T, Mayr LM, Leder L (2006) Evaluation of two novel tag-based labelling technologies for site-specific modification of proteins. Int J Biol Macromol 39(1–3):66–76
13. Sharma B, Deo SK, Bachas LG, Daunert S (2005) Competitive binding assay using fluorescence resonance energy transfer for the identification of calmodulin antagonists. Bioconjug Chem 16(5):1257–1263
14. Tian H, Ip L, Luo H, Chang DC, Luo KQ (2007) A high throughput drug screen based on fluorescence resonance energy transfer (FRET) for anticancer activity of compounds from herbal medicine. Br J Pharmacol 150(3):321–334
15. Rogers MS, Cryan LM, Habeshian KA, Bazinet L, Caldwell TP, Ackroyd PC, Christensen KA (2012) A FRET-based high throughput screening assay to identify inhibitors of anthrax protective antigen binding to capillary morphogenesis gene 2 protein. PLoS One 7(6):e39911

Chapter 14

Analysis of Endogenous Protein Interactions of Polycomb Group of Proteins in Mouse Embryonic Stem Cells

Lluis Morey* **and Luciano Di Croce**

Abstract

PRC1 complexes contain four core subunits: Pcgf, Phc, Ring1, and Cbx proteins. Interestingly, mammalian genomes have several paralogues for each subunit, which are differentially expressed depending on the cell type, differentiation program, and cellular stimuli. Therefore, identification and characterization of the specific architecture of different PRC1 complexes during cellular differentiation are essential to better understand the function and recruitment mechanism of PRC1 complexes. In this chapter we describe several methods to study Polycomb architecture, and identification of novel interactors in both pluripotent and differentiating mouse embryonic stem cells.

Key words Polycomb, Immunoprecipitation, Pluripotent stem cells, Differentiation

1 Introduction

Polycomb group of protein complexes are epigenetic factors essential for stem cell differentiation, cell fate commitment, development, and cancer [1–6]. They are classified into two main complexes, the Polycomb-repressive complexes 1 (PRC1) and 2 (PRC2) [7]. The PRC2 complex comprises three core subunits: EED, Suz12, and the histone methyltransferases Ezh1/2, which deposit di- and tri-methylation on lysine 27 of the histone H3 (H3K27me2/3) [8]. Depending on its subunit composition, the PRC1 complex can be classified into two main types of complexes, the canonical PRC1 (cPRC1) and the noncanonical PRC1 (ncPRC1). cPRC1 specifically contains Cbx proteins, Pcgf2/4/6, Phc1/2/3, and the E3-ligases Ring1A/B, which monoubiquitinate lysine 119 at histone H2A (H2AK119ub). ncPRC1 complexes lack a Cbx protein and are mainly characterized by the presence of RYBP/YAF2, Pcgf1/3/5, Fbxl10, and Ring1A/B [9–13]. This large heterogeneity in the architecture of the PRC1 complexes (Fig. 1) suggests that

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_14,

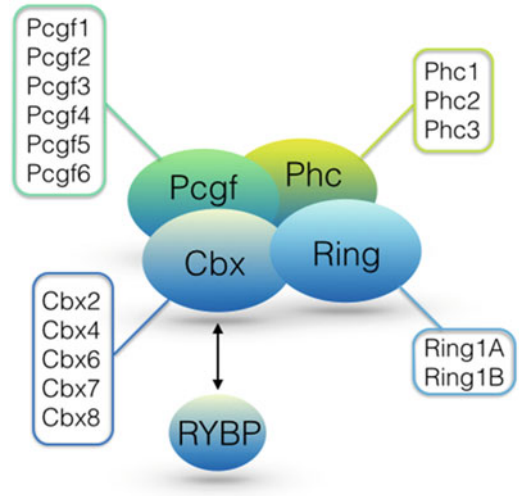


Fig. 1 PRC1 complex subunits. Core PRC1 complexes contain Pcgf, Cbx, Phc, and Ring proteins. Each subunit has several paralogues, and cPRC1 and ncPRC1 complexes contain either Cbx or RYBP proteins, respectively

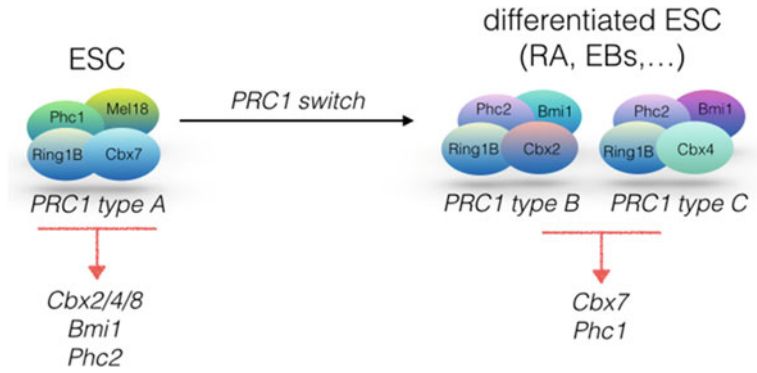


Fig. 2 Different PRC1 complexes are assembled in pluripotent and differentiating ESCs. In pluripotent ESCs the cPRC1 contains Cbx7, Mel18, Phc1, and Ring1B. This complex directly represses the expression of other PRC1 subunits that during differentiation are de-repressed and assembled into different PRC1 complexes depending on the cellular type

each of these complexes has specific recruitment mechanisms, and therefore gene target selectivity, and probably enzymatic activities. Recently, we have shown that specific PRC1 complexes are assembled in pluripotent and differentiating cells (Fig. 2), resulting in specific PRC1-mediated biological functions [2, 9].

Here we report some of the methods we currently apply to characterize the composition of PRC1 complexes in mouse embryonic stem cells (ESCs) and differentiating ESCs.

2 Materials

2.1 Cells

Mouse embryonic stem cells (strain E14Tg2a).

2.2 Media Components and Reagents

E14Tg2a ESCs can be cultured feeder free in media containing serum supplemented with the leukemia inhibitor factor (LIF) or in media containing GSK and MEK inhibitors supplemented with LIF (2i + LIF).

2.2.1 ESC Cultured in Serum + LIF

1. Gelatin 0.1% (EmbryoMax, Millipore).
2. Glasgow minimum essential medium (Sigma).
3. Fetal bovine serum—FBS (HyClone).
4. Sodium pyruvate (Gibco).
5. Nonessential amino acids-MEM (NEAA) (Gibco).
6. Glutamax (Gibco).
7. β-Mercaptoethanol (Gibco).
8. Antibiotics (Pen/Strep, Gibco).
9. Leukemia inhibitor factor—LIF (Millipore).
10. Trypsin.

2.2.2 ESC Cultured in 2i + LIF

1. PD0325901 (Selleck).
2. CHIR-99021 (Selleck).
3. DMEM/F12 medium (Gibco).
4. Neurobasal medium (Gibco).
5. N2 supplement (Gibco).
6. B-27 supplement (Gibco).
7. Glutamax.
8. Nonessential amino acids-MEM (NEAA).
9. β-Mercaptoethanol.
10. 30% Bovine serum albumin—BSA (Gemini).
11. LIF (ESGRO mouse LIF medium, Millipore).
12. Accutase solution (Sigma).

2.3 Buffer Components and Reagents

2.3.1 For Cell Lysis, Sonication, Immunoprecipitation (Ip) Washes, and Protein Quantification

1. 1 M Tris pH 7.5.
2. 5 M NaCl.
3. Glycerol.
4. NP-40.
5. Complete EDTA-free tablets (Roche).
6. Bradford.

2.3.2 For IPs and Western Blot Assays

1. Protein A and G sepharose beads (GEHealthcare).
2. Primary antibodies (*see* Table 1).
3. Secondary antibodies for western blot.
4. TBST 1× buffer.
5. Laemmli sample buffer.

2.3.3 For Protein Elution After IPs

1. BS3 (Thermo Scientific).
2. UREA.
3. ABC.

2.3.4 For Silver Staining of SDS-PAGE Gels

1. Precast gels NuPage 4–12% Bis–Tris Gel (Novex NP).
2. SilverQuest Staining Kit (Invitrogen).

2.4 Equipment

1. Cell culture hood (i.e., biosafety cabinet).
2. Inverted microscope with 4× and 10× objectives.
3. Incubator set at 37 °C, 5% CO_2.
4. Water bath set at 37 °C.
5. Refrigerated centrifuge.
6. Micropipettes (1–10, 2–20, 20–200, 200–1000 μl).
7. Pipettor.

Table 1
Antibodies to study PRC1 proteins by IP and western blot

Antibody	WB dilution	IP	Supplier
Ring1B	1:2000	5 μg	MBL (D139-3)
Ring1A	1:1000		Millipore (09-706)
Cbx2	1:2000	5 μg	Bethyl (A302-524A)
Cbx4	1:2000	5 μg	Sigma (HPA008228)
Cbx7	1:500	5 μg	Abcam (ab21873)
Phc1	1:2000	5 μg	Active Motif (39723)
Phc2	1:1000		Active Motif
Phc3	1:1000		Santa Cruz (sc99586)
Pcgf1/NsPC1	1:1000		Abcam (ab183499)
Pcgf2/Mel18	1:2000	5 μg	Santa Cruz (sc-10744)
Pcgf4/Bmi1	1:2000	5 μg	Active Motif (39993)
RYBP	1:2000	5 μg	Millipore (AB3637)
IgG		5 μg	Abcam

8. Freezers: −20 and −80 °C.
9. Complete electrophoresis and western blot apparatus.
10. Bioruptor.
11. Branson sonifier.
12. Rotating shaker.
13. Thermo-block.

2.5 Disposables

1. Sterile plastic pipettes (5, 10 mL).
2. 15 and 50 ml conical tubes.
3. 10 cm Tissue culture-treated dishes.
4. Filter pipette tips (0.5–10, 2–20, 20–200, 200–1000 μl).
5. Glass Pasteur pipettes, 9 in., sterilized by autoclave.
6. 1.5 ml Eppendorfs.
7. 1 ml Syringes.

2.6 Solutions

To prepare 600 ml of serum + LIF medium mix the following:

2.6.1 Serum + LIF Medium

Glasgow Minimum Essential Medium	474.5 ml
FBS	100 ml (20%)
Sodium pyruvate	6 ml
MEM NEAA	6 ml
Glutamax	6 ml
β-Mercaptoethanol	1.5 ml
Antibiotics (Pen/Strep)	6 ml

Add LIF (1 μl in 10 ml of medium) to the medium that will be used the same day.

Stock solution:

Leukemia inhibitor factor	10 μg
Deionized water	1 ml

2.6.2 2i + LIF Medium

To prepare 516 ml of 2i stock medium mix the following:

DMEM/F12 medium	250 ml
Neurobasal medium	250 ml
N2-supplement	2.5 ml
B27-supplement	5 ml
Glutamax	5 ml
MEM NEAA	5 ml
β-Mercaptoethanol	0.5 ml
BSA	0.5%

To prepare 50 ml of complete medium mix the following:

LIF 10 μg/ml	100 μl
10 mM PD	5 μl
10 mM CHIR	15 μl

2.6.3 Embryoid Body (EB) Medium

The medium to differentiate ESCs into EBs contains 10% of serum, instead of 20%, and is not supplemented with LIF.

2.6.4 ESC Medium Containing Retinoic Acid (ATRA)

The medium to differentiate the ESCs with ATRA contains 10% of serum, instead of 20%, no LIF, and 1 μM of ATRA dissolved in ethanol (stock 10 μM).

2.6.5 Lysis, Sonication, IP, and Washing Buffer (IP300) (see Note 1)

To prepare 500 ml of IP300 buffer mix the following:

1 M Tris pH 7.5	25 ml (final concentration 50 mM)
5 M NaCl	30 ml (final concentration 300 mM)
10% Glycerol	5 ml (final concentration 10%)
10% NP-40	10 ml (final concentration 0.2%)
Deionized water	430 ml

Keep the buffer at 4 °C.

2.6.6 BS3-Cross-Linking Buffer

1. Conjugation buffer: 20 mM Hepes pH 8.0 + 150 mM NaCl. Keep at room temperature (RT).
2. 100 mM BS3 stock: 8 mg of BS3 in 140 μl of conjugation buffer. Prepare it fresh.
3. For each IP use 250 μl of BS3-cross-linking buffer at 5 mM take 12.5 μl of BS3 stock + 237.5 μl of conjugation buffer.

2.6.7 Elution Buffer with Glycine

Prepare glycine 0.1 M pH 2.5 and store it at RT.

2.6.8 Elution Buffer with UREA + ABC Buffer

Prepare fresh 6 M urea and 200 mM ABC in deionized water.

2.7 Preparation of Gelatin-Coated Dishes

1. Prepare gelatinized plates by covering the entire dish surface with a 0.1% gelatin solution and incubate for 15 min at 37 °C. For a 10 cm dish 4–5 ml of gelatin is enough.
2. Completely remove the gelatin solution from the dishes. Gelatinized plates should be used immediately.

3 Methods

3.1 Culture of mESCs

3.1.1 Culture of mESCs in Serum + LIF

1. Gelatinize a 10 cm plate with 5 ml of gelatin 0.1%. Keep the plate at 37 °C for 15 min. Remove gelatin and add 10 ml of ESC medium containing serum and LIF.
2. In a bath at 37 °C thaw a cryo-tube containing 2×10^6 of cells, wash cells once with PBS1×, and culture them in a 10 cm dish with 10 ml of media in a gelatin-coated dish.
3. Change media after 24 h.
4. Next day trypsinize cells for 2–3 min at 37 °C, and seed 2×10^6 cells in a new 10 cm gelatin-coated dish.
5. Split cells (2×10^6 cells for a 10 cm plate) every 48 h.
6. To perform endogenous co-IPs, prepare 2 × 10 cm plates of cells. Around 5–6 mg of proteins will be extracted following the lysis method described in Subheading 3.4.

3.1.2 Culture of mESCs in 2i + LIF

ESCs need to adapt to 2i + LIF culture conditions, and are ready to be used after 8–10 passages.

1. Gelatinize a 10 cm plate with 5 ml of gelatin 0.1%. Keep the plate at 37 °C for 15 min. Remove gelatin and add 10 ml of complete *2i + LIF medium*.
2. Remove medium from the plate and wash the cells once with PBS1×.
3. Dissociate cells using 1 ml of Accutase instead of trypsin.
4. Collect cells from the plate with 9 ml of PBX 1×.
5. Centrifuge at 1200 rpm for 5 min.
6. Resuspend the cells in *2i + LIF medium*.
7. Count the cells.
8. Seed 2×10^6 cell in a 10 cm plate.

3.2 Differentiation of mESCs with ATRA

To study either new Polcyomb interactors or newly assembled Polycomb complexes during RA-mediated differentiation, culture the ESCs in *ESC medium containing retinoic acid (ATRA)*. To differentiate ESC with RA culture the cells with serum + LIF and after 24 h change to the *ESC medium containing retinoic acid (ATRA)* for 2 days. Change media after 48 h.

RA induces cell differentiation but also cell death; therefore prepare 3–4 10 cm plates.

3.3 Differentiation of mESCs into EBs

We also observed that Polycomb architecture changes when ESCs are differentiated into EBs. To generate uniform-sized EBs use the hanging drop method [14]:

1. Trypsinize ESCs, wash once with PBS 1×, and prepare *embryoid body (EB) medium*.
2. Resuspend 2×10^6 ESCs in 40 ml *EB medium*.
3. In a lid of a 150 cm plate, and using a multichannel pipette, place drops of 20 μl containing 1000 ESCs and incubate cells at 37 °C for 48 h. Use as many lids as needed to make all 20 μl drops from the 40 ml.
4. Flush out the drops from lid using a 10 ml pipette and *EB medium*. Usually one drop contains one EBs.
5. Check under the microscope the EBs. They should appear round and uniform sized.
6. Transfer the EBs into a 50 ml conical tube.
7. Let the EBs to precipitate by gravity to the bottom of the conical tube. Do not centrifuge, as EBs will disaggregate.
8. Carefully collect EBs with a 10 ml pipette containing *EB medium*.
9. Transfer EBs into four non-treated 10 cm plates.
10. Change medium every 2 days (repeat **steps 6–9**).
11. Collect EBs at days 4, 6, and/or 10.
12. Wash the EBs with PBS 1×.
13. Freeze them at −20 °C or directly use them for IPs.

3.4 Protein Extraction

As mentioned above, 5–6 mg of protein is extracted from two 10 cm plates of pluripotent ESCs at 80% confluency [15] (*see* **Note 2**). For IPs using differentiated ESCs with RA use four 10 cm plates to extract 5–6 mg of protein. To perform IPs from EB extracts generate at least 40 ml of EBs (*see* Subheading 3.3). Prepare lysates of pluripotent ESCs and differentiated ESCs following the same method:

1. Trypsinize cells.
2. Inhibit trypsin with media containing serum.

3. Centrifuge cells at 1200 rpm for 5 min at RT.
4. Wash cells with PBS 1×, and centrifuge at 1200 rpm for 5 min.
5. Discard supernatant.
6. Keep cell pellet on ice.
7. For each 10 cm dish cell pellet add 1 ml of IP300 buffer (*see* **Note 1**) supplemented with protease inhibitors (Complete EDTA-free tablets resuspended in deionized water).
8. Gently resuspend pellet until DNA is extracted (a viscous, mucus-like substance will appear).
9. Keep the cell lysate on ice for 10 min.
10. Sonicate cells: If using a Bioruptor sonicate for 5 min at maximum amplitude (30″ on/off cycles); if using a Branson sonifier sonicate at 10% output two times for 10 s.
11. Centrifuge for 15 min at 12,000 rpm/4 °C.
12. Collect supernatant and discard pellet.
13. Quantify protein using a Bradford solution.
14. Use the sample directly for IP. Do not freeze the lysate.

3.5 IPs

1. Reserve 1 mg of protein for each IP assay.
2. Take 10% of IP material as input sample and keep it on ice.
3. Add 4× Laemmli buffer to a final concentration of 2×.
4. Denature input sample for 10 min in a thermo-block at 100 °C.
5. Store input samples at −20 °C.
6. For each IP use 1 mg of cell extract and 5 μg of an antibody (Ab) against the protein of interest. Include a control sample with 5 μg of an IgG Ab.
7. Incubate protein extract/Ab overnight at 4 °C in a rotating shaker.
8. Wash/equilibrate protein A or G beads three times with 1 ml of IP300 buffer.
9. Add 30 μl of equilibrated beads to each IP.
10. In a rotating shaker incubate the beads with the protein extract/Ab for 2 h.
11. Spin down the bead-containing immunocomplexes at 2000 rpm for 2 min in a refrigerated centrifuge.
12. Wash the bead-containing immunocomplexes with 1 ml of IP300 by turning up side down the Eppendorf ten times.
13. Centrifuge at 2000 rpm for 2 min in a refrigerated centrifuge.
14. Repeat **steps 12** and **13** three times.
15. Carefully remove buffer from the last wash with a 1 ml syringe.
16. Samples are ready for protein elution (*see* Subheading 3.6.1).

3.5.1 IPs in the Presence of Ethidium Bromide (EtBr)

To investigate whether a protein directly interacts with the PRC1 complex or in contrast the interaction is DNA mediated co-IP assays should include EtBr [16]:

1. After cellular sonication add 100 μg/ml of EtBr to the protein extract.
2. Incubate for 30 min at 4 °C in a rotating shaker.
3. Centrifuge for 15 min at 12,000 rpm/4 °C.
4. Collect supernatant and discard pellet.
5. Proceed with the IP as in Subheading 3.5.
6. During IP washes add 100 μg/ml of EtBr to the IP300 buffer.
7. Proceed as in Subheading 3.5.

3.5.2 IPs Using Cross-Linked Antibodies to the Beads

If endogenous co-IPs will be subjected to mass-spectrometry analysis it is recommended to cross-link the antibody to the beads to avoid immunoglobulins in the eluted material.

To cross-link the Ab to the beads:

1. Prepare 50 μl of IP300 washed beads for each IP (including IgG as a control).
2. Couple 10 μg of Ab to 50 μl of beads in 500 μl of IP300 buffer for 2 h in a rotating shaker at 4 °C.
3. Prepare BS3 stock solution (*see* Subheading 2.6.6).
4. Wash beads/Ab three times with 1 ml of conjugation buffer.
5. Incubate 250 μl of 5 mM BS3 solution with beads/Ab for 1 h at RT in a rotating shaker.
6. Quench the cross-linking reaction with 12.5 μl of 1 M Tris pH 7.5 (final concentration 50 mM) for 15 min at RT in a rotating shaker.
7. Incubate cross-linked Ab/beads with 1 mg of cell lysate prepared as in Subheading 3.4.
8. Perform IPs and washes as in Subheading 3.5.
9. Elute immunocomplexes with glycine pH 2.5 or urea+ABC (*see* Subheadings 3.6.2.1 and 3.6.2.2)

3.6 Elution of Immunoprecipitated Proteins

After the last wash of the beads containing the immunoprecipitated material:

3.6.1 Laemmli Elution

1. Remove wash buffer with a 1 ml syringe.
2. Add 60 μl 2× Laemmli buffer and transfer the Eppendorfs to a thermo-block at 100 °C for 10 min.
3. Spin down the beads at 12,000 rpm for 1 min.
4. Take supernatant containing the eluted material and transfer to a new Eppendorf.
5. Load 1/3 of the eluted material into a SDS-PAGE gel for western blot analysis including the Input samples.

3.6.2 Elution of Immunocomplexes Recovered Using Cross-Linked Antibodies

Using Acid Buffer

1. After the last wash with the IP300 buffer carefully remove all buffer with a 1 ml syringe.
2. Add 1:1 volume of 0.1 M glycine pH 2.5 to the beads.
3. Place the sample containing the beads + glycine in a thermoblock shacking at 1000 rpm for 5 min at RT.
4. Centrifuge at 5000 rpm for 2 min.
5. Transfer the supernatant to a new Eppendorf.
6. Add 1:1 volume of 1 M Tris pH 8 to neutralize the pH.
7. Repeat **steps 2–5** three times, but perform the second and third elution for 30 min and the last one for 5 min.
8. Combine all four eluates in one Eppendorf.
9. If eluted material will be subjected to mass spectrometry (*see* **Note 3**) store it at −20 °C.
10. If eluted material will be used for western blot analysis dilute it 1:1 with 4× Laemmli buffer and load 1/3 of the material into an SDS-PAGE gel.

Using urea + ABC Buffer

Urea and ABC can be used to elute immunocomplexes from the beads.

1. Prepare urea and ABC solutions prior to use (*see* Subheading 2.6.8).
2. Add 1:1 volume of urea + ABC solution into the beads.
3. In a shaker at 1000 rpm incubate the beads + urea + ABC solution at RT for 30 min.
4. Spin down the beads at 2000 rpm for 3 min.
5. Collect eluates and discard the beads.
6. If eluted material will be subjected to mass spectrometry (*see* **Note 3**) store it at −20 °C.
7. If eluted material will be used for western blot analysis dilute it 1:1 with 4× Laemmli buffer and load 1/3 of the material into an SDS-PAGE gel.

4 Notes

1. The interactions of the core PRC1 subunits are stable in buffers containing up to 1 M NaCl; yet if novel interactors are studied, cell lysate and IPs can be performed in a buffer containing 150 mM NaCl (IP150) instead of 300 mM NaCl (IP300 buffer).
2. We successfully characterized PRC1 complexes by mass-spectrometry analysis using 2 mg of protein extracts; yet eluted proteins are not visible by silver staining of the SDS-PAGE gel.

To control the IP efficiency perform a western blot using 10% of the material prior to mass spectrometry. To visualize the eluted proteins by silver stain perform the IPs with 30 mg of protein. IPs using this large amount of material need to be performed using 100 µl of bead-slurry for 10 mg of protein. The amount of antibody needs to be tested by the researcher, as every antibody has different efficiency.

3. Before mass-spectrometry analysis it is highly recommended to carefully check the best digestion method prior to injecting the samples into the mass-spectrometer machine, as well as the amount of eluted protein needed to successfully identify the proteins of interest.

References

1. Morey L, Helin K (2010) Polycomb group protein-mediated repression of transcription. Trends Biochem Sci 35(6):323–332. doi:10.1016/j.tibs.2010.02.009, S0968-0004(10)00030-7 [pii]
2. Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, Di Croce L (2012) Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell 10(1):47–62. doi:10.1016/j.stem.2011.12.006
3. O'Loghlen A, Munoz-Cabello AM, Gaspar-Maia A, Wu HA, Banito A, Kunowska N, Racek T, Pemberton HN, Beolchi P, Lavial F, Masui O, Vermeulen M, Carroll T, Graumann J, Heard E, Dillon N, Azuara V, Snijders AP, Peters G, Bernstein E, Gil J (2012) MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. Cell Stem Cell 10(1):33–46. doi:10.1016/j.stem.2011.12.004
4. Li M, Liu GH, Izpisua Belmonte JC (2012) Navigating the epigenetic landscape of pluripotent stem cells. Nat Rev Mol Cell Biol 13(8):524–535. doi:10.1038/nrm3393
5. Laugesen A, Helin K (2014) Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell 14(6):735–751. doi:10.1016/j.stem.2014.05.006
6. Simon JA, Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 10(10):697–708. doi:10.1038/nrm2763, nrm2763 [pii]
7. Di Croce L, Helin K (2013) Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol 20(10):1147–1155. doi:10.1038/nsmb.2669
8. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298(5595):1039–1043. doi:10.1126/science.1076997, 1076997 [pii]
9. Morey L, Aloia L, Cozzuto L, Benitah SA, Di Croce L (2013) RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell Rep 3(1):60–69. doi:10.1016/j.celrep.2012.11.026
10. Wu X, Johansen JV, Helin K (2013) Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Mol Cell 49(6):1134–1146. doi:10.1016/j.molcel.2013.01.016
11. Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell 45(3):344–356. doi:10.1016/j.molcel.2012.01.002
12. Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW, Koseki H, Brockdorff N, Ponting CP, Kessler BM, Klose RJ (2012) KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife 1:e00205. doi:10.7554/eLife.00205
13. Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LL, Ito S, Cooper S, Kondo K, Koseki Y, Ishikura T, Long HK, Sheahan TW, Brockdorff N, Kessler BM, Koseki H, Klose RJ (2014) Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and

polycomb domain formation. Cell 157(6):1445–1459. doi:10.1016/j.cell.2014.05.004

14. Keller GM (1995) In vitro differentiation of embryonic stem cells. Curr Opin Cell Biol 7(6):862–869
15. Morey L, Santanach A, Blanco E, Aloia L, Nora EP, Bruneau BG, Di Croce L (2015) Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms. Cell Stem Cell 17(3):300–315. doi:10.1016/j.stem.2015.08.009
16. Lai JS, Herr W (1992) Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. Proc Natl Acad Sci U S A 89(15):6958–6962

Chapter 15

Determination of Polycomb Group of Protein Compartmentalization Through Chromatin Fractionation Procedure

Federica Marasca, Fabrizia Marullo, and Chiara Lanzuolo*

Abstract

Epigenetic mechanisms modulate and maintain the transcriptional state of the genome acting at various levels on chromatin. Emerging findings suggest that the position in the nuclear space and the cross talk between components of the nuclear architecture play a role in the regulation of epigenetic signatures. We recently described a cross talk between the Polycomb group of proteins (PcG) epigenetic repressors and the nuclear lamina. This interplay is important for the maintenance of transcriptional repression at muscle-specific genes and for the correct timing of muscle differentiation. To investigate the synergism between PcG factors and nuclear architecture we improved a chromatin fractionation protocol with the aim to analyze the PcG nuclear compartmentalization. We thus separated PcG proteins in different fractions depending on their solubility. We surprisingly found a consistent amount of PcG proteins in the matrix-associated fraction. In this chapter we describe the chromatin fractionation procedure, a method that can be used to study the nuclear compartmentalization of Polycomb group of proteins and/or PcG targets in murine and *Drosophila* cells.

Key words Polycomb, Chromatin fractionation, Nuclear compartmentalization, Lamin

1 Introduction

Gene expression is dependent on the cell type and orchestrated through different levels of regulation from interactions between DNA sequences and transcription factors to DNA methylation, histone modifications and chromatin remodeling, including the interplay with the nuclear architecture [1–3]. Among others, lamin proteins are intermediate filaments that, interacting with each other, give rise to a scaffold at the nuclear periphery of the nucleus (nuclear lamina) [4]. Interestingly lamins were not only found at the nuclear periphery, but also within the nucleoplasm where they exist as a detergent-soluble pool [5, 6]. Pioneering studies in *Drosophila melanogaster* by Bas van Steensel and colleagues identified the lamina-associated domains (LADs) as late replicative genomic regions,

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_15,

enriched in transcriptionally inactive genes and heterochromatic histone marks [7]. Then, growing evidences confirmed that, beyond the structural function of lamins in nucleus, these proteins are implicated in chromatin dynamics, coordinating gene expression during differentiation processes [8–11].

Polycomb group (PcG) of proteins are epigenetic regulators that exert their functions in multiprotein complexes called Polycomb repressive complexes (PRCs). PRCs promote gene repression on a large number of genes primarily involved in development and differentiation, contributing to the acquisition and the maintenance of cellular identity [12, 13]. In the nucleus, PcG proteins are organized in distinct microscopically visible aggregates called Polycomb bodies [14], where multiple PcG targets are clustered [13, 15].

Recently, we have shown that an intact nuclear Lamin A/C is necessary for the correct positioning of the PcG bodies and for their repressive functions, thus ensuring the correct timing of gene expression during muscle differentiation [11].

To unravel a possible interplay between lamins and PcG proteins, we studied PcG subnuclear localization both in chromatin and unsoluble, lamins soaked, nuclear matrix compartments using a biochemical technique called chromatin fractionation. Classic biochemical studies of nuclear fractionation have demonstrated that transcriptionally active chromatin is easily solubilized after nuclease treatment of isolated nuclei while the rest is tightly bound to the nuclear matrix [16–18]. The nuclear matrix was previously defined as the internal scaffold of the nuclei that remains after salt extraction and DNase treatment [19]. In the past, many different protocols have been developed in the attempt to obtain clean nuclear matrix to study its form and composition. For this purpose in 1974 Berezney and Coffey began to treat isolated nuclei with DNase I and 2 M NaCl [20]; in opposition, Mirkovitch and collaborator in 1984 devised a milder extraction method of nuclei using strong detergent as lithium diiodosalycilate (LIS) to remove digested chromatin, thus avoiding possible artifact due to salt precipitation [21]. Then Fey et al. (1986) and He et al. (1990) introduced whole-cell extraction with nonionic detergent, followed by a step of DNase chromatin digestion and ammonium sulfate extraction [22, 23]. Finally, the latest procedure had been employed by Reyes et al. (1997) that adopted the method to study the subnuclear compartmentalization of chromatin remodeler complex SWI/SNF, giving rise to a complete, four-step fractionation procedure that includes whole-cell nonionic detergent extraction, chromatin DNase digestion followed by ammonium sulfate extraction, 2 M NaCl extraction, and unsoluble pellet resuspension in 8 M of urea [24]. We adapted this protocol, adding some washes and improving the fraction purity. We found that PcG proteins are present in the chromatin fractions as well as in the nuclear matrix

fraction. This evidence and others lay the groundwork for a novel model that envisages a role for Lamin A/C protein in contributing to differentiation program regulation through reinforcing PcG protein-repressive functions [11].

2 Materials

2.1 Cells

C2C12 mouse skeletal myoblasts (ATCC, CRL-1772).

Drosophila embryonic S2 cells.

2.2 Cell Culture Media Components and Reagents

C2C12 are cultured as replicative myoblasts at low confluence (30–50%) in growth medium and can be differentiated in multinucleated, post-mitotic myotubes at higher cellular confluence, in differentiating medium.

D. melanogaster embryonic S2 cells are cultured at 25 °C in serum-free insect culture medium.

2.2.1 C2C12 Growth Medium

1. Dulbecco's modified Eagle's medium (DMEM) (4.5 g/L D-glucose, glutamax).
2. 10% Fetal bovine serum (FBS).
3. 100 I.U./mL Penicillin and 100 μg/mL streptomycin.

2.2.2 C2C12 Differentiating Medium

1. Dulbecco's modified Eagle's medium (DMEM) (4.5 g/L D-glucose, glutamax).
2. 2% Horse serum (HS).
3. 100 I.U./mL Penicillin and 100 μg/mL streptomycin.

2.2.3 S2 Cell Medium

1. Serum-free insect culture medium.
2. 100 I.U./mL Penicillin and 100 μg/mL streptomycin.

2.3 Buffer Components and Reagents

2.3.1 Protein Extraction from Chromatin Fractionation

1. 1× Phosphate-buffered saline (PBS).
2. 500 mM Pipes pH 6.8.
3. 5 M NaCl.
4. 1 M Sucrose.
5. 1 M $MgCl_2$.
6. 500 mM EGTA.
7. 10% Triton X-100 (*see* **Note 1**).
8. 8 M Urea.
9. 25× Complete EDTA-Free Protease Inhibitor Cocktail tablets (PIC).
10. 100× Phenylmethanesulfonyl fluoride (PMSF).
11. 1 M Dithiothreitol (DTT).

12. RNase-free DNase.
13. 4 M $(NH_4)_2SO_4$.

2.3.2 DNA Extraction from Chromatin Fractionation

1. 1 M Tris HCl pH 8.
2. 500 mM EDTA.
3. 10 mg/mL RNase A.
4. 20 mg/mL Proteinase K.
5. 24:24:1 Phenol:chlorophorm:isoamyl alcohol.
6. 98% Ethanol.
7. 3 M Sodium acetate pH 5.2.
8. Glycogen.
9. 70% Ethanol.
10. 10 mM Tris–HCl pH 7.5.

2.3.3 Protein Quantification and Western Blot Assays

1. Protein quantification assay kit.
2. 4× Sample loading buffer (8% sodium dodecyl sulfate (SDS), 40% glycerol, 240 mM Tris–HCl pH 6.8, 0.02% bromophenol blue, 0.4% 2-mercaptoethanol).
3. Antibodies (*see* Table 1).
4. PBS.
5. Tween-20.

2.3.4 DNA Quantification and qPCR Assays

1. DNA quantification assay kit.
2. SYBR Green master mix.
3. Locus-specific primers.

2.4 Equipment

1. Cell culture hood (i.e., biosafety cabinet).
2. Inverted microscope.
3. Incubator set at 37 °C, 5% CO_2.
4. Incubator set at 25 °C.
5. Micropipettes (1–10, 2–20, 20–200, 200–1000 μl).
6. Pipette aid.
7. Burker camera.
8. Refrigerated centrifuge.
9. Wheel at 4 °C.
10. Thermo-block.
11. Freezers: –20 and –80 °C.
12. Complete electrophoresis and western blot apparatus.
13. Oscillating shaker.
14. Real-time qPCR machine.

Table 1
List of antibodies used for western blot assay to study PRC1 and PRC2 protein cellular localization

Antibody	WB dilution	Supplier
Ez	1:2000	Home made, rabbit
Pc	1:2000	Home made, rabbit
LmnC	1:2000	Home made, mouse
LmnDm0	1:2000	Home made, mouse
β-Tubulin (E7)	1:2000	University of Iowa, Hybridoma Bank, Mouse
Ezh2	1:2000	Cell Signaling, 3147S, mouse
Bmi1	1:1000	Millipore, 05-637, mouse
YY1	1:2000	Santa Cruz, SC-281, rabbit
α-Tubulin	1:2000	Sigma, T5168, mouse
H3	1:10,000	Abcam, Ab1791, rabbit
Lamin B	1:5000	Santa Cruz, sc6216, goat
Lamin A/C	1:2000	Santa Cruz, sc6215, goat
Anti-rabbit HRP	1:2000	Sigma, A9169
Anti-mouse HRP	1:4000	Sigma, A9044
Anti-goat HRP	1:20,000	Sigma, A5420

2.5 Disposables

1. For adherent cells: 10 cm^2 tissue culture.
2. For *Drosophila* embryonic S2 cells 25 and 75 cm^2 flasks.
3. Sterile plastic pipettes (5, 10, 25 mL).
4. 15 mL and 50 mL conical tubes.
5. Filter pipette tips (0.5–10, 2–20, 20–200, 200–1000 μl).
6. 1.5 mL Eppendorf tubes.
7. 1 mL Syringes.

2.6 Solutions

2.6.1 Cytoskeleton (CSK) Triton Buffer

10 mM Pipes pH 6.8.
100 mM NaCl.
1 mM EGTA.
300 mM Sucrose.
3 mM $MgCl_2$.
1 mM DTT.
0.5% Triton X-100.
1× PMSF.

1× PIC.
ddH_2O.
Keep the buffer at 4 °C just before use.

2.6.2 Cytoskeleton (CSK) Buffer

10 mM Pipes pH 6.8.
100 mM NaCl.
1 mM EGTA.
300 mM Sucrose.
3 mM $MgCl_2$.
1× PMSF.
1X PIC
ddH_2O.
Keep the buffer at 4 °C.

2.6.3 Cytoskeleton (CSK)/NaCl Buffer

10 mM Pipes pH 6.8.
2 M NaCl.
1 mM EGTA.
300 mM Sucrose.
3 mM $MgCl_2$.
1× PMSF.
1× PIC.
ddH_2O.
Keep the buffer at 4 °C.

2.6.4 TE Buffer

10 mM Tris–HCl pH 8.
1 mM EDTA.
ddH_2O.

3 Methods

3.1 Culture of C2C12 and S2 Cell Lines

3.1.1 Culture of C1C12 Myoblasts

1. In a bath at 37 °C thaw a cryo-tube containing 5×10^5 of cells and culture them in a 10 cm^2 dish with growth medium.
2. Change media after 24 h.
3. Next day check the cells; if they reach 60–70 % of density, trypsinize cells for 2–3 min at 37 °C, and seed 5×10^5 cells in a new 10 cm^2 dishes.
4. Split cells (5×10^5 cells for a 10 cm^2 dish) every 48 h.
5. To perform chromatin fractionation, prepare 8×10 cm^2 dishes of cells for myoblasts. Around 8×10^6 cells will be obtained the day after, ready for lysis or differentiation procedure.

3.1.2 Differentiation of C1C12 Myoblasts in Myotubes

1. Check C2C12 myoblast confluence; at 80–90 % of density shift them from growth medium to differentiating medium.
2. Collect myotubes at preferred time point, considering that myotubes reach a high fusion index after 96 h (4 days of differentiation).
3. To perform chromatin fractionation of myotubes, prepare 4×10 cm^2 dishes of cells. Around 8×10^6 myotubes will be obtained, ready for lysis.

3.1.3 Culture of D. melanogaster Embryonic S2 Cells

1. In a bath at 25 °C thaw a cryo-tube containing 2×10^6 of cells, add directly the cells in 5 mL of medium in a 25 cm^2 flask under the hood. Let the cells deposit on the bottom of the flask for 30 min and then gently remove the medium with a pipette. Add 5 mL of fresh medium and incubate cells at 25 °C.
2. Dilute cells (2×10^6 cells for a 25 cm^2 flask) every 48–72 h simply adding 1 mL of culture to 4 mL of fresh medium in a new 25 cm^2 flask.
3. To perform chromatin fractionation, prepare 15×25 cm^2 flasks of S2 cells. Around 35×10^6 cells will be obtained 2 days after plating, ready for lysis.

3.2 Chromatin Fractionation

3.2.1 Chromatin Fractionation (Fig. 1)

1. Trypsinize cells (*see* **Note 2**).
2. Inhibit trypsin resuspending the cells with media containing serum.
3. Count the cells with Burker camera or preferred method (*see* **Note 3**).
4. Centrifuge cells at $150 \times g$ for 5 min at 4 °C.
5. Wash cells twice with cold 1× PBS, and centrifuge at $150 \times g$ for 5 min at 4 °C.
6. Discard supernatant.
7. Keep cell pellet on ice.
8. Resuspend pellet in 600 µL of CSK Triton buffer (*see* **Note 4**).
9. Syringe ten times (*see* **Note 5**).
10. Keep on wheel for 10 min at 4 °C.
11. Centrifuge for 3 min at $950 \times g$ at 4 °C.
12. Collect supernatant designing it as S1 (*see* **Note 6**).
13. Resuspend the pellet in 600 µL of CSK Triton buffer (*see* **Note 7**).
14. Keep on wheel for 10 min at 4 °C.
15. Centrifuge for 3 min at $950 \times g$ at 4 °C.
16. Discard the supernatant (*see* **Note 8**).
17. Resuspend gently the pellet in 100 µL of CSK buffer.

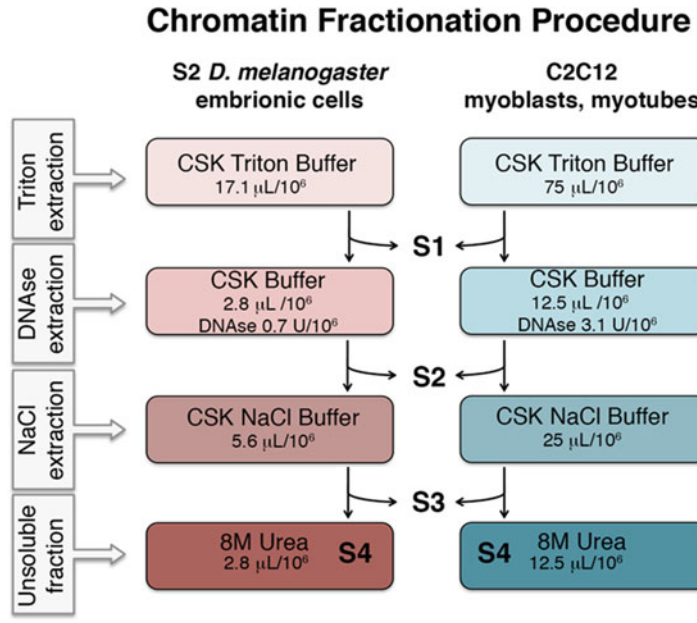


Fig. 1 Flow chart of chromatin fractionation procedure. Flow chart of chromatin fractionation procedure in S2 and C2C12 cells. For each buffer is indicated the volume required per million of cells. S1 fraction is obtained after Triton extraction of cytosolic and nuclear soluble proteins; S2 fraction is obtained after a step of DNase digestion of nuclei and chromatin solubilization; S3 fraction is obtained after a stringent NaCl extraction that removes all DNA and histones from the nuclei; S4 fraction is obtained after nuclear matrix solubilization in 8 M urea

18. Add 25 U of RNase-free DNase and incubate at 37 °C for 1 h.
19. Add $(NH_4)_2SO_4$ to a final concentration of 250 mM, and incubate for 5 min in ice.
20. Centrifuge for 3 min at 2350 × *g* at 4 °C.
21. Collect supernatant designing it as S2 (*see* **Note 9**).
22. Resuspend the pellet with 200 µL of CSK buffer.
23. Keep on wheel for 10 min at 4 °C.
24. Centrifuge for 3 min at 2350 × *g* at 4 °C.
25. Discard the supernatant.
26. Resuspend gently the pellet in 100 µL of CSK NaCl buffer (*see* **Note 10**).
27. Keep on wheel for 10 min at 4 °C.
28. Centrifuge for 3 min at 2350 × *g* at 4 °C.
29. Collect supernatant designing it as S3 (*see* **Note 11**).
30. Wash twice with 600 µL of CSK NaCl buffer.
31. Keep on wheel for 10 min at 4 °C.

32. Centrifuge for 3 min at 2350 × *g* at 4 °C.
33. Resuspend the pellet in 100 µL 8 M urea.
34. Keep the tube at RT for 10 min, pipetting after 5 min.
35. Collect supernatant designing it as S4 (*see* **Note 12**).
36. Quantify proteins with preferred assay.
37. Store at −80 °C. If you are interested in DNA fractions proceed to Subheading 3.2.2.

3.2.2 DNA Extraction from Fractions Obtained by Chromatin Fractionation

1. Take an aliquot corresponding to 30% of volume of S2, S3, and S4 fractions (*see* **Notes 13** and **14**).
2. Add TE buffer to each fraction to reach a final volume of 300 µL.
3. Add 60 µg of RNase A and incubate for 90 min at 37 °C.
4. Add 60 µg of proteinase K and incubate for 150 min at 55 °C.
5. Add 300 µL of 25:24:1 phenol:chlorophorm:isoamyl alcohol and mix well.
6. Centrifuge at 18,000 × *g* for 4 min.
7. Take aqueous phase.
8. Add 300 µL of TE buffer to phenol:chlorophorm:isoamyl alcohol phase and mix well (*see* **Note 15**)
9. Centrifuge at 18,000 × *g* for 4 min.
10. Take aqueous phase mixing to the phase obtained in point 6.
11. Precipitate aqueous phase adding 2 volumes of 98% ethanol, 1/10 volumes of 3 M sodium acetate pH 5.2, 20 µg of glycogen and put at −20 °C overnight.
12. Centrifuge at 18,000 × *g* for 30 min.
13. Wash DNA pellet with 500 µL of 70% ethanol.
14. Centrifuge at 18,000 × *g* for 10 min.
15. Remove ethanol, let the pellet air-dry, and resuspend in 25 µL of 10 mM Tris–HCl pH 7.5.
16. Quantify DNA with preferred assay.

3.3 Western Blot Assay

Analyze chromatin fractionation phases by SDS-PAGE (Fig. 2) using a 12% gel for histones and an 8–10% gel (37.5:1 Acryl/Bis Acrylamide) for the remaining proteins (*see* **Note 16**). *See* Table 1 for the complete list of antibodies used in Fig. 2.

The purity of chromatin fractionation experiment must be checked before further analysis: fraction-specific proteins should be present only in the indicated fractions:

1. α-Tubulin, specific for S1 fraction.
2. H3, specific for S2 and S3 fractions.
3. Lamin B, specific for S4 fraction (*see* **Note 17**, Fig. 3).

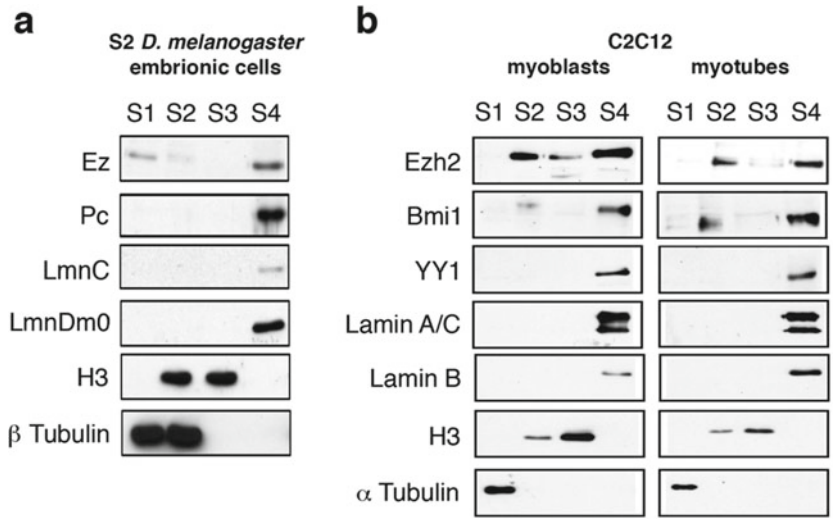


Fig. 2 PcG proteins distribute in chromatin-bound and matrix-enriched fraction. (a) Western blot for Ez, and Pc in S1, S2, S3, and S4 fractions extracted from S2 cells. LmnC, LmnDm0, H3, and β-tubulin were used for loading controls. (b) Western blot for Ezh2, Bmi1, YY1, and lamin A/C in S1, S2, S3, and S4 fractions extracted from C2C12 myoblasts and myotubes. Lamin B, H3, and α-tubulin were used for loading controls

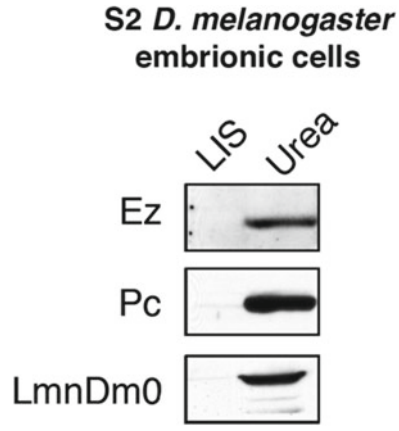


Fig. 3 LIS chromatin fractionation procedure in *Drosophila* revealed that PcG proteins distribute in matrix-enriched fraction. Western blot for Ez and Pc in chromatin-bound fraction obtained after S2 nuclei treatment with LIS detergent (LIS), and in matrix-enriched fraction extracted with urea (Urea). LmnDm0 was used for loading control

3.4 DNA Fraction Analysis

qPCR reactions were performed in triplicate; the fraction-specific relative enrichment expressed in percentage can be calculated as follows:

1. Amplify equal amount of DNA for each fraction and for a pool of the fraction (S2 + S3 + S4).
2. Perform $2^{-\Delta Ct}$.

3. Divide the value of fraction-specific $2^{-\Delta Ct}$ for the sum of $2^{-\Delta Ct}$ of all fractions to obtain the frequency. For instance: S2 frequency = $2^{-\Delta Ct}$ S2/($2^{-\Delta Ct}$ S2 + $2^{-\Delta Ct}$ S3 + $2^{-\Delta Ct}$ S4).

4 Notes

1. For an optimal cellular permeabilization and protein extraction we suggest to prepare a fresh 10% Triton X-100 solution; this is in principle to avoid variability between experiments due to detergent deterioration. The solution can be stored in the dark at room temperature up to 2 weeks.
2. To collect the cells avoid the use of scrapers in order to preserve cellular and nuclear structures and protein compartmentalization. It is preferable to use cold PBS (stored at 4 °C) to wash cells to prevent protein degradation. Since *Drosophila* S2 cells grow in suspension, cells are collected without trypsin.
3. For quantitative analysis and for comparison of distinct populations we suggest to count cells and to start from the same amount of cells. For C2C12 myoblasts or myotubes we use 8×10^6 cells; for *Drosophila* S2 cells we collect 35×10^6 cells.
4. The indicated volumes work finely with the amount of cells (*Drosophila* or C2C12) indicated above. When using different cell numbers, adjust the volume accordingly.
5. Use the syringe only for cells with abundant cytoplasm that are more difficultly lysed. For *Drosophila* S2 cells, syringe step is not needed.
6. S1 fraction is enriched of the soluble nuclear and cytoplasmic proteins. Triton X-100 in a concentration between 0.1 and 1% is able to permeabilize plasma membrane and partially nuclear membrane; thus CSK Triton extraction allows the solubilization both of cytosolic and nuclear soluble proteins.
7. Washes are essential to obtain uncontaminated fraction. In case of cross-contamination between fractions you can increase the numbers of washes and/or the volumes.
8. Remove gently the supernatant, and keep attention in withdrawing all the supernatant using if needed p20 pipette to be more precise. This is required to obtain high-pure fractions.
9. S2 fraction is enriched in chromatin-binding proteins. To avoid cross-contamination between S2 and S3 fractions, pay attention in carefully removing S2 phase avoiding contact of the viscous pellet (enriched in genomic DNA) with the microtips.
10. In CSK NaCl buffer biological material tends to attach to the plastic tip. In order to preserve your sample and to have quantitative reliable results, avoid extensive pipetting. Resuspend the pellet using a cut tip and perform subsequent washes just

placing an appropriate amount of CSK NaCl buffer in the tube and flapping the tube several times to resuspend the pellet.

11. The treatment in CSK NaCl buffer removes entirely the histones from the DNA. S3 fraction is enriched in histones and histone-bound proteins.
12. Urea treatment dissolves cellular membranes releasing membrane-associated proteins. S4 fraction is enriched in nucleo-skeleton and membrane-associated proteins and it is considered as nuclear matrix-containing fraction.
13. S1 fraction contains a negligible amount of DNA; thus it is not processed for DNA analysis.
14. If you are interested in DNA analysis only, we recommend to check the purity of the fractions running western blots with fraction-specific controls indicated above.
15. To improve the yield of the DNA extracted and for reproducibility between experiments, we strongly suggest to perform a back extraction of DNA adding equal volume of TE buffer to the phenol:chlorophorm:isoamyl fraction and repeat step 5, 6 and 7 of paragraph 3.2.2.
16. To perform quantitative western blot we suggest loading a range between 1 and 5 μg of proteins for histones, and between 5 and 20 μg for other proteins.
17. Alternative protocol for chromatin fractionation has been developed by Mirkovitch and colleagues in 1984 [21] that devised a low-salt nuclear matrix preparation protocol; this, in principle, can avoid proteins precipitation due to high salt extraction. For this purpose, Reyes and colleagues demonstrated that the distribution of SWI/SNF components using high-salt and low-salt extraction protocols was unaltered, suggesting the absence of artifacts with the high-salt extraction protocol [24]. Mirkovitch protocol foresees isolation of nuclei using a buffer complemented with spermine, spermidine, and 0.1% digitonin (3.75 mM Tris–HCl pH 7.4; 0.05 mM spermine; 0.125 mM spermidine; 0.5 mM EDTA; 20 mM KCl; 5 mM $MgCl_2$; 0.1% digitonin; 1× PIC). The nuclei were then DNase digested in the absence of EDTA (3.75 mM Tris–HCl pH 7.4; 0.05 mM spermine; 0.125 mM spermidine; 20 mM KCl; 5 mM $MgCl_2$; 0.1% digitonin; 1× PIC), centrifugated, and resuspended in a low-salt buffer complemented with 25 mM 3,5-diiodosalicylicacid lithium salt (LIS) (5 mM Hepes pH 7.4; 0.25 mM spermidine; 2 mM EDTA; 2 mM KCl; 0.1% digitonin; 25 mM LIS; 1× PIC). By centrifugation chromatin-depleted nuclei were thus recovered and the pellet was solubilized in 8 M urea. We performed this procedure to check the compartmentalization of PcG proteins in *Drosophila* S2 cells. Using the two protocols in parallel we obtained similar results, confirming the reliability of both protocols (Fig. 2A and 3).

Acknowledgments

We thank P. Fisher, J. Muller and R. Paro for providing antibodies used in this study.
This work was supported by grants from the Italian Ministry of Education, University, and Research (Futuro in Ricerca RBFR106S1Z_001) and the flagship Consiglio Nazionale delle Ricerche project (Epigen).

References

1. Bianchi A, Lanzuolo C (2015) Into the chromatin world: role of nuclear architecture in epigenome regulation. AIMS Biophys 2(4):585–612. doi:10.3934/biophy.2015.4.585, http://dx.doi.org/
2. Lanzuolo C, Orlando V (2007) The function of the epigenome in cell reprogramming. Cell Mol Life Sci 64(9):1043–1062
3. Mao YS, Zhang B, Spector DL (2011) Biogenesis and function of nuclear bodies. Trends Genet 27(8):295–306. doi:10.1016/j.tig.2011.05.006
4. Gruenbaum Y, Foisner R (2015) Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation. Annu Rev Biochem 84:131–164. doi:10.1146/annurev-biochem-060614-034115
5. Hozak P, Sasseville AM, Raymond Y, Cook PR (1995) Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. J Cell Sci 108(Pt 2):635–644
6. Kolb T, Maass K, Hergt M, Aebi U, Herrmann H (2011) Lamin A and lamin C form homodimers and coexist in higher complex forms both in the nucleoplasmic fraction and in the lamina of cultured human cells. Nucleus 2(5):425–433. doi:10.4161/nucl.2.5.17765, http://dx.doi.org/10.4161/nucl.2.5.17765
7. Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B (2006) Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nat Genet 38(9):1005–1014. doi:10.1038/ng1852
8. Lund E, Oldenburg AR, Delbarre E, Freberg CT, Duband-Goulet I, Eskeland R, Buendia B, Collas P (2013) Lamin A/C-promoter interactions specify chromatin state-dependent transcription outcomes. Genome Res 23(10):1580–1589. doi:10.1101/gr.159400.113
9. Mejat A, Decostre V, Li J, Renou L, Kesari A, Hantai D, Stewart CL, Xiao X, Hoffman E, Bonne G, Misteli T (2009) Lamin A/C-mediated neuromuscular junction defects in Emery-Dreifuss muscular dystrophy. J Cell Biol 184(1):31–44. doi:10.1083/jcb.200811035
10. Scaffidi P, Misteli T (2008) Lamin A-dependent misregulation of adult stem cells associated with accelerated ageing. Nat Cell Biol 10(4):452–459. doi:10.1038/ncb1708
11. Cesarini E, Mozzetta C, Marullo F, Gregoretti F, Gargiulo A, Columbaro M, Cortesi A, Antonelli L, Di Pelino S, Squarzoni S, Palacios D, Zippo A, Bodega B, Oliva G, Lanzuolo C (2015) Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. J Cell Biol 211(3):533–551. doi:10.1083/jcb.201504035
12. Di Croce L, Helin K (2013) Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol 20(10):1147–1155. doi:10.1038/nsmb.2669
13. Lanzuolo C, Orlando V (2012) Memories from the polycomb group proteins. Annu Rev Genet 46:561–589. doi:10.1146/annurev-genet-110711-155603
14. Cmarko D, Verschure PJ, Otte AP, van Driel R, Fakan S (2003) Polycomb group gene silencing proteins are concentrated in the perichromatin compartment of the mammalian nucleus. J Cell Sci 116(Pt 2):335–343
15. Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, Cavalli G (2011) Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell 144(2):214–226. doi:10.1016/j.cell.2010.12.026
16. Jackson DA, Cook PR (1985) Transcription occurs at a nucleoskeleton. EMBO J 4(4):919–925
17. Jackson DA, Hassan AB, Errington RJ, Cook PR (1993) Visualization of focal sites of transcription within human nuclei. EMBO J 12(3):1059–1065
18. Stratling WH (1987) Gene-specific differences in the supranucleosomal organization of rat liver chromatin. Biochemistry 26(24):7893–7899

19. Berezney R, Mortillaro MJ, Ma H, Wei X, Samarabandu J (1995) The nuclear matrix: a structural milieu for genomic function. Int Rev Cytol 162A:1–65
20. Berezney R, Coffey DS (1974) Identification of a nuclear protein matrix. Biochem Biophys Res Commun 60(4):1410–1417
21. Mirkovitch J, Mirault ME, Laemmli UK (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell 39(1):223–232
22. Fey EG, Krochmalnic G, Penman S (1986) The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. J Cell Biol 102(5):1654–1665
23. He DC, Nickerson JA, Penman S (1990) Core filaments of the nuclear matrix. J Cell Biol 110(3):569–580
24. Reyes JC, Muchardt C, Yaniv M (1997) Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix. J Cell Biol 137(2):263–274

Chapter 16

An Automatic Segmentation Method Combining an Active Contour Model and a Classification Technique for Detecting Polycomb-group Proteins in High-Throughput Microscopy Images

Francesco Gregoretti, Elisa Cesarini, Chiara Lanzuolo, Gennaro Oliva, and Laura Antonelli*

Abstract

The large amount of data generated in biological experiments that rely on advanced microscopy can be handled only with automated image analysis. Most analyses require a reliable cell image segmentation eventually capable of detecting subcellular structures.

We present an automatic segmentation method to detect Polycomb group (PcG) proteins areas isolated from nuclei regions in high-resolution fluorescent cell image stacks. It combines two segmentation algorithms that use an active contour model and a classification technique serving as a tool to better understand the subcellular three-dimensional distribution of PcG proteins in live cell image sequences. We obtained accurate results throughout several cell image datasets, coming from different cell types and corresponding to different fluorescent labels, without requiring elaborate adjustments to each dataset.

Key words Cell segmentation, Fluorescence microscopy, High-throughput imaging, Polycomb group of proteins, Variational models, Thresholding techniques

1 Introduction

Nowadays, biological studies make intensive use of high-throughput microscopy image analysis in order to perform live cell experiments of intracellular phenomena [1, 2]. Manual processing of such image sequences is often subjective and error prone and also very laborious, creating a strong demand for automated imaging techniques [3–5]. In particular, image segmentation is a fundamental task because it produces a partition of the image separating cells from background, highlighting subcellular compartments and enabling extraction and quantification of features and properties (i.e., shape, area, signal distribution, etc.). The image features that have to be highlighted are strictly related to the kind of phenomenon under

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_16,

investigation produced by the biological experiment. So far, existing tools for segmentation are still somehow limited in their scope and capacity for cell imaging. On the other hand, segmentation of fluorescence microscopy images still put significant challenges in mathematical modeling and algorithms [6], e.g., the variety of fluorescence proteins and labeling techniques lead to considerable differences in the appearance of cells; often the images are corrupted by noise and can contain autofluorescence; moreover, since cells are sensible to photodamage, fluorescence microscopy produces images with very low contrast [7]. Unluckily, there is no general way to determine what any particular segmentation method might work better: methods and techniques differ in the way they emphasize one or more of the desired image features/properties and in the way they balance and compromise one selected feature/property against another [8]. These methods and techniques can be classified in a different way, according to Aubert et al. [9] they are classified into two groups depending on the different approach used to produce a partition of image. The first one comprises the Region-based Segmentation methods, like the one based on the famous Mumford–Shah model [10] and its variants [11], region growing algorithms [12], classification methods based on global or local thresholding techniques [13]. These methods provides an image partition as a map of regions with constant intensities (i.e., a simplified image). The second group comprises the Active Contour (AC) methods [14–16] or snakes, initially proposed by Kass et al. [17]. Unlike the Region-based methods, the aim of AC is detecting object contours by means of a curve that separates objects from the image background.

We recently presented an alternate cell segmentation method combining an AC method with a thresholding technique for the analysis of fluorescence microscopy images with particular focus on the detection of Polycomb group (PcG) protein bodies. PcG bodies are prominent discrete nuclear structures: PcG proteins are concentrated in nuclear foci that can change in number and dimension mirroring cell dynamics (cell cycle, stress, differentiation, etc.) [18]. Our segmentation method assumes that you have two fluorescence images of the cells stained respectively to label nuclei and PcG bodies and carries out a "trinary" segmentation of the fluorescent cell images: regions are segmented as background, nuclei, and PcG bodies. It serves as a tool to improve the knowledge about how PcG bodies are organized inside the nuclei and, as matter of fact, it has been employed for a high-throughput cell-image analysis in order to study the role of Lamin A/C in PcG intranuclear compartmentalization [19]. In that work the segmented stack has been used to perform the 3D reconstruction of the PcG foci estimating their volume.

The first algorithm of our combined method separates nuclei regions from background implementing a new AC method based

on a modified version of the Local Chan–Vese (LCV) model [20]. Since the model in Wang et al. [20] is efficient and reliable for high-dimensional images with low-contrast boundaries and robust with respect to the noise, it is particularly suitable for the segmentation of fluorescence cell images. We refer to the first algorithm as Modified Chan–Vese (MCV) and to the algorithm implementing the original model as LCV. MCV provides an optimal binary partition of images (nuclei regions and background) by solving the following minimization problem:

$$min_{c1,\,c2,\,d1,\,d2,\,C}\,[E^{MCV}(c_1, c_2, d_1, d_2, C)] \text{ where:}$$
$$E^{MCV}(c_1, c_2, d_1, d_2, C) = \mu\, Length(C) + \alpha\,(\int_{inside(C)} |I - c_1|^2\, dxdy + \int_{outside(C)} |I - c_2|^2\, dxdy) +$$
$$+ \beta\,(\int_{inside(C)} |I^* - I - d_1|^2\, dxdy + \int_{outside(C)} |I^* - I - d_2|^2\, dxdy).$$

The functional E^{MCV} controls the evolution of the curve C by means of the three terms weighted by positive parameters μ, α, and β respectively. The first term (internal forces of the curve) is the smoothness term measuring the perimeter of C. The second two terms (external forces) represent the data fidelity term or the object edge-detector (external forces) that attracts the curve toward the nuclei boundaries. In detail, the term weighted by α is the sum of the intensity variances of the regions of I inside and outside C respectively, where I is the nuclear fluorescence image and c_1 and c_2 are the mean intensity values of regions of I inside and outside C respectively. Instead in the term weighted by β, different from the one used in [20], we enhanced the contrast between the nuclei regions and image background by incorporating statistical information on both nuclear and PcG images. This has the aim to preserve the integrity of nuclei regions that is a fundamental task in order to have a correct detection of PcG bodies. In fact, I^* is obtained as sum of $g_k(I)$ and $g_k(I^G)$, where g_k is the averaging convolution operator with $k \times k$ window and I^G is the PcG fluorescence image. This term represents the sum of the intensity variances of the regions of the composite image $I^*–I$ inside and outside C respectively, where d_1 and d_2 are the mean intensity values of $I^*–I$ inside and outside C respectively. Notice that the edge-detector function of E^{MCV} acts by adaptively balancing the intensity variances of the nuclei regions and image background of I and $I^*–I$ respectively. Moreover, being not based on image intensity gradient, it is well suitable for cell image segmentation when nuclei regions have irregular intensities and cannot be clearly distinguished from the background by a sharp variation of the image intensity. The evolution of C stops when an optimal partition of the image into two regions is achieved.

The minimization problem is usually solved by means of the level set methods [21], where the curve C is represented by the

zero level set of a Lipschitz function ϕ defined in the image domain Ω with values in the set of real numbers such that:

$$C = \{(x,y) \text{ in } \Omega \text{ such that } \phi(x,y) = 0\}$$
$$inside(C) = \{(x,y) \text{ in } \Omega \text{ such that } \phi(x,y) < 0\}$$
$$outside(C) = \{(x,y) \text{ in } \Omega \text{ such that } \phi(x,y) > 0\}$$

The expression of the functional E^{MCV} in terms of function ϕ is analog to the one of the functional E^{LCV} described in [20]. The functional E^{MCV} is minimized by computing the Euler Lagrange (EL) equation and then solving the resulting PDE by means of the gradient descent method. Differently from the work of Wang et al. [20] where an explicit scheme is employed, we used a time-dependent semi-implicit scheme in order to discretize EL equations. Although the use of an implicit time discretization scheme requires a greater computational effort with respect to an explicit one, it is compensated by the unconditional stability of the numerical algorithm. Furthermore, usually segmentation of high dimensional images (like fluorescence images) requires a lot of iterations to converge to numerical solution. By adopting an implicit scheme we can reduce the number of iterations allowing the choice of larger time steps. We solved the time dependent PDE shown in (Fig. 1) by using the iterative implicit scheme shown in (Fig. 2).

$$\begin{aligned}
\frac{\phi_{ij}^{n+1} - \phi_{ij}^{n}}{dt} = \delta_\epsilon(\phi_{ij}^n) \Bigg[\frac{\mu}{h^2} \Bigg(& \frac{\phi_{i+1j}^{n} - \phi_{ij}^{n+1}}{\sqrt{\nu^2 + \left(\frac{\Delta_+^x \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^y \phi_{ij}^n}{2h}\right)^2}} - \frac{\phi_{ij}^{n+1} - \phi_{i-1j}^{n}}{\sqrt{\nu^2 + \left(\frac{\Delta_-^x \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^y \phi_{i-1j}^n}{2h}\right)^2}} \\
& + \frac{\phi_{ij+1}^{n} - \phi_{ij}^{n+1}}{\sqrt{\nu^2 + \left(\frac{\Delta_+^y \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^x \phi_{ij}^n}{2h}\right)^2}} - \frac{\phi_{ij}^{n+1} - \phi_{ij-1}^{n}}{\sqrt{\nu^2 + \left(\frac{\Delta_-^y \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^x \phi_{ij-1}^n}{2h}\right)^2}} \Bigg) \\
& - \lambda_1 \left(I_{ij} - c_1(\phi^n)\right)^2 + \lambda_2 \left(I_{ij} - c_2(\phi^n)\right)^2 \Bigg]
\end{aligned}$$

Fig. 1 The time dependent PDE solved by means of MCV, where h is the space step, dt is the time step, $(x_i, y_j) = (ih, jdt)$ are the grid points for $1 \le i, j \le M$ and $\phi_{ij}^n = \phi(ndt, x_i, y_j)$ is an approximation of $\phi(t,x,y)$ with $n \ge 0$ and $\phi^0 = \phi_0$. Moreover, $\Delta_\pm$ denotes the forward (or the backward) difference operator in x (or in y) direction, i.e. $\Delta_\pm^x = \pm(\phi_{i\pm1j} - \phi_{ij})$ and $\Delta_\pm^y = \pm(\phi_{ij\pm1} - \phi_{ij})$, instead of Δ_0 denotes the central difference operator in x (or in y direction), i.e. $\Delta_0^x = \pm(\phi_{i+1j} - \phi_{i-1j})$ and $\Delta_0^y = \pm(\phi_{ij+1} - \phi_{ij-1})$. δ_ε is a smoothed version of the one-dimensional Dirac function

$$\phi_{ij}^{n+1} = \left(\frac{h}{h + dt\left(\delta_\epsilon\left(\phi_{ij}^n\right)\mu - 1\right)\left(C1 + C2 + C3 + C4\right)} \right)$$
$$\left[\phi_{ij}^n + dt\left(\delta_\epsilon\left(\phi_{ij}^n\right)\mu\left(C1\,\phi_{i+1j}^{n+1} + C2\,\phi_{i-1j}^{n+1} + C3\,\phi_{ij+1}^{n+1} + C4\,\phi_{ij-1}^{n+1}\right)\right.\right.$$
$$\left.\left. - dt\,\delta_\epsilon\left(\phi_{ij}^n\right)\left(\lambda_1\left(I_{ij} - c_1\left(\phi^n\right)\right)^2 + \lambda_2\left(I_{ij} - c_2\left(\phi^n\right)\right)^2\right)\right)\right]$$

where:

$$C1 = \frac{1}{\sqrt{\nu^2 + \left(\frac{\Delta_+^x \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^y \phi_{ij}^n}{2h}\right)^2}} + \frac{1}{h\left(\mu\delta_\epsilon\left(\phi_{ij}^n\right) - 1\right)}$$

$$C2 = \frac{1}{\sqrt{\nu^2 + \left(\frac{\Delta_-^x \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^y \phi_{i-1j}^n}{2h}\right)^2}} + \frac{1}{h\left(\mu\delta_\epsilon\left(\phi_{ij}^n\right) - 1\right)}$$

$$C3 = \frac{1}{\sqrt{\nu^2 + \left(\frac{\Delta_+^y \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^x \phi_{ij}^n}{2h}\right)^2}} + \frac{1}{h\left(\mu\delta_\epsilon\left(\phi_{ij}^n\right) - 1\right)}$$

$$C4 = \frac{1}{\sqrt{\nu^2 + \left(\frac{\Delta_-^y \phi_{ij}^n}{h}\right)^2 + \left(\frac{\Delta_0^x \phi_{ij-1}^n}{2h}\right)^2}} + \frac{1}{h\left(\mu\delta_\epsilon\left(\phi_{ij}^n\right) - 1\right)}$$

Fig. 2 Iterative implicit scheme implemented in MCV algorithm

The sketch of MCV algorithm follows:

MCV(*in*: I, I^G; *out*: ϕ^{k^*})
n = 0 initialize ϕ^0 by ϕ_o
compute $I^* = g_k(I) + g_k(I^G)$
m = 1
for $n = 1,2,\ldots,K$ (outer loop)
  compute $c_1(\phi^{n-1}(I))$, $c_2(\phi^{n-1}(I))$, $d_1(\phi^{n-1}(I^*-I))$ *and* $d_2(\phi^{n-1}(I^*-I))$
  compute ϕ^n (inner loop)([1])
  if $|Length(C^n) - Length(C^{n-1})| < T_L$ *then*
    if $m = T_i$ *then*
      stop evolution curve
    $m = m + 1$
  else
    $m = 1$

where $K^* \leq K$ is the last iteration step and the last six lines represents the termination criterion used in [20]: if the absolute value of the change of the curve length remains smaller than a given threshold T_L over a fixed number of iterations T_i, the evolution of curves will be stopped.

[1] The inner loop is an iterative algorithm solving the scheme in (Fig. 2).

The second algorithm of our combined method uses a classification method in order to separate PcG bodies from nuclei regions estimating the classifier threshold value by means of the ISODATA [22] and by using relevant values computed by the MCV algorithm. We adopted a global thresholding because the presence of an high number of PcG bodies give rise to a blurry region around each PcG body that increases the noise and can hide small PcG bodies. Instead with a local thresholding strategy those small PcG bodies would be selected only in cells with few PcG foci.

Classical ISODATA [23] is an unsupervised classification algorithm that splits and merges nonhomogeneous image regions into two subregions, based on a global threshold that changes during an iterative refinement of the two subregions. Initial value of the threshold is arbitrary, but a bad choice of this parameter may lead the algorithm to spiral out of control leaving only one region. We avoided this drawback by following these observations: usually ISODATA assigns the mean intensity value of the entire image as initial value of the threshold when background and objects occupy comparable area in the image [24], but it was misleading for our purpose of detecting PcG bodies that correspond to small areas within nuclei regions with highest intensity values; the mean intensity value of nuclei regions seemed like a more reliable choice for the initial threshold [24]. To sum up, we set the initial threshold value as the mean intensity value of the nuclei regions, c_1, computed by the MCV algorithm. We will refer to this algorithm as Modified ISODATA (MID) while we will refer to the original ISODATA segmentation algorithm as Isodata. In order to enhance high intensity areas, the MID algorithm initializes the first nonhomogeneous regions to be split as the nuclei regions of image $\bar{I}$. $\bar{I}$ is obtained by subtracting from the original PcG image, I^G, the corresponding smoothed image obtained using an averaging convolution operator with $k \times k$ window, $g_k(I^G)$.
The sketch of MID algorithm follows:

MID(*in*: $c_1(\phi^{k^*}(I^G))$, I^G; *out:* t)
initialize $t = c_1(\phi^{k^*}(I^G))$
compute $\bar{I} = |I^G - g_k(I^G)|$
repeat
    split $\bar{I}$ *into* R_1 *and* R_2 *regions by means of* t
    compute m_1=*mean of* R_1 *intensity values*
    compute m_2=*mean of* R_2 *intensity values*
    $t = average(m_1, m_2)$
until the difference in t *in successive iterations is less than a predefined tolerance*

Given the image stacks I_j and I_j^G for $1 \le j \le N$, the sketch of combined MCV-MID algorithm follows:

ϕ_j= MCV(I_j , I_j^G)
detect nuclei regions and background for the entire stack images by using ϕ_j
compute $c_1(\phi_j(I_j^G))$
t_j = MID($c_1(\phi_j(I_j^G))$, I_j^G)
for j=1,2,…,N
    compute $c_1(\phi_j(I_j^G))$
    t_j = MID($c_1(\phi_j(I_j^G))$, I_j^G)
    detect PcG bodies in nuclei regions for stack section j *by using max*(t_j, t_j)

where $\hat{j}$ is the middle section of the stack (if not given as input).

In order to assess the accuracy of our segmentation tool we tested it using different cell image datasets from different cell lines (mouse muscle stem cells, C_2C_{12} myoblasts, Drosophila Schneider 2 cells) and corresponding to different fluorescent labels [Figs. 3, 4, 5, 6, 7]. We demonstrate that the combined use of the two modified algorithms lead to segmentation results more accurate than using their respective original formulations without requiring elaborate adjustments to each dataset. In fact, by keeping the third

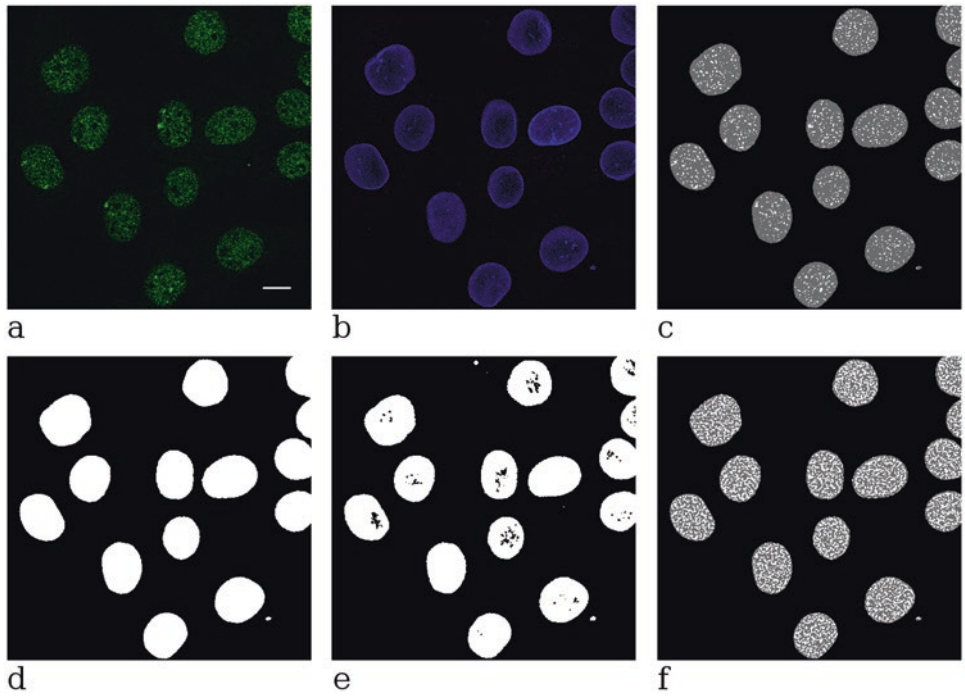


Fig. 3 Representative images of C_2C_{12} myoblasts immunostained using Bmi1 (**a**) and Lamin B (**b**) antibodies: segmentation result using MCV-MID (**c**); nuclei regions as segmented by MCV (**d**); nuclei regions as segmented by LCV (**e**); segmentation result using MCV-Isodata (**f**). Bar, 10 micron. MCV/LCV parameters: $\alpha = 1$, $\beta = 1$, $\mu = 0.5$

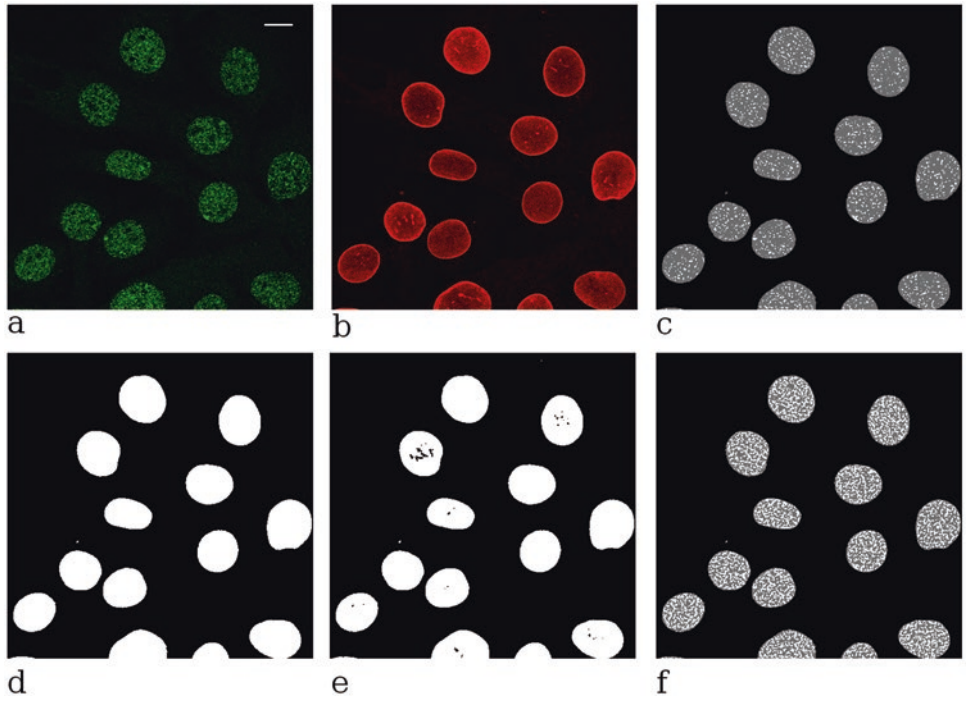


Fig. 4 Representative images of C_2C_{12} myoblasts immunostained using Bmi1 (**a**) and Lamin B (**b**) antibodies: segmentation result using MCV-MID (**c**); nuclei regions as segmented by MCV (**d**); nuclei regions as segmented by LCV (**e**); segmentation result using MCV-Isodata (**f**). Bar, 10 micron MCV/LCV parameters: $\alpha = 1$, $\beta = 1$, $\mu = 0.5$

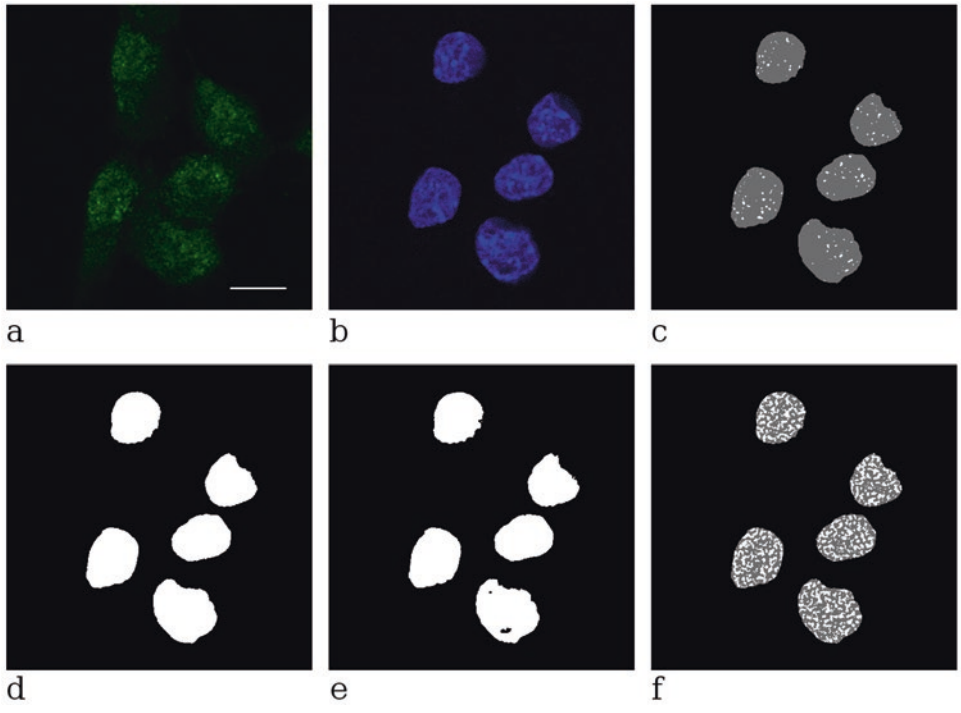


Fig. 5 Representative images of mouse muscle stem cells (satellite) immunostained using Bmi1 antibodies (**a**) and dapi (**b**): segmentation result using MCV-MID (**c**); nuclei regions as segmented by MCV (**d**); nuclei regions as segmented by LCV (**e**); segmentation result using MCV-Isodata (**f**). Bar, 10 micron. MCV/LCV parameters: $\alpha = 1$, $\beta = 1$, $\mu = 0.5$

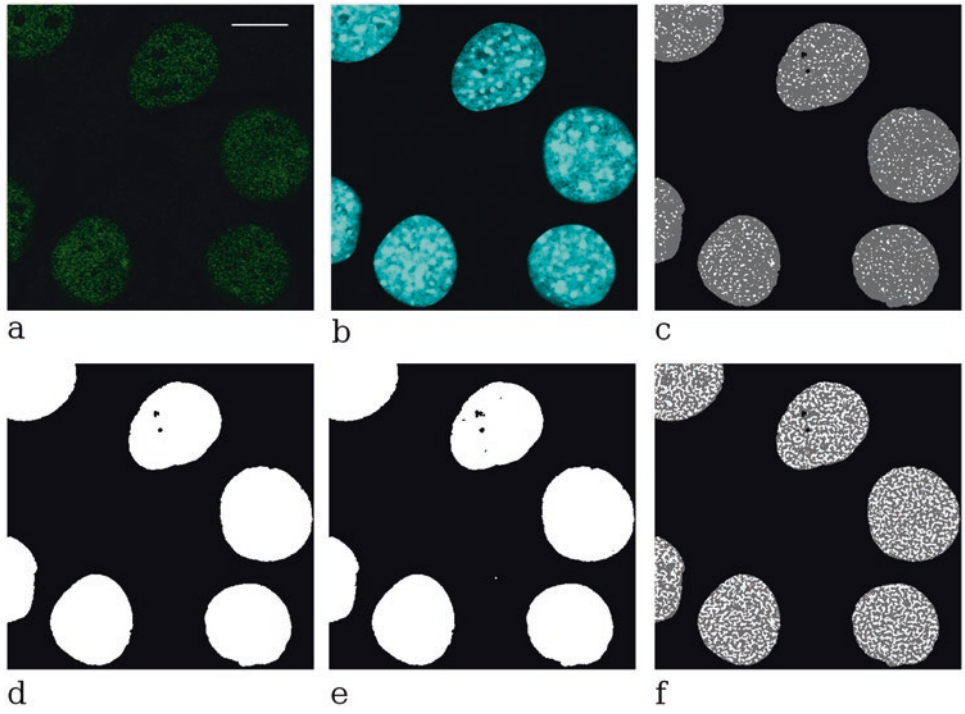


Fig. 6 Representative images of C_2C_{12} myoblasts immunostained using Ezh2 antibodies (**a**) and dapi (**b**): segmentation result using MCV-MID (**c**); nuclei regions as segmented by MCV (**d**); nuclei regions as segmented by LCV (**e**); segmentation result using MCV-Isodata (**f**). Bar, 10 micron. MCV/LCV parameters; $\alpha = 1$, $\beta = 1$, $\mu = 0.5$

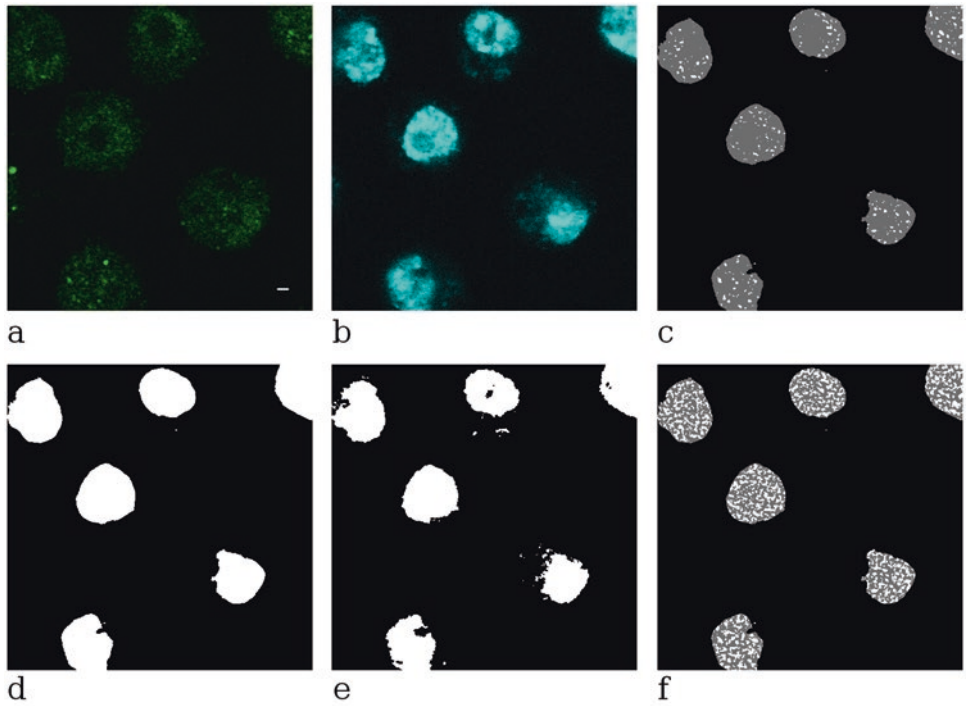


Fig. 7 Representative images of Drosophila Schneider 2 (S_2) cells immunostained using PC antibodies (**a**) and dapi (**b**): segmentation result using MCV-MID (**c**); nuclei regions as segmented by MCV (**d**); nuclei regions as segmented by LCV (**e**); segmentation result using MCV-Isodata (**f**). Bar, 1 micron. MCV/LCV parameters: $\alpha = 0.1$, $\beta = 1$, $\mu = 0.5$

term of the functional as it is in LCV model, i.e., $\beta(\int_{inside(C)} |g_k(I) - I - d_1|^2 \, dxdy + \int_{outside(C)} |g_k(I) - I - d_2|^2 \, dxdy$, we obtained a segmentation result of nuclei regions less accurate (see [Fig. 3e, 4e, 5e, 6e, 7e]). Curves provided by the original model tend to match some protein group contours besides nuclei contours. This produces holes in nuclei regions also decreasing contour regularity. Moreover, by setting the initial threshold, t, as the mean of I^G intensity values (as it is in Isodata), we obtained weak segmentation results with PcG bodies more numerous and greater than they are (see [Figs. 3f, 4f, 5f, 6f, 7f]).

2 Materials

1. *32-bit RGB TIFF confocal images of size 1024×1024 pixels.*
   We used fluorescent images from different cell lines (mouse muscle stem cells, C_2C_{12} myoblasts, Drosophila Schneider 2 cells) incubated with primary Polycomb antibodies (Bmi1 or Ezh2 or PC) and then hybridized with fluorescent-conjugated secondary antibodies. Images were taken with the confocal Leica SP5 supported by LAS-AF software; however, any confocal microscope can be used. Moreover, our tool can be used to detect any similar protein forming aggregates in the nucleus corresponding to image regions with the highest intensity values. The number of stack sections on the z axis varies between 24 and 36.
2. *Hardware test bed*
   We carried out our tests on a Linux workstation with Intel Core i5 CPU, 2.67 GHz processor speed, and 4 GB RAM.
3. *ImageJ*
   We used ImageJ [25, 26] to visualize and evaluate input and segmented images and to estimate the volume of the PcG bodies.

3 Methods

The proposed tool, MCV-MID, works on two cell image stacks in TIFF 16-bit format that include respectively the staining to visualize the nuclear contour (in our tests Lamin B or dapi) and the staining to highlight PcG bodies (in our tests Bmi1, Ezh2, or PC). In order to apply the tool input images need to be preprocessed. Therefore starting from cell image stacks we are able to detect PcG proteins areas by following a three step process: 3.1 preprocessing; 3.2 MCV-MID segmentation tool installation; 3.3 MCV-MID segmentation tool execution. Moreover, in subsection 3.4 we show how the output of the tool is used as input of the "3D Object Counter" ImageJ plugin [27] to estimate the volume of the PcG foci.

3.1 Image Preprocessing

1. Use an ImageJ macro as a preprocessing step that converts RGB images to 16-bit grayscale and apply a despeckle filter for luminance noise removal.

   The macro has to contain the following instructions:

```
changeValues(0xffffff, 0xffffff, 0x000000); (see Note 1)
run("16-bit");
run("Despeckle");
```

3.2 MCV-MID Segmentation Tool: Installation

The combined MCV-MID algorithm described in the Introduction has been implemented in C programming language. It makes use of the *libtiff* library for the properly handling of the TIFF image files [28]. The installation process is achieved by following these steps:

1. Install the prerequisite software:
   - C compiler
   - *libtiff* library
2. Download the MCV-MID tool from http://jcb.rupress.org/content/suppl/2015/11/04/jcb.201504035.DC1/JCB_201504035_S1.zip
3. Unpack the source code with the command

   ```
   unzip JCB_201504035_S1.zip
   ```
4. Change the working directory to the newly created directory with the command

   ```
   cd mcvmid-1.0.0
   ```
5. Configure the installation with the command

   ```
   ./configure
   ```

   This will check the system and automatically configure MCV-MID (*see* **Note 2**).
6. Compile MCV-MID with the command

   ```
   make
   ```
7. Install the binary with the command (optional)

   ```
   make install
   ```

3.3 MCV-MID Segmentation Tool: Execution

3.3.1 To run the MCV-MID tool, users need to provide two cell image stacks in TIFF 16-bit format that include respectively the staining to highlight nuclei and the staining to highlight PcG bodies.

Users can use the executable file `mcvmid` followed by a set of options (*see* **Note 3**):

- -b Lamin B/dapi stack first section filename: set the Lamin B/dapi stack first section filename.
- -c Bmi1/Ezh2/PC stack first section filename: set the Bmi1/Ezh2/PC stack first section filename.
- -f first section number: set the number of the first section of the stack to be considered for the segmentation (default 1).
- -l last section number: set the number of the last section of the stack to be considered for the segmentation (default 1).

- -D MCV parameter *dt*: set the MCV parameter time step spacing *dt* (default 0.1).
- -1 MCV parameter α: set the first MCV data-fidelity parameter α (default 1.0) (*see* **Note 4**).
- -2 MCV parameter β: set the second MCV data-fidelity parameter β (default 1.0).
- -M MCV parameter μ: set the MCV contour length weighting parameter μ (default 0.5).
- -H MCV parameter h: set the MCV parameter pixel spacing h (default 1.0).
- -K maximum iteration number of the outer loop in the MCV algorithm: set the maximum number of outer loop iterations K (default 300).
- -L maximum iteration number of the inner loop in the MCV algorithm: set the maximum number of inner loop iterations (default 5).
- -V choice of the convolution operator (g_k) size used in MCV algorithm: set to use an averaging convolution operator with $V \times V$ size window (default V=3).
- -T choice of a preprocessing for Lamin B/dapi image: set to use an averaging convolution operator with $T \times T$ size window for preprocessing (default do no preprocessing, T=0).
- -I choice of the convolution operator (g_k) size used in MID algorithm: set to use an averaging convolution operator with $I \times I$ size window (default I=4) (*see* **Note 5**).
- -Z section number $\hat{\jmath}$ used in MCV-MID algorithm: set the section to be used by MCV to detect nuclei regions and background for the entire stack (default use the middle section).
- -s threshold for PcG bodies detection: set a fixed threshold to filter PcG bodies from nuclei regions instead of applying MID (*see* **Note 6**).

3.3.2 The produced output is a "trinary" segmented images stack: regions are segmented as black background (black colored), nuclei (gray colored), and PcG bodies (white colored) as shown in [Figs. 3–7 (see Note 7)]

For example, suppose that first input image stack files are named *dapi_z00.tif*, *dapi_z01.tif*, …, *dapi_z35.tif* while second input image stack files are named *bmi1_z00.tif*, *bmi1_z01.tif*, …, *bmi1_z35.tif*, the MCV-MID tool can be executed via a command line like this:

```
mcvmid -b dapi_z00.tif -c bmi1_z00.tif -f 00 -l 35 -L
30 -D 0.001 -V 7
```

3.4 PcG Bodies Volume Estimation

Starting from the segmented stack users can estimate the volume of PcG bodies in a specific cell by using the following procedure:

1. Install the "3D Object Counter" ImageJ plugin following the instructions in Chapter 1 of the manual that can be downloaded from "3D Objects Counter" homepage [27].
2. Open ImageJ.
3. Click **File** > **Import**, navigate to the directory containing the cell image stack to be analyzed and click **Open**, then select the first image in the sequence.
4. A new windows appears, all relevant fields like the number of images are already filled correctly, so click **OK.**
5. All images will be imported into a stack named as the directory name. The window's status bar shows the number of the current section and total sections.
6. Surround the nucleus of the cell to be analyzed with the rectangular selection tool then click **Image** > **Crop**.
7. Set voxel dimensions according to your microscope: click **Image** > **Properties**, enter *micron* for Unit of length and set Pixel width, Pixel height, and Voxel depth according to the microscope scanning setup, then click **OK**.
8. Launch the "3D Object Counter": **Plugins > 3D Object Counter > 3D Object Counter…**
9. A new window appears, set the Size filter to exclude bodies that are too small to be properly evaluated according to the microscope scanning setup (*see* **Note 8**). Leave the other default options as they are and click **OK.**
10. The analysis starts, images are generated and results logged: a results table named "Statistics for "+the image's title display all statistics, including the volume, of the found bodies. A results table example is shown in Table 1. In that example we were analyzing a cell, within the Drosophila Schneider 2 cells dataset (Fig. 8), numbered as 4, therefore all the bodies are labeled as a numbered object in the objects map of 4.

4 Notes

1. Some images came out with a micron-scale stamp. Our preprocessing macro removes this stamp to prevent it from hindering nuclei and PcG bodies detection. The first instruction performs the scale stamp removal by changing all white pixels to black pixels.
2. Run configure command with the –help option to obtain a list of command-line arguments to modify the installation.

Table 1

Example of a result table containing some of the statistics about all objects (PcG bodies) measured by the 3D Object Counter in a cell stack of the Drosophila Schneider 2 cells dataset.

	Label	Volume (micron^3)	Surface (micron^2)	No. of obj. voxels	No. of surf. voxels
1	Objects map of 4_Object_1	0.046	1.593	43	41
2	Objects map of 4_Object_2	0.023	0.797	22	21
3	Objects map of 4_Object_3	0.143	4.05	134	126
4	Objects map of 4_Object_4	0.06	2.32	56	56
5	Objects map of 4_Object_5	0.026	1.003	24	24
6	Objects map of 4_Object_6	0.038	1.466	36	36
7	Objects map of 4_Object_7	0.022	0.733	21	20
8	Objects map of 4_Object_8	0.031	1.018	29	28
9	Objects map of 4_Object_9	0.291	7.906	273	242
10	Objects map of 4_Object_10	0.074	2.811	69	68
11	Objects map of 4_Object_11	0.1	2.861	94	89
12	Objects map of 4_Object_12	0.034	1.246	32	32
13	Objects map of 4_Object_13	0.031	1.231	29	29
14	Objects map of 4_Object_14	0.022	0.875	21	21
15	Objects map of 4_Object_15	0.018	0.584	17	17
16	Objects map of 4_Object_16	0.042	1.509	39	38
17	Objects map of 4_Object_17	0.034	1.26	32	32
18	Objects map of 4_Object_18	0.186	5.616	175	168
19	Objects map of 4_Object_19	0.063	2.441	59	58
20	Objects map of 4_Object_20	0.096	2.284	90	75
21	Objects map of 4_Object_21	0.129	3.089	121	101
22	Objects map of 4_Object_22	0.114	3.623	107	105
23	Objects map of 4_Object_23	0.025	0.932	23	23
24	Objects map of 4_Object_24	0.025	1.181	23	23
25	Objects map of 4_Object_25	0.099	2.399	93	86
26	Objects map of 4_Object_26	0.079	2.142	74	65
27	Objects map of 4_Object_27	0.092	2.775	86	81
28	Objects map of 4_Object_28	0.018	0.726	17	17
29	Objects map of 4_Object_29	0.042	1.465	39	38
30	Objects map of 4_Object_30	0.018	0.626	17	17

(continued)

Table 1
(continued)

	Label	Volume (micron^3)	Surface (micron^2)	No. of obj. voxels	No. of surf. voxels
31	Objects map of 4_Object_31	0.019	0.768	18	18
32	Objects map of 4_Object_32	0.12	4.534	113	113
33	Objects map of 4_Object_33	0.206	5.572	193	176
34	Objects map of 4_Object_34	0.069	2.42	65	62
35	Objects map of 4_Object_35	0.019	0.804	18	18
36	Objects map of 4_Object_36	0.027	0.847	25	24
37	Objects map of 4_Object_37	0.022	0.69	21	21
38	Objects map of 4_Object_38	0.031	0.976	29	29
39	Objects map of 4_Object_39	0.046	1.786	43	43
40	Objects map of 4_Object_40	0.032	0.918	30	28
41	Objects map of 4_Object_41	0.019	0.662	18	18
42	Objects map of 4_Object_42	0.018	0.598	17	17
43	Objects map of 4_Object_43	0.21	5.217	197	169
44	Objects map of 4_Object_44	0.02	0.868	19	19
45	Objects map of 4_Object_45	0.023	1.081	22	22
46	Objects map of 4_Object_46	0.022	0.676	21	21
47	Objects map of 4_Object_47	0.02	0.762	19	19
48	Objects map of 4_Object_48	0.025	0.719	23	23
49	Objects map of 4_Object_49	0.082	3.124	77	74
50	Objects map of 4_Object_50	0.028	1.281	26	26
51	Objects map of 4_Object_51	0.02	0.662	19	19
52	Objects map of 4_Object_52	0.034	0.947	32	30
53	Objects map of 4_Object_53	0.033	0.954	31	30
54	Objects map of 4_Object_54	0.02	0.768	19	18
55	Objects map of 4_Object_55	0.018	0.697	17	17

3. At a minimum, required options are -b and -c.
4. In the case of the Drosophila Schneider 2 cells dataset we set $\alpha = 0.1$ in order to tackle with the intensity inhomogeneity of the cell image stained with PC (see [Fig. 7]).
5. User can set this parameter accordingly to the dimensions of image regions corresponding to the PcG bodies.

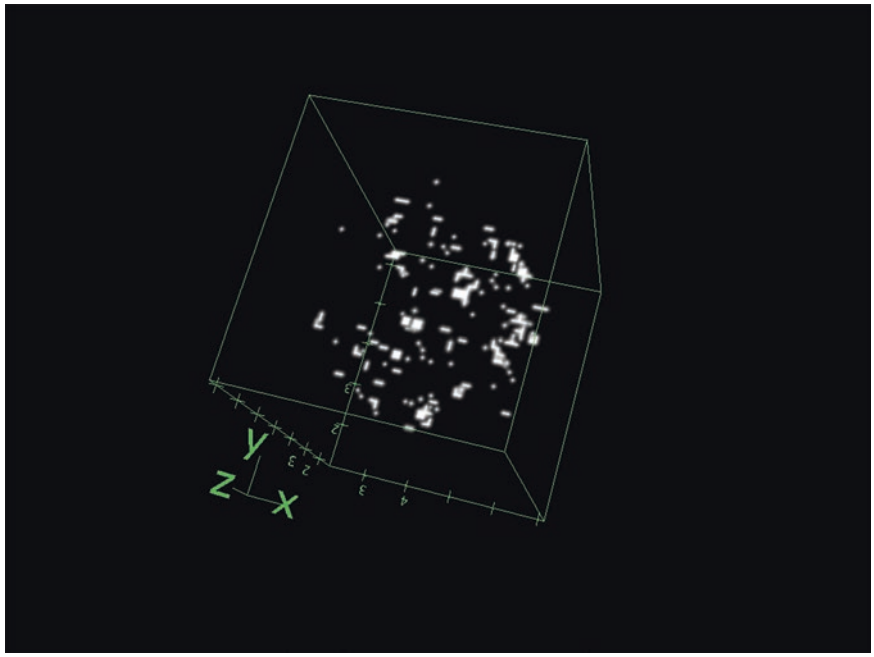


Fig. 8 3D view of the PcG bodies identified by the 3D object counter in a cell stack of the Drosophila Schneider 2 cells dataset. *tick*-marks every micron

6. This option can be useful in case of several datasets within the same biological replicate: user can decide to use the same fixed threshold for all datasets in order to have homogeneous results within populations.
7. Although the low contrast between the background and cell nuclei MCV-MID is able to segment nuclei regions with accurate results because the edge detector is based on the intensity variation between background and nuclei regions of the image rather than the intensity variation across nuclei boundaries (see [Fig. 7c].
8. In our tests we excluded bodies with volume size below 0.3 micron^3, being distance between consecutive sections in the stack 0.3 micron.

Acknowledgments

Work partially supported by FIRB 2010 Project n.~RBFR106S1Z002.

References

1. Walter T, Held M, Neumann B, Hériché JK, Conrad C, Pepperkok R, Ellenberg J (2010) Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. J Struct Biol 170:1–9
2. Handfield L, Chong YT, Simmons J, Andrews BJ, Moses AM (2013) Unsupervised clustering of subcellular protein expression patterns in high-throughput microscopy images reveals protein complexes and functional relationships between proteins. PLoS Comput Biol 9(6):e1003085
3. Shamir L, Delaney JD, Orlov N, Eckley DM, Goldberg IG (2010) Pattern recognition software and techniques for biological image analysis. PLoS Comput Biol 6(11):e1000974
4. Huang K, Murphy RF (2004) Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. BMC Bioinformatics 5:78–97
5. Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, Martone ME, Murphy RF, Peng H, Plant AL, Roysam B, Stuurman N, Swedlow JR, Tomancak P, Carpenter A (2012) Biological imaging software tools. Nat Methods 9(7): 697–710.
6. Zhou X, Wong STC (2006) Informatics challenges of high-throughput microscopy. IEEE Signal Process Mag 23(3):63–72
7. Dzyubachyk O, van Cappellen WA, Essers J, Niessen WJ, Meijering E (2010) Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. IEEE Trans Med Imaging 29(3):852–867
8. Haralick RM, Shapiro LG (1985) Image segmentation survey. CVGIP 29:100–132
9. Aubert G, Kornprobst P (2006) Mathematical problems in image processing. partial differential equations and calculus of variations, 2nd edn. Springer, New York
10. Mumford D, Shah J (1989) Optimal approximation by piecewise smooth functions and associated variational problems. Commun Pure Appl Math 42:577–685
11. Chan T, Vese L (2001) Active contours without edges. IEEE Trans Image Process 10:266–277
12. Zucher S (1976) Region growing: childhood and adolescence. Comput Graphics Image Process 5:382–399
13. Sahoo PK, Soltani S, Wong KC, Chen YC (1988) A survey of thresholding techniques. Comput Vis Graphics Image Process 41:233–260
14. Blake A, Isard M (1998) Active contours. Springer, New York
15. Sapiro G (2001) Geometric partial differential equations and image analysis. Cambridge University Press, Cambridge
16. Osher S, Fedkiw RP (2000) Level set methods. Technical Report 00–08, UCLA CAM Report
17. Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. Int J Comput Vis 1(4):321–331
18. Portoso M, Cavalli G (2008) The role of RNAi and noncoding RNAs in polycomb mediated control of gene expression and genomic programming. In: Morris KV (ed) RNA and the regulation of gene expression: a hidden layer of complexity. Caister Academic, Poole, UK, pp 29–44
19. Cesarini E, Mozzetta C, Marullo F, Gregoretti F, Gargiulo A, Columbaro M, Cortesi A, Antonelli L, Di Pelino S, Squarzoni S, Palacios D, Zippo A, Bodega B, Oliva G, Lanzuolo C (2015) Lamin A/C sustain PcG proteins architecture maintaining transcriptional repression at target genes. J Cell Biol 211(3):533–551
20. Wang X, Huang D, Xu H (2010) An efficient local Chan-Vese model for image segmentation. Pattern Recogn 43:603–618
21. Osher S, Sethian JA (1988) Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. J Comput Phys 79:12–49
22. El-Zaart A (2010) Images thresholding using ISODATA technique with Gamma Distribution. Pattern Recogn Image Anal 20(1):29–41
23. Ball GH, Hall DJ (1965) A novel method of data analysis and pattern classification. Technical report, Stanford Research Institute, Menlo Park, CA
24. Gonzalez R, Woods R (2006) Digital image processing, 3rd edn. Prentice-Hall, Upper Saddle River, NJ
25. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675
26. ImageJ – Home page. http://imagej.nih.gov/ij/
27. 3DObjects Counter. http://fiji.sc/3D_Objects_Counter
28. LibTIFF – TIFF library and utilities. http://www.remotesensing.org/libtiff/

Part IV

Chapter 17

Polymer Physics of the Large-Scale Structure of Chromatin

Simona Bianco, Andrea Maria Chiariello, Carlo Annunziatella, Andrea Esposito, and Mario Nicodemi*

Abstract

We summarize the picture emerging from recently proposed models of polymer physics describing the general features of chromatin large scale spatial architecture, as revealed by microscopy and Hi-C experiments.

Key words Chromatin, Polymer physics, Microscopy, Chromosome

1 Introduction

Chromosomes in the cell nucleus of eukaryotes are organized in a complex three-dimensional (3D) structure. As shown by microscopy, they occupy distinct nuclear regions, called chromosomal territories [1, 2]. More recently, different approaches based on chromosome conformation capture technique (3C), such as Hi-C [3], have allowed to derive detailed genome-wide frequencies of physical contacts between all genomic regions. These technologies and their applications are reviewed in the chapters below; in particular, 3C methods for Drosophila are discussed in Chapter 18 by H.-B. Li, 3C and 4C methods in mammals respectively in Chapters 19 and 20 by A. Cortesi and B. Bodega, and by M. Matelot and D. Noordermeer, and the corresponding associated data analysis techniques in Chapter 21 by B. Leblanc, I. Comet, F. Bantignies, and G. Cavalli. The emerging picture is an intricate network of interactions with interesting properties. A striking discovery is that each chromosome is partitioned into Mb-sized topologically associated domains (TADs), characterized by enriched levels of internal interactions [4, 5]. However, the origin of TADs and their underlying mechanisms are still unclear. More generally, a fundamental open challenge in molecular biology is the understanding of the nucleus structure, the origin of the observed patterns, the factors that shape them and how they are regulated by the cell for functional

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_17,

purposes. Beyond their conceptual interest, these issues have important practical implications, as disruptions of chromosome folding are linked to a number of diseases, including cancer [6].

2 Patterns in Hi-C Data

Hi-C data have shown, in particular, that the average contact probability between pairs of loci decreases with their genomic distance, with a power law decay at least within the 0.5–7 Mb range [3]. However, the measured power law exponent, α, is different in different systems (Fig. 1; [7]). Human chromatin from embryonic stem cells (ESCs) has, for example, on average $\alpha \sim 1.6$ [7], while human lymphoblastoid cells at interphase have $\alpha \sim 1.1$ [3] and human chromosomes in metaphase from a cancer (HeLa) cell line have $\alpha \sim 0.5$ [8]. Furthermore, the exponent α varies widely in different species: in mouse ESCs it is found $\alpha \sim 0.7$–0.9 [5], in Drosophila $\alpha \sim 0.7$–0.8 [9], in Yeast $\alpha \sim 1.5$ [10]. Actually, even in a given cell system different chromosomes can have different exponents [7, 11]. All these observations have contradicted the early idea that a single universal architecture, originally contemplated in the Fractal Globule model [3], where $\alpha = 1$, might describe chromosome folding.

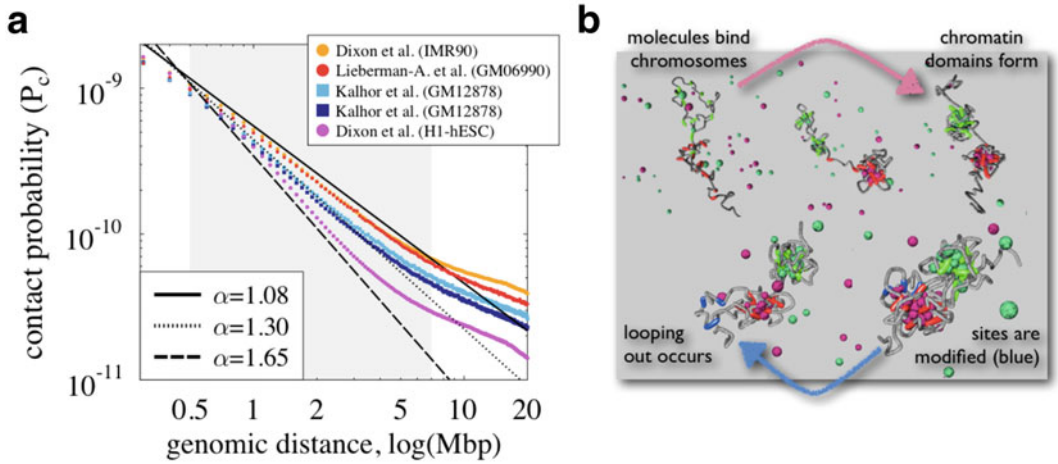


Fig. 1 Chromosome contact probability and the SBS model. (**a**) The average contact probability, *Pc(s)*, from Hi-C data reveals a pattern of chromatin interactions extending across genomic scales. In the 0.5–7 Mb range, *Pc(s)* can be fitted by power-laws with an exponent α. Independent experiments across different human cell lines give varying exponents reflecting distinct chromatin behaviors in different system. (**b**) In the SBS model, here in a pictorial view, chromatin is represented as a polymer chain having binding sites for diffusing molecules. In a mixture of open and compact SBS polymers, the contact probability, *Pc(s)* and its exponent α, depend on the mixture composition. In particular, the entire range of Hi-C data exponents can be recovered (superimposed lines in Fig. 1a)

3 Polymer Models of Chromatin

An alternative approach has emerged from polymer physics, which may help understanding the variety of observed chromosome conformations. The basic idea is that chromatin folding is driven by its interactions with other nuclear elements, such as DNA-binding molecules. This kind of scenario has been considered, for example, by the strings and binders switch (SBS) model [7, 12] and later on in the dynamic loop (DL) model [13, 14].

The simplest version of the SBS model describes a chromatin filament as a self-avoiding walk (SAW) polymer chain having specific binding site for diffusing molecules. Different chromatin folding states are stable thermodynamic phases. The switch between conformations, in turn, can be controlled by standard biochemical strategies, such as protein upregulation or chromatin modification. The SBS can be extended to take into account further complications, for instance crowding and off-equilibrium effects as encountered in real complex fluids [15–19]. An interesting aspect of the model is the finding of a variety of transient conformations, including, under special conditions, the fractal globule.

4 Chromatin is a Mixture of Differently Folded Regions

As predicted by polymer physics, the model has three major folding classes, independently of the specific biochemical details (which set the class a polymer folds into): the open polymer state, the compact globule state, and the Θ-point state at the transition between the open and compact states. The exponent is $\alpha \sim 2.1$ in the open state, where the polymer is randomly folded, $\alpha \sim 1.5$ in the Θ-point state, and $\alpha \sim 0$ in the compact state. In this way, the range of effective exponents found experimentally in different chromosomes, cell cycle stages, and cell types, can easily explained by mixture systems made of different fractions of open and compact (and Θ-point) conformations [7], as seen in Fig. 1.

The SBS model is also able to describe the behavior of the mean square distance, R^2, with the genomic distance between two loci in a chromosome and, importantly, its plateauing at large genomic separations [7]. Another significant quantity for direct comparison of data and modeling is also the distance moment ratio, $<R^4>/<R^2>^2$. This dimensionless number is useful to distinguish between different folding states of chromosomes. The SBS model predict that $<R^4>/<R^2>^2$ is ~1.5 in both the open and compact states, while a sharp peak, up to ~5, is found in the Θ-point state. Interestingly, the same range is found in FISH experiments, indicating that chromosome conformations in cells include, beside open and compact regions, corresponding to euchromatin and heterochromatin, another kind of conformation close to the Θ-point [7].

5 Origin of Topological Associating Domains (TADs)

The SBS model can naturally explain the "topological domains" (TADs) of chromatin, for example through specialized binding sites (Fig. 2). In addition, changes in specialization of binding sites, upon domain formation, con conduct to chromatin looping events, observed in experiments across short time scales [7]. The binding specificity at distinct loci or domains can be obtained, for example, by a combination of single molecular factors. Even more complex architectures, with nested layers of organization could be easily achieved. Other mechanisms could contribute to the chromatin folding in compact states, like supercoiling [20] and plectoneme formation [21]. The above results, illustrated within the SBS model, are supported by similar findings in the DL model. In particular, the DL model has been used to frame the effects of entropy in mixtures of long, looped polymers, which may result in effective repulsive forces [13, 14]. The above findings explained within the SBS model are supported by analogous results in the DL model. The DL model has been used to enclose the effects of entropy in mixtures of long looped polymers, which may result in effective repulsive forces [13, 14].

6 Discussion

The emerging new picture is that chromatin [7] is a complex mixture of diversely folded regions regulated by simple and basic mechanisms of polymer physics. In that picture folding classes of

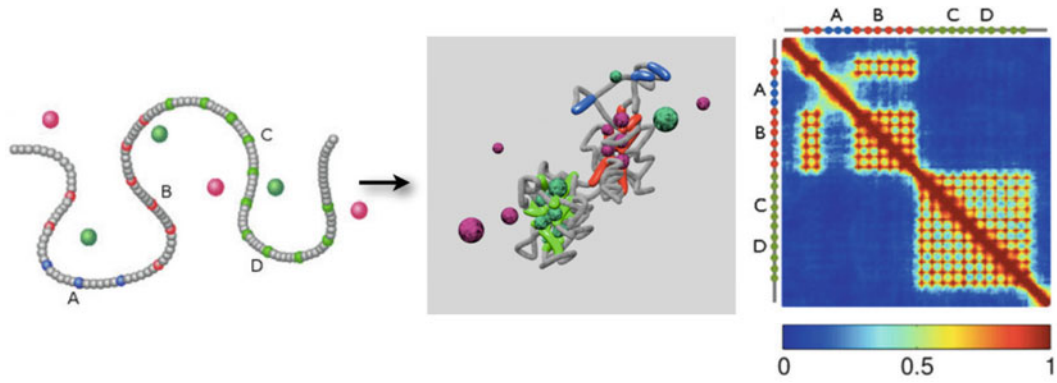


Fig. 2 Topological domains and the SBS model. In the SBS model, as different types of binding sites are considered, here shown as *red* and *green* polymer sites (*left*), different domains naturally emerge (*center*, see also Fig. 1b). Looping out of a given locus from one of the domains can result from binding site specialization, in this case, through the loss of affinity for the red binding molecules represented by a change in polymer sites from *red* to *blue*. The corresponding SBS contact matrix (*right*) closely resembles those observed for chromatin topological domains from experiments

chromatin correspond to stable emergent phases. The contact probability and its exponent found by Hi-C experiments is an average, since specific loci can be folded in different conformations across a population of cells, and different loci can be folded in different conformations inside a single cell. Other quantities, such as also the distance moment ratio $<R^4>/<R^2>^2$, can help better dissecting the folding states of chromatin. The SBS model explains some basic physical mechanisms of self-organization and can describe the range of available data from FISH and Hi-C, e.g., the contact probability, the mean square distance, moment ratios, the formation of TADs, and much more.

Models like the SBS, informed with biological details, can also be used for prediction of the folding of specific genomic loci. We, and others, currently employ high-resolution single-cell mapping of interactions across single loci, together with SBS modeling of epigenetics features to reveal some of folding determinants. Our ongoing work supports a picture where a deep connection exists between chromatin architecture and transcription. A variant of the SBS model has also been used to depict and make predictions about the folding of the *Xist* locus [22], and meiotic chromosome recognition [23]. SBS models can be also used for "inverting" Hi-C contact matrices, i.e., to find from Hi-C data the corresponding spatial conformation of chromatin and the involved key interaction regions. As recent technological advances are revealing the complexity and extent of genomic organization [24], the combination of experiments and models promises to return a deeper understanding of the principles of genome function and, in the future, of associated diseases.

References

1. Lanctot C, Cheutin T, Cremer M, Cavalli G, Cremer T (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat Rev Genet 8:104–115
2. Pombo A, Branco MR (2007) Functional organisation of the genome during interphase. Curr Opin Genet Dev 17:451–455
3. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstorm R et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293
4. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376–380
5. Nora EP et al (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485:381–385
6. Misteli T (2011) The inner life of the genome. Sci Am 304:66
7. Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, Nicodemi M (2012) Complexity of chromatin folding is captured by the Strings & Binders Switch model. Proc Natl Acad Sci U S A 109:16173–16178
8. Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J (2013) Organization of the mitotic chromosome. Science 342:948–953
9. Sexton T et al (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148:458–472
10. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau

CA, Noble WS (2010) A three-dimensional model of the yeast genome. Nature 465: 363–367

11. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotechnol 30:90–98
12. Nicodemi M, Prisco A (2009) Thermodynamic pathways to genome spatial organization in the cell nucleus. Biophys J 96:2168–2177
13. Bohn M, Heermann DW (2010) Diffusion-driven looping provides a consistent framework for chromatin organization. PLoS One 5:e12218
14. Tark-Dame M, van Driel R, Heermann DW (2011) Chromatin folding—from biology to polymer models and back. J Cell Sci 124: 839–845
15. Cataudella V et al (1994) Critical clusters and efficient dynamics for frustrated spin models. Phys Rev Lett 72:1541–1544
16. Tarzia M et al (2004) Glass transition in granular media. Europhys Lett 66:531–537
17. Coniglio A, Nicodemi M (2000) The jamming transition of granular media. J Phys Cond Matt 12:6601–6610
18. Coniglio A et al (1998) Segregation of granular mixtures in presence of compaction. Europhys Lett 43:591–597
19. Nicodemi M, Jensen HJ (2001) Creep of superconducting vortices in the limit of vanishing temperature: a fingerprint of off-equilibrium dynamics. Phys Rev Lett 86: 4378–4381
20. Benedetti F, Dorier J, Burnier Y, Stasiak A (2014) Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. Nucleic Acids Res 42:2848–2855
21. Le TBK, Imakaev MV, Mirny LA, Laub MT (2013) High-resolution mapping of the spatial organization of a bacterial chromosome. Science 342:731–734
22. Scialdone A et al (2011) Conformation regulation of the X chromosome inactivation center: a model. PLoS Comput Biol 7:e1002229
23. Nicodemi M, Panning B, Prisco A (2008) A thermodynamic switch for chromosome colocalization. Genetics 179:717–721
24. Bickmore W, van Steensel B (2013) Genome architecture: domain organization of interphase chromosomes. Cell 152:1270–1284

Chapter 18

Chromosome Conformation Capture in Drosophila

Hua-Bing Li*

Abstract

Linear chromatin fiber is packed inside the nuclei as a complex three-dimensional structure, and the organization of the chromatin has important roles in the appropriate spatial and temporal regulation of gene expression. To understand how chromatin organizes inside nuclei, and how regulatory proteins physically interact with genes, chromosome conformation capture (3C) technique provides a powerful and sensitive tool to detect both short- and long-range DNA–DNA interaction. Here I describe the 3C technique to detect the DNA–DNA interactions mediated by insulator proteins that are closely related to PcG in Drosophila, which is also broadly applicable to other systems.

Key words Chromosome conformation capture, DNA–DNA interaction, Chromatin, Insulator proteins, Polycomb

1 Introduction

Each cell contains approximately 2 m long DNA, which is folded thousands of times to fit into a nuclei of 6 μm in diameter. The tight packaging by chromatin-associated factors is highly regulated and plays important roles in the appropriate spatial and temporal regulation of gene expression. To understand how chromatin organizes inside nuclei, and how regulatory proteins physically interact with genes, the microscopy-based fluorescence in situ hybridization (FISH) and in vivo live imaging, and the molecular technique of chromosome conformation capture (3C), have been developed to help us understand the higher order chromatin structures, such as Polycomb bodies and transcription factories [1].

FISH has been extensively used to probe the chromosome structure since its first introduction in 1969 [2]. The principle of current FISH protocol is simple: the fluorescent probes recognize and hybridize to their endogenous sequences, which then could be visualized under microscope. The FISH technique has been successfully used to address many important questions, such as the positions of genes on chromosomes, the discovery of chromosome

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_18,

territories, diagnosis of chromosomal abnormalities, and DNA long-distance interaction in interphase. The advantage of this technique is that it provides information on the frequency of interaction in a single-cell level. However, the protocol is complex, and it is difficult to obtain high-quality fluorescence images, and based on fixed cells, in addition it has poor resolution (usually >0.2 μm), which prevents its use in the detection of relatively short-distance DNA-DNA interaction.

To define roles of the DNA elements and protein factors that may mediate Polycomb body formations in Drosophila, we adapted and developed live-imaging systems and 3C techniques to detect DNA–DNA interactions [3, 4]. The live-imaging system was first developed by Belmont and colleagues in yeast, based on the specific and tight binding of the bacterial lac repressor (LacI) to its target sequence lac operator (LacO) [5, 6]. The advantage of the LacO/LacI live imaging is that you could study the chromosomal dynamics under its physiological conditions, and acquire time-lapse images of single cells. However, the problem with this technique is that the GFP could be easily photo-bleached, rendering the acquisition of high-quality images very difficult.

To cross validate the observations from live-imaging systems that Polycomb response elements (PREs) from different chromosomes co-localize with each other, we also adapted 3C to study the long- as well as short-range interactions between PREs in Drosophila imaginal discs. The technique of 3C was first invented by Dekker and colleagues to study the conformation of a complete chromosome in yeast [7], and was quickly adapted to investigate the interactions between complex gene loci and loops. Within a few years, it soon gained popularity, evolved into several 3C-based variants and became a standard research tool to study chromosomal interactions [8]. The principle of 3C technology is based on the formaldehyde fixation of cells which cross-link the protein–DNA complexes in close proximity, followed by proper restriction digestion, and then ligation in diluted condition in favor of intramolecular ligation. After reversion of cross-links, the ligation products are quantified by PCR, with PCR primers specific for the fragments of interest. This technique could be used to study the genomic region of any size, and between different chromosomes. However, to draw meaningful conclusions, proper controls must be implemented in the experiment design [9]. 3C and 3C-based technology only provide information about the DNA interaction frequency of a large population of cells. So, it is beneficial and complementary to use both 3C and FISH or live imaging to draw meaningful conclusions.

2 Materials

Prepare all solutions using ultrapure water, and store at appropriate temperatures. Maintain a clean area to avoid DNA contamination when doing PCR and running agarose electrophoresis gels.

1. 1.25 M Glycine, store at 4 °C.
2. Lysis buffer: 10 mM Tris–HCl (pH 8.0), 10 mM NaCl, 0.2% NP-40, Roche protease inhibitor cocktail added freshly.
3. 10× NEBuffer EcoRI: 1000 mM Tris–HCl (pH 7.5 at 25 °C), 500 mM NaCl, 100 mM $MgCl_2$, 0.25% Triton X-100. Aliquot and store at −20 °C freezer. Dilute to 1.2× NEBuffer EcoRI with dH_2O.
4. Nuclei lysis buffer: 450 µl 1.2× EcoR1 buffer + 15 µl 10% SDS.
5. 10× Ligation buffer: 400 mM Tris–HCl (pH 7.8 at 25 °C), 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP. Aliquot and store at −20 °C freezer.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

Day 1

1. Dissect the imaginal discs from about 50 third instar lava under the dissection scope. Include all discs and brain parts and transfer into 5 ml cold PBS; let sit in ice. This step should be done as quickly as possible; it needs a lot of practice on dissection technique (*see* **Note 1**).
2. Add 285 µl 37% formaldehyde (final concentration of 2% of formaldehyde), and rotate at RT for 15 min.
3. Add 0.588 ml 1.25 M glycine (final concentration of 0.125 M of glycine) to quench fixation, mix, and cool down on ice for 5 min.
4. Spin 300 × *g* for 6 min at 4 °C, and remove supernatant.
5. Resuspend in 5 ml cold lysis buffer, incubate for 10 min on ice, followed by 20 strokes inside a 5 ml Dounce homogenizer; let sit on ice for 15 min again.
6. Centrifuge for 6 min at 600 × *g* at 4 °C and remove the supernatant.
7. Wash the pelleted nuclei with 400 µl 1.2× EcoR1 buffer once, transfer the nuclei to 1.7 ml tube, spin 500 × *g*/5 min, and remove supernatant (*see* **Note 2**).

8. Prepare nuclei lysis buffer and preheat to 37 °C, and use it to resuspend the pelleted nuclei (final 0.3% SDS).
9. Incubate and shake at 300 rpm at 37 °C for 1 h. If aggregates form, gently pipette up and down, and avoid the formation of bubbles.
10. Add 50 μl 20% Triton X-100 to neutralize SDS (final 2% Triton X-100), at 37 °C shake 300 rpm for 1 h. (Take 5 μl for digestion efficiency analysis: un-digested, keep in 4 °C temporarily, then **step 16**.)
11. Add 10 μl NEB EcoR1 (20 U/μl), and shake 300 rpm at 37 °C overnight (*see* **Note 3**).

Day 2

12. Add 80 μl of 10% SDS, at 65 °C for 25 min. (Take 5 μl for digestion efficiency analysis: digested, keep at 4 °C, then **step 16**.)
13. Aliquot 100 μl per tube and store at −20 °C freezer. Only need one 100 μl aliquot for one 3C analysis.
14. Add 70 μl 10× ligation buffer to the 100 μl sample, and 630 μl ddH_2O to make up 800 μl (*see* **Note 4**).
15. Add 40 μl of 20% triton, and incubate at 37 °C for 1 h.
16. Lower temperature by incubation in ice for 5 min, add 4 μl of 100 mM ATP, then add 10 μl ligase, and ligate at 16 °C for 4 h, and then at RT for 1 h.
17. After ligation, add 10 μl of 10 mg/ml proteinase K; also include the digestion control in **steps 10** and **11**. Incubate at 65 °C overnight to de-cross-link (*see* **Note 5**).

Day 3

18. Add 10 μl of 1 mg/ml RNase, and incubate for 45 min at 37 °C.
19. Add 0.8 ml of phenol–chlorophorm, mix vigorously, and centrifuge for 15 min at 13,000 rpm (16,000g or maximum speed of microcentrifuge) at RT.
20. Take 0.8 ml of supernatant, add 0.8 ml of chlorophorm, mix, and centrifuge at 13,000 rpm/15 min.
21. Take 750 μl supernatant, add 2 μl glycogen and 70 μl 3 M NaAc, split into two tubes, 400 μl each, add 1000 μl 100% ethanol, and mix at −70 °C 1 h.
22. Centrifuge for 20 min at top speed at 4 °C, and remove supernatant.
23. Use 800 μl 70% ethanol to wash the pellet once, and resolve in the following digestion solution to linearize the ligated circles (*see* **Note 6**):

10× NEB buffer **3**:	2 μl
100× BSA:	0.2 μl
dH_2O:	15.8 μl
Pst1:	2 μl

24. Incubate at 37 °C for 4 h. Then add 2 μl RNase and incubate at 37 °C for another 1 h.
25. Add 130 μl TE buffer, and 150 μl phenol–chlorophorm. Mix, and spin for 15 min at RT.
26. Take the supernatant, and add 150 μl phenol–chlorophorm to extract again.
27. Combine the supernatant, add 300 μl chlorophorm, mix, and spin for 10 min.
28. Add 25 μl 3 M NaAc, 2 μl glycogen, and then 900 μl ethanol, and mix at −70 °C for 1 h.
29. Spin at maximum speed for 20 min. Wash pellet, and resolve in 20 μl 10 mM Tris buffer. The 3C DNA could be used for the following 3C-PCR, and could be stored at −20 °C freezer for months (*see* **Note 7**).
30. 3C PCR primers should be optimal and efficient, and closed to the restriction sites. It is better to use qPCR to quantify the ligated DNA products. Importantly, proper controls should be used, and ref. [9] provides excellent explanations on how to make the controls (*see* **Notes 8** and **9**).

4 Notes

1. Dissect specific imaginal discs types; it may decrease background, but take much longer time to get enough cells to work with. If work with culture cells, 10 million cells will be enough and fixation could be carried out on the culture plate.
2. Choose proper restriction enzyme: first need to consider the bait sequence so that the restriction sites contain the entire interested region. Second, choosing the commonly used restriction enzyme will get higher cutting efficiency.
3. Digestion should not go over 18 h; otherwise the fixation may be reversed.
4. It is important to dilute the cutting mixture, so that the ligation only happens in between DNAs that are bound by specific proteins.

5. Determine digestion efficiency: should use two or more pairs of primers across randomly picked restriction sites, and compare the PCR product abundance of cut samples versus uncut samples by qPCR. Usually around 80 % cut efficiency is enough.
6. Proper enzymes need to be selected that are cut in the middle of the bait sequence; here my bait Mcp PRE is cut by Pst1; this step is optional.
7. The resulting 3C-DNA could be diluted and used for many PCR reactions depending on the PCR primer efficiency and the strength of the 3C interactions.
8. Refs. [3] and [4] have the example results and all the primer information used for PRE 3C experiments.
9. The protocol provided here could be easily modified for 4C experiments, which provides un-biased genome-wide information of interactors of the bait sequence. While for 3C experiment, we need to guess the potential interactions between the bait and the interactor.

Acknowledgement

I sincerely thank Dr. Vincenzo Pirrotta for his guidance to develop the technique in his lab at the department of Molecular Biology and Biochemistry of Rutgers University.

References

1. Pirrotta V, Li HB (2012) A view of nuclear Polycomb bodies. Curr Opin Genet Dev 22(2): 101–109. doi:10.1016/j.gde.2011.11.004
2. Gall JG, Pardue ML (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci U S A 63(2):378–383
3. Li HB, Ohno K, Gui H, Pirrotta V (2013) Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. PLoS Genet 9(4):e1003436. doi:10.1371/journal.pgen.1003436
4. Li HB, Muller M, Bahechar IA, Kyrchanova O, Ohno K, Georgiev P, Pirrotta V (2011) Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster. Mol Cell Biol 31(4):616–625. doi:10.1128/MCB.00849-10
5. Robinett CC, Straight A, Li G, Willhelm C, Sudlow G, Murray A, Belmont AS (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol 135(6 Pt 2):1685–1700
6. Straight AF, Belmont AS, Robinett CC, Murray AW (1996) GFP tagging of budding yeast chromosomes reveals that protein–protein interactions can mediate sister chromatid cohesion. Curr Biol 6(12):1599–1608
7. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. doi:10.1126/science.1067799
8. Dekker J (2014) Two ways to fold the genome during the cell cycle: insights obtained with chromosome conformation capture. Epigenetics Chromatin 7:25. doi:10.1186/1756-8935-7-25
9. Dekker J (2006) The three 'C's of chromosome conformation capture: controls, controls, controls. Nat Methods 3(1):17–21. doi:10.1038/Nmeth823

Chapter 19

Chromosome Conformation Capture in Primary Human Cells

Alice Cortesi and Beatrice Bodega*

Abstract

3D organization of the genome, its structural and regulatory function of cell identity, is acquiring prominent features in epigenetics studies; more efforts have been done to develop techniques that allow studying nuclear structure. Chromosome conformation capture (3C) has been set up in 2002 from Dekker and from that moment great investments were made to develop genomics variants of 3C technology (4C, 5C, Hi-C) providing new tools to investigate the shape of the genome in a more systematic and unbiased manner. 3C method allows scientists to fix dynamic and variable 3D interactions in nuclear space, and consequently to study which sequences interact, how a gene is regulated by different and distant enhancer, or how a set of enhancer could regulate transcriptional units; to follow the conformation that mediates regulation change in development; and to evaluate if this fine epigenetic mechanism is impaired in disease condition.

Key words Chromosome conformation capture (3C), 3D interactions, Nuclear structure, Primary human cells

1 Introduction

Structure of the cell nucleus has sparked interest from years. The development of chromosome conformation capture (3C) technology has allowed to deeply investigate genome topology [1, 2].

3C approach is a biochemical method that enables to study the three-dimensional organization of the genome in cells. Using formaldehyde, physical interactions are fixed; a subsequent digestion leads the formation from interacting sequences of free ends that are in close proximity. Then a ligation reaction in diluted conditions is a disadvantage for intermolecular ligation, promoting intramolecular fusions (Fig. 1). As a result, a collection of products are generated that are composed of DNA fragments that were originally physically near together in the nuclear space. Using a bait primer on the region of interest and the other one on the supposing region of interaction, in a simple PCR it could be possible to visualize the band, if the interaction occurs [2].

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_19,

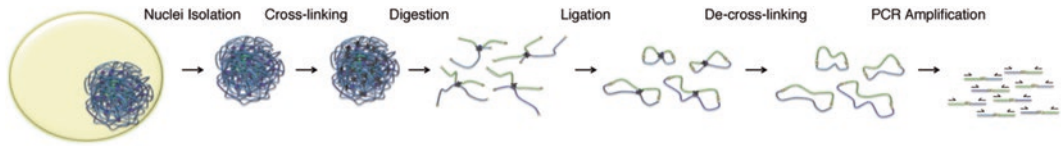


Fig. 1 Chromosome conformation capture (3C) procedure. Schematic representation of 3C steps: isolation of nuclei, cross-linking of 3D interactions, digestion with a specific restriction enzyme, creating free ends, ligation of them to obtain interacting sequences, purification, and PCR amplification to evaluate the occurrence of a specific interaction

First study, using 3C to explore the three-dimensional organization of chromosomes at high resolution, describes intrachromosomal interactions between telomeres as well as interchromosomal interactions between centromeres and between homologous chromosomes in yeast [2]. The idea that the 3D genome topology is in some way functional to the regulation of nuclear activities takes place and a lot of work is done in this direction. 3C method was applied to analyze physical connections between genes and *cis* enhancers (mouse and human β-*globin* locus model [3–5]). Then, highly specific associations between loci located on separate chromosomes are described. These *trans*-interactions can be between a distant enhancer with different target genes (olfactory receptor genes [6]). In other cases, *trans*-interactions appear to play a role in a higher level of gene control to coordinately regulate multiple loci with a set of both intrachromosomal and interchromosomal interactions (T helper 2 cytokine locus [7]), or provide additional levels of gene regulation by allowing combinatorial association of genes and sets of regulatory elements (imprinted loci [8]). It is also reported that specific DNA-binding sites could mediate the formation of this topologically complex structure (Polycomb response elements in Drosophila [9]). Finally, it is demonstrated that 3D structure could also have a role in developmental processes (mammalian X-chromosome inactivation [10, 11]).

Here, we describe the 3C technique adapted to study primary human cells, allowing to capture 3D chromatin interactions.

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents.

Cell lysis buffer: 10 mM Tris–HCl pH 8.0, 10 mM NaCl, 5 mM $MgCl_2$, 100 μM EGTA. Before the use complement with protease inhibitors, Protease Inhibitor Cocktail (PIC) and phenylmethanesulfonylfluoride (PMSF).

Cross-linking buffer: 10 mM Tris–HCl pH 8, 50 mM NaCl, 10 mM $MgCl_2$. Before the use complement with protease inhibitors, PIC and PMSF, and 1 mM dithiothreitol (DTT).

Formaldehyde solution 36.5–38 % in H_2O.

Glycine 2.5 M.

Phosphate-buffered saline (PBS) 1× (pH 7.4).

Sodium dodecyl sulfate (SDS) 10 %.

Digestion buffer: Buffer of the specific restriction enzyme diluted at 1.5×. Before the use complement with 0.3 % SDS.

Triton X-100 100 %.

Ligation buffer (1.15×): 57 mM Tris–HCl pH 7.5, 11 mM $MgCl_2$, 11 mM DTT, 1.1 mM ATP, 1 % Triton X-100. Prepare at the time of use.

T4 DNA ligase.

Phenol-chloroform-isoamyl alcohol 25:24:1.

Ethanol (100 and 70 %).

3 M Sodium acetate pH 5.2.

Glycogen.

Proteinase K.

RNase cocktail (500 U/ml of RNase A and 20,000 U/ml of RNase T1).

Resuspension buffer: 10 mM Tris–HCl pH 7.5.

Qubit dsDNA Broad Range Assay Kit.

3 Methods

3.1 Nuclei Isolation

1. Grow cells in appropriate culture conditions.
2. Count cells using an automated cell counter. Consider to use 3.5×10^6 cells for each experiment.
3. Collect cells by centrifugation at 200 × *g* for 5 min at room temperature.
4. Resuspend cells, pipetting up and down, in 10 ml of cold cell lysis buffer (*see* **Note 1**).
5. Incubate on ice for 10 min.
6. Check the lysis at microscope (*see* **Note 2**).

3.2 Cross-Linking

1. Collect nuclei by centrifugation at 370 × *g* for 5 min at 4 °C.
2. Resuspend nuclei in 5 ml of cross-linking buffer.
3. Cross-link by the addition of 270 µl of 36.5–38 % formaldehyde solution (2 %) (*see* **Note 3**).

4. Incubate for 10 min at room temperature on a shaker.
5. Quench the reaction by the addition of 250 µl of 2.5 M glycine (125 mM).
6. Incubate for 5 min at room temperature on a shaker.
7. Collect cross-linked nuclei by centrifugation at 370 × *g* for 10 min at room temperature.
8. Wash twice with ice-cold PBS 1×.
9. Collect cross-linked nuclei by centrifugation at 370 × *g* for 5 min at 4 °C.

3.3 Digestion

1. Resuspend cross-linked nuclei in 500 µl of digestion buffer.
2. Incubate for 1 h at 37 °C with bland agitation.
3. Add 110 µl of 100% Triton X-100 (1.8%) to sequester SDS.
4. Incubate for 1 h at 37 °C with bland agitation.
5. Collect 60 µl aliquot of the sample and label as undigested genomic DNA control (UND) (Fig. 2a).
6. Digest with 500 U of restriction enzyme at appropriate conditions overnight with bland agitation (*see* **Note 4**).
7. Collect 60 µl aliquot of the sample and label as digested genomic DNA control (DIG) (Fig. 2a).
8. Inactivate the restriction enzyme by the addition of 95 µl of 10% SDS (1.6%).
9. Incubate for 25 min at 65 °C with bland agitation.

3.4 Ligation

1. Transfer the reaction in a 50 ml tube.
2. Dilute DNA to a final concentration of 2.5 ng/µl in 10 ml of ligation buffer (*see* **Note 5**).

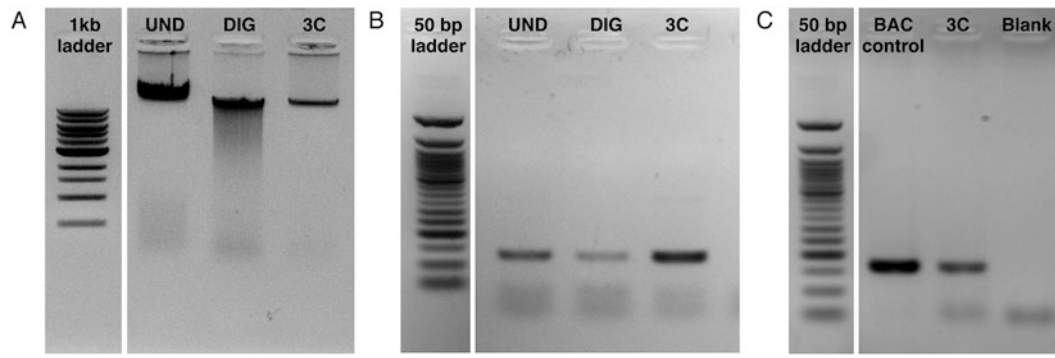


Fig. 2 3C controls. (**a**) Quality check of samples: UND, cross-linked genomic DNA, DIG, digested cross-linked DNA, and 3C, ligated cross-linked DNA after digestion. (**b**) Quality check on a specific restriction site: PCR product is positive in UND sample, less efficient in DIG sample, and recovered in 3C sample. (**c**) Chimeric PCR on a specific interaction in 3C sample and BAC (positive control)

3. Incubate for 1 h at 37 °C with bland agitation.
4. Add 2000 U of T4 DNA ligase.
5. Incubate for 8 h at 16 °C with bland agitation.
6. Leave sample at room temperature for 30 min.
7. Treat with 1 mg of proteinase K.
8. Incubate O.N. at 65 °C with constant agitation to reverse the formaldehyde cross-links.

3.5 Purification

1. Add 25 μl of RNase cocktail.
2. Incubate for 40 min at 37 °C with bland agitation to remove RNA.
3. Purify DNA, adding one volume of phenol-chloroform-isoamyl alcohol, vortexing for 2 min, and centrifuging at 3.220 × *g* for 15 min at RT.
4. Transfer the supernatant to a new tube.
5. Dilute 1:2 in bi-distillated water.
6. Precipitate with three volumes of 100 % ethanol, 1/10 volume of 3 M sodium acetate pH 5.2, and 10 μl of glycogen, leaving at −80 °C for at least 1 h.
7. Centrifuge at 1.810 × *g* for 45 min at 4 °C.
8. Discard the supernatant.
9. Wash the pellet with 10 ml of 70 % ethanol.
10. Centrifuge at 1.810 × *g* for 15 min at 4 °C.
11. Air-dry the pellet.
12. Dissolve in 100 μl of resuspension buffer.
13. Quantify DNA by Qubit dsDNA Broad Range Assay Kit.

3.6 Purification of Control Aliquots (UND and DIG)

1. Process aliquots of UND and DIG, bringing volume to 100 μl with resuspension buffer.
2. Add 1 μl of RNase cocktail.
3. Incubate for 30 min at 37 °C with bland agitation to remove RNA.
4. Treat with 100 μg of proteinase K.
5. Incubate for 1 h at 65 °C with constant agitation to reverse the formaldehyde cross-links.
6. Purify DNA, adding one volume of phenol-chloroform-isoamyl alcohol, vortexing for 2 min, and centrifuging at 9.390 × *g* for 4 min at RT.
7. Transfer the supernatant to a new tube.
8. Precipitate with two volumes of 100 % ethanol, 1/10 volume of 3 M sodium acetate pH 5.2, and 1 μl of glycogen, leaving at −80 °C for at least 1 h.

9. Centrifuge at 18.400 × *g* for 30 min at 4 °C.
10. Discard the supernatant.
11. Wash the pellet with 100 µl of 70% ethanol.
12. Centrifuge at 18.400 × *g* for 5 min at 4 °C.
13. Air-dry the pellet.
14. Dissolve in 20 µl of resuspension buffer.
15. Quantify DNA by Qubit dsDNA Broad Range Assay Kit.
16. Perform PCR amplification using a Taq DNA polymerase with provided reagents in 50 µl of reaction, using 0.3 µM of primers and 10 ng of template (*see* **Note 6**).

3.7 PCR

1. Design primers so that they will be in close proximity (proximal to 100–50 bp) to the restriction site for the regions to be amplified (*see* **Note 7**).
2. Select BACs covering the regions that are supposed to interact, and quantify them by Qubit dsDNA Broad Range Assay Kit. Digest equimolar amounts of different BACs with the restriction enzyme of interest, then mix, and ligate in 50 µl; after a step of purification by phenol extraction and ethanol precipitation, resuspend the BAC mix in 50 µl of resuspension buffer, obtaining the BAC control. It represents the PCR template that contains all possible ligation products that are relevant for the genomic regions of interest, and as such is considered a positive control.
3. Perform PCR amplification using a Taq DNA polymerase with provided reagents in 50 µl of reaction, using 0.3 µM of primers and 25 ng of template (*see* **Note 8**). Remember to perform the PCR on 3C sample and BAC control in parallel, using for the BAC control 1/10 of the quantity used for 3C sample.
4. Run PCR products on 2% agarose gels, stained with ethidium bromide.
5. Quantify with the Image J (*see* **Note 9**).

4 Notes

1. 3C method could be performed using directly intact cells or nuclei. It depends on the cellular type in study and also on the percentage of cross-linking that you use. Indeed if you use cells with a great cytoskeleton or you need to cross-link with high percentage of formaldehyde to reveal less frequent interactions, it could be better to perform 3C experiment on isolated nuclei to promote next steps of the procedure and consequently the detection of higher signals of interaction.

2. If the lysis was incomplete you could use a Dounce homogenizer, performing approximately 15 strokes.
3. Formaldehyde is used to cross-link protein–protein and protein–DNA interactions by means of their amino and imino groups. Advantages of this cross-linking agent are that it works over a relatively short distance (2 Å) and that cross-links can be reversed at higher temperatures [12–14]. The percentage of formaldehyde used depends on the frequency and stability of the interactions analyzed and has to be set for every new 3C experiment.
4. The restriction enzyme used depends on the locus to be analyzed. If small loci (<10–20 kb) are studied it is needed to use frequently cutting restriction enzymes, while larger loci are investigated, six-base cutters (six-cutters) can also be used. It has to be considered that a high percentage of cross-linking could decrease digestion efficiency [15]. Indeed it is recommend to be sure that at least 60–70 % of the DNA, but preferably 80 % or more, be digested before continuing with the ligation step [16].
5. Ligation reaction has to be performed in a condition of diluting DNA (2.5 ng/μl), in a way that intramolecular ligation events between cross-linked DNA fragments are favored. Moreover, independently of the restriction site analyzed, two types of junction are always over-represented. The first most abundant junction is with the neighboring DNA sequences, resulting from incomplete digestion, and can constitute up to 20–30 % of all the junctions. The second most abundant junction is with the other end of the same restriction fragment, after circularization, and can account for up to 5–10 % of all the junctions. The formation of other junctions from fragments that are close together in the nuclear space represents only 0.2–0.5 % of the junctions and with increasing genomic site separation decreases to <0.1 %. It is therefore clear that to accurately quantify such rare events, it is necessary to include many genome equivalents in a PCR reaction [16].
6. Biological parameters, such as the heterogeneity of the cells, and technical parameters, as different efficiency in digestion and ligation step, also have to be considered. Therefore it is necessary to quantify the efficiency of digestion and ligation for each sample. It could be quantified by PCR or real-time PCR, amplifying fragments containing a restriction site of the enzyme used in UND, DIG, and 3C samples and a region without restriction site as control to normalize the results of the PCR (Fig. 2b). The Ct values (cycle thresholds) were used to calculate the digestion efficiency according to the formula reported in Hagege et al. [17]: $\%restriction = 100 - 100 / 2^{(Ct_{DIG} - Ct_{UND})}$.

7. To assess if the interaction detected in 3C reflects a functional 3D contact, the frequency of random collision has to be determined. This is important especially for *in cis* interactions. The reason is that the flexibility of the chromatin fiber could cause an engagement of DNA segments from the same fiber in random collisions, with a frequency inversely proportional to the genomic distance between them. For this purpose, it is necessary to design primers that scan the entire region that is supposed to interact and to evaluate if two regions interact more frequently with each other than with neighboring sequences. Only in this case it is possible to affirm that you have a specific interaction. Instead *in trans* interactions random collisions are not expected; indeed any interaction that is detected indicates a specific contact.
8. The standard 3C PCR protocol uses a standard number of PCR cycles and a standard amount of DNA template for the analysis of all different ligation products. This approach is only semiquantitative and prone to inaccuracies. In fact PCR products that come to saturation, going outside from the linear range of the amplification reaction, are not suitable for differential analysis. To overcome this limitation, a real-time PCR approach using TaqMan probes, called 3C-qPCR, was developed [18, 19].
9. Interaction frequencies are determined by dividing the intensity of PCR product bands obtained from 3C sample with that from the BAC control, for each pair of primers (Fig. 2c).

References

1. de Laat W, Dekker J (2012) 3C-based technologies to study the shape of the genome. Methods 58(3):189–191. doi:10.1016/j.ymeth.2012.11.005
2. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. doi:10.1126/science.1067799
3. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10(6):1453–1465
4. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W (2003) The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet 35(2):190–194. doi:10.1038/ng1244
5. Dekker J (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods 3(1):17–21. doi:10.1038/nmeth823
6. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R (2006) Interchromosomal interactions and olfactory receptor choice. Cell 126(2):403–413. doi:10.1016/j.cell.2006.06.035
7. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA (2005) Interchromosomal associations between alternatively expressed loci. Nature 435(7042):637–645. doi:10.1038/nature03574
8. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A 103(28):10684–10689. doi:10.1073/pnas.0600326103
9. Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V (2007) Polycomb response elements mediate the formation of chromosome

higher-order structures in the bithorax complex. Nat Cell Biol 9(10):1167–1174. doi:10.1038/ncb1637

10. Xu N, Tsai CL, Lee JT (2006) Transient homologous chromosome pairing marks the onset of X inactivation. Science 311(5764): 1149–1152. doi:10.1126/science.1122984
11. Bacher CP, Guggiari M, Brors B, Augui S, Clerc P, Avner P, Eils R, Heard E (2006) Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat Cell Biol 8(3):293–299. doi:10.1038/ncb1365
12. Solomon MJ, Varshavsky A (1985) Formaldehyde-mediated DNA-protein cross-linking: a probe for in vivo chromatin structures. Proc Natl Acad Sci U S A 82(19): 6470–6474
13. Orlando V, Strutt H, Paro R (1997) Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods 11(2):205–214. doi:10.1006/meth.1996.0407
14. Jackson V (1999) Formaldehyde cross-linking for studying nucleosomal dynamics. Methods 17(2):125–139. doi:10.1006/meth.1998.0724
15. Splinter E, Grosveld F, de Laat W (2004) 3C technology: analyzing the spatial organization of genomic loci in vivo. Methods Enzymol 375:493–507
16. Simonis M, Kooren J, de Laat W (2007) An evaluation of 3C-based methods to capture DNA interactions. Nat Methods 4(11):895–901. doi:10.1038/nmeth1114
17. Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc 2(7):1722–1733. doi:10.1038/nprot.2007.243
18. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev 20(17):2349–2354. doi:10.1101/gad.399506
19. Wurtele H, Chartrand P (2006) Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosome Res 14(5):477–495. doi:10.1007/s10577-006-1075-0

Chapter 20

Determination of High-Resolution 3D Chromatin Organization Using Circular Chromosome Conformation Capture (4C-seq)

Mélody Matelot and Daan Noordermeer*

Abstract

3D chromatin organization is essential for many aspects of transcriptional regulation. Circular Chromosome Conformation Capture followed by Illumina sequencing (4C-seq) is among the most powerful techniques to determine 3D chromatin organization. 4C-seq, like other modifications of the original 3C technique, uses the principle of "proximity ligation" to identify and quantify ten thousands of genomic interactions at a kilobase scale in a single experiment for predefined loci in the genome.

In this chapter we focus on the experimental steps in the 4C-seq protocol, providing detailed descriptions on the preparation of cells, the construction of the circularized 3C library and the generation of the Illumina high throughput sequencing library. This protocol is particularly suited for the use of mammalian tissue samples, but can be used with minimal changes on circulating cells and cell lines from other sources as well. In the final section of this chapter, we provide a brief overview of data analysis approaches, accompanied by links to publicly available analysis tools.

Key words Circular Chromosome Conformation Capture, 4C-seq, 3D chromatin organization, DNA interactions, Chromatin loops, Chromatin compartmentalization, Nuclear organization, High-throughput sequencing

1 Introduction

3D chromatin organization is an essential component of transcriptional regulation [1, 2]. The function of enhancers and insulators requires the formation of 3D chromatin loops ([3, 4] and Fig. 1, left), whereas the recently discovered Topological Associated Domains (TADs) appear to spatially structure and separate gene regulatory domains [5–7]. Genomic regions that bind the repressive Polycomb group proteins (PcG-proteins) and that carry the associated H3K27me3 histone mark form specialized 3D chromatin structures as well. The first 3C studies (Chromosome Conformation Capture) in human and *Drosophila* cells identified complex loop structures at the GATA-4 gene (in cultured human Tera-2

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_20, © Springer Science+Business Media New York 2016

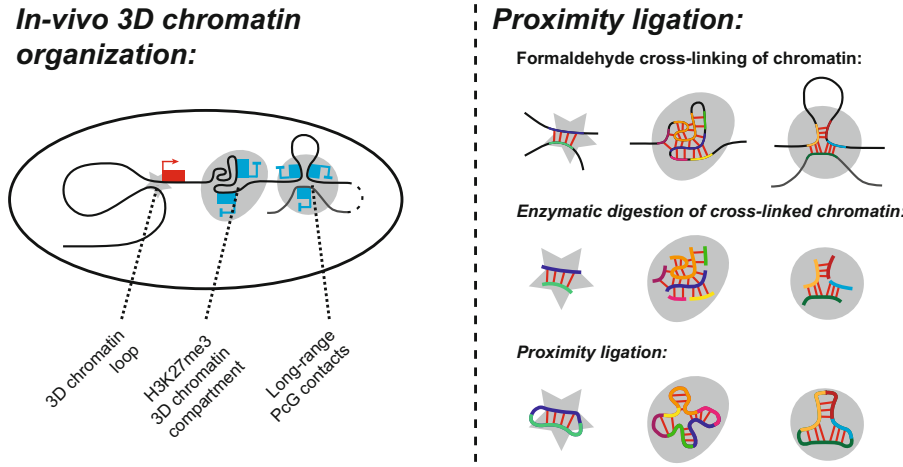


Fig. 1 The principle of proximity ligation for the detection of 3D chromatin interactions. Chromatin fragments that are in spatial proximity are cross-linked in vivo using formaldehyde (*red bars*). After enzymatic digestion, only cross-linked fragments are kept together. Proximity ligation circularizes those fragments that are together due to their shared cross-links

embryonic carcinoma cells, [8]) and the bithorax complex (BX-C in *Drosophila* embryonic cells, [9]). More recent 4C-seq studies in mouse embryos reported that the repressed *Hox* gene clusters form local 3D chromatin compartments. In these dynamic 3D compartments the H3K27me3 marked chromatin clusters together, following the temporal and spatial repressed state of the *Hox* genes ([10, 11] and Fig. 1, left). Moreover, in mouse and *Drosophila* cells, PcG targets form long-range contacts among each other despite being separated by many megabases on the same chromosome or being located on different chromosomes ([10, 12–14] and Fig. 1, left). Multiple specialized and 3D chromatin structures are therefore dynamically associated with PcG-mediated repression in mammalian and insect cell systems.

Among the most used techniques to study 3D genome organization at high resolution is the 4C-seq technique (Circular Chromosome Conformation Capture followed by Illumina sequencing). 4C-seq is a genome-wide adaptation to the original 3C technique [15] that was originally developed for readout with microarrays (4C, [16]). In a 4C-seq experiment, the genome-wide 3D interactions of preselected genomic sites (the so-called "viewpoints") are identified and quantified. Due to this focus on individual viewpoints, typically several tens of thousands of interacting sites can be identified in a single experiment, which makes 4C the most comprehensive approach for individual genomic sites available (as compared to other genome-wide adaptations to the 3C approach like 5C, HiC, and ChIA-PET). The combination of 4C with Illumina sequencing allows sequencing of up to 20 viewpoints at a time, thereby considerably improving the throughput of the experiments.

4C-seq, like all 3C-based assays, relies on "proximity ligation" to detect contacts between DNA fragments (Fig. 1, right). First, 3D chromatin organization is fixed in vivo over short distance using formaldehyde cross-linking. Next, chromatin is fragmented using a restriction enzyme, only keeping fragments together that were cross-linked due to their initial spatial proximity. In a final step, the DNA is ligated under diluted conditions, thereby promoting ligation between fragments that are present in the same cross-linked complexes. Using the frequency of ligation between pairs of restriction fragments as readout, 3D chromatin interactions can next be determined. By interrogating the ligation events of a viewpoint, typically in around 100,000 cells, an average snapshot of 3D chromatin organization within the cell population can be obtained. Depending on the downstream bioinformatics analysis, local 3D organization, long-range interactions or differences in 3D organization between cell types or experimental conditions can be determined.

In this chapter, we provide a detailed description of the 4C-seq approach, with particular emphasis on the experimental procedures (*see* Fig. 2). This protocol generates a very high-resolution description of 3D chromatin organization for individual viewpoints, by using two sequential rounds of frequent cutting restriction enzymes (4 bp recognition sites, resulting in an average distance between informative restriction fragments of 1–1.5 kb). The first section of this protocol describes the preparation of cross-linked nuclei from tissue samples, which can be used with minor modifications for cultured or circulating cells as well (*see* **Notes 1** and **6**). In the second section, a circularized 3C library is generated using a first round of digestion and (proximity) ligation on cross-linked chromatin followed by a second round of digestion and ligation on naked DNA. In the third section, a 4C-seq library for Illumina sequencing is generated by PCR amplification of the circularized 3C library with viewpoint-specific inverse primers. In the final section, an overview of the data analysis strategy is provided, accompanied by links to recently published and publicly available resources that can be used for the bioinformatics analysis of 4C-seq data.

2 Materials

2.1 Isolation and Preparation of Tissue Samples

1. Isolation buffer: Phosphate buffered saline (PBS) (pH 7.2–7.6) supplemented with 10% filter-sterilized fetal bovine serum (FBS). Can be stored at 4 °C for 1 week.
2. 12.5% w/v collagenase solution: collagenase powder from *Clostridium histolyticum* dissolved in PBS.
3. Cell strainer: 35 μm cell strainer for round bottom tube or 40 μm cell strainer for 50 ml tube.

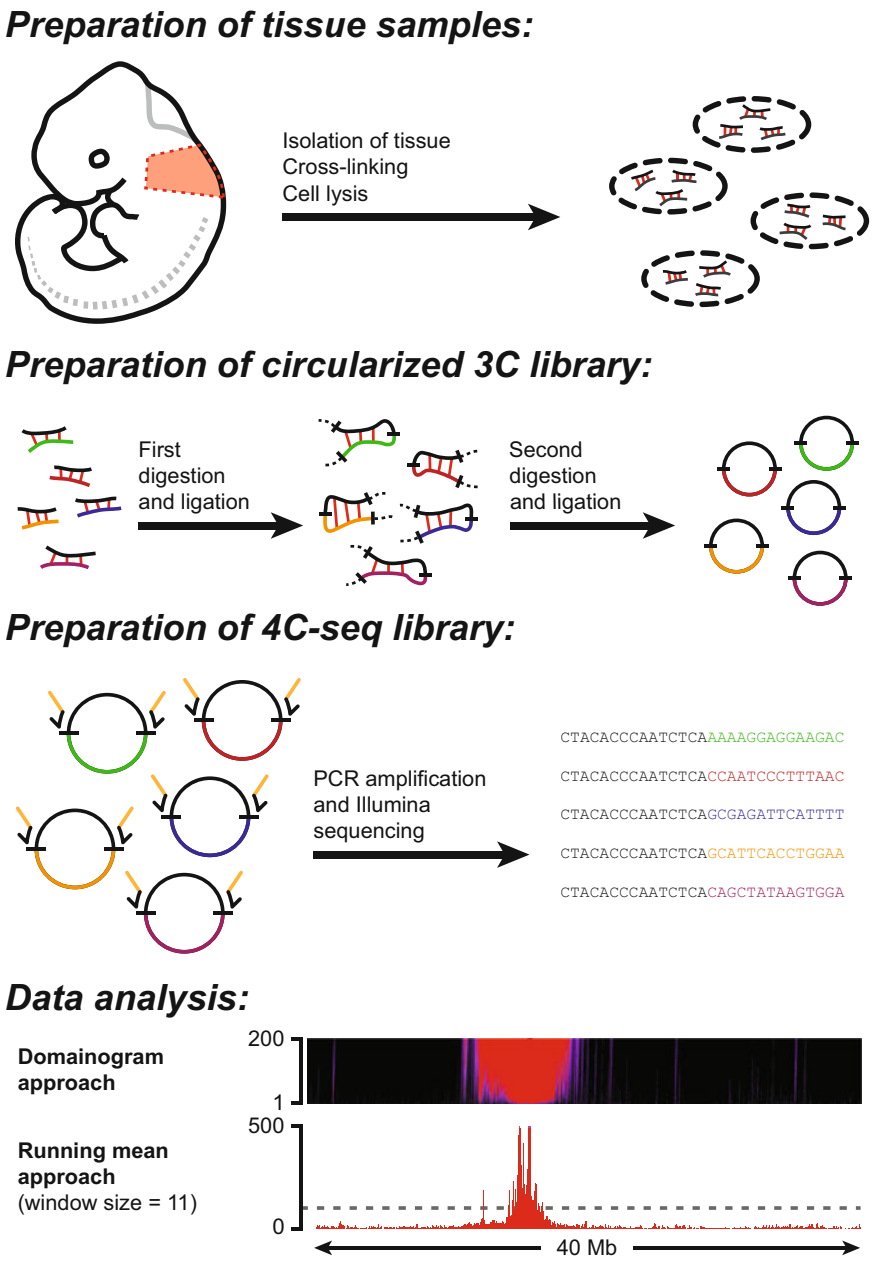


Fig. 2 Schematic outline of the steps in the 4C-seq assay

4. 37% w/v formaldehyde solution.
5. 2% cross-linking solution: 650 µl 37% formaldehyde solution, 11.35 ml isolation buffer. Prepare the solution fresh on the day of use. Keep at 20–25 °C (room temperature).
6. 1 M glycine solution. Keep at 4 °C.
7. Cell lysis buffer: 50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% v/v NP-40, 1% v/v Triton X-100, 1×

complete protease inhibitor cocktail, EDTA free. Prepare the solution on the day of use and keep at 4 °C.

Depending on the cell type, the cell lysis buffer may need to be optimized. *See* **Note 6** for other published lysis buffers.

2.2 Preparation of Circularized 3C Library

1. 1.2× restriction buffer: 60 µl 10× restriction buffer (as supplied with restriction enzyme), 12 µl bovine serum albumin (BSA stock) (10 mg/ml stock), 428 µl ultrapure water.
2. 20% sodium lauryl sulfate (SDS) solution.
3. 20% Triton X-100 solution.
4. Selected restriction enzymes, preferably in high concentrated form.
5. 5 M NaCl solution.
6. 1 M Tris–HCl, pH 7.5 solution.
7. 0.5 M EDTA solution.
8. 20 mg/ml Proteinase K solution.
9. 10 mg/ml RNaseA solution.
10. Phenol–chloroform–IAA solution (25:24:1): 50% v/v phenol, saturated with 10 mM Tris, pH 8.0 and 1 mM EDTA, 48% v/v chloroform, 2% v/v isoamyl alcohol (commercially available).
11. 2 M sodium acetate (NaAc), pH 5.6 solution.
12. 70% and 100% ethanol solutions.
13. 10× ligation buffer: 660 mM Tris–HCl, pH 7.5, 50 mM $MgCl_2$, 50 mM dithiothreitol (DTT), and 10 mM adenosine triphosphate (ATP). 1 ml aliquots can be stored at −20 °C for 6 months.
14. T4 DNA ligase, preferably in high concentrated form.
15. PCR Purification Kit (e.g., Qiagen, prior to the first use, add four volumes of 100% ethanol to wash buffer PE).

2.3 Optimization of PCR Conditions and Preparation of 4C Sequencing Libraries

1. Expand Long Template PCR System (e.g., Roche life sciences).
2. 10 mM dNTP solution. Can be stored for 1 month at −20 °C. Repeated freeze–thaw cycles and prolonged storage at −20 °C can severely decrease PCR efficiency.
3. Locus specific inverse Forward and inverse Reverse primers, diluted at 100 µM (*see* **Note 4**).
4. PCR Purification Kit (e.g., Qiagen, additional PB binding buffer can be ordered separately).

2.4 Available Pipelines for 4C Data Analysis

1. HTSstation: online data analysis and Python scripts; [17] and http://htsstation.epfl.ch/
2. 4Cseqpipe: R-package; [18] and http://compgenomics.weizmann.ac.il/tanay/?page_id=367/

3. FourCSeq: R/Bioconductor-package; [19] and http://www.bioconductor.org/packages/release/bioc/html/FourCSeq.html
4. Basic4Cseq: R/Bioconductor-package; http://www.bioconductor.org/packages/release/bioc/html/Basic4Cseq.html
5. r3Cseq: R/Bioconductor-package; [20] and http://www.bioconductor.org/packages/release/bioc/html/r3Cseq.html

3 Methods

See **Notes 1–4** and Fig. 3 for important guidelines on tissue quantities, cell types, choice of restriction enzymes and primer design that should be considered prior to starting the experiment.

3.1 Isolation and Preparation of Tissue Samples

1. Pool dissected tissue fragments in 495 µl of isolation buffer. Keep at 4 °C during dissection. *See* **Note 5** if samples of less than 1×10^7 cells are used.
2. Add 5 µl of 12.5% w/v collagenase solution (final concentration 0.125%) and incubate tissue fragments for 45 min at 37 °C in a shaker at 750 rpm. Dissociate cells by pipetting up and down several times with a blue tip.
3. Make single cell by forcing the solution through a cell strainer and transfer cells to a 15 ml conical tube. Add 9.5 ml of 2% cross-linking solution and incubate the cells for 10 min on a rotating wheel or rocking platform at 20–25 °C (room temperature).
4. Immediately transfer the tube with cells to ice. Quench the cross-link reaction by adding 1.43 ml of a cold 1 M Glycine solution. Centrifuge the cells for 8 min at 225 × *g*, 4 °C.
5. Remove the cross-linking solution. Resuspend the cells in 5 ml cold cell lysis buffer and incubate the cells for 10 min on ice followed by pipetting up and down several times with a blue tip. Centrifuge the nuclei for 5 min at 400 × *g*, 4 °C. *See* **Note 6**.
6. Remove 4.5 ml of the cell lysis buffer and resuspend the nuclei in the remaining 500 µl of volume. Transfer the remaining volume to a 1.5 ml plastic micro tube and centrifuge the nuclei for 1 min at 230 × *g*, 4 °C.
7. Remove the remaining 500 µl of the cell lysis buffer. At this stage, cells can either be frozen in liquid nitrogen and stored at −80 °C until further use, or the protocol can be immediately continued at Subheading 3.2.

3.2 Preparation of Circularized 3C Library

Multiple preparations of cross-linked nuclei, as prepared in Subheading 3.1, can be pooled at this point. In this case, we advise to pool and wash the samples: add a total volume of 500 µl 1.2×

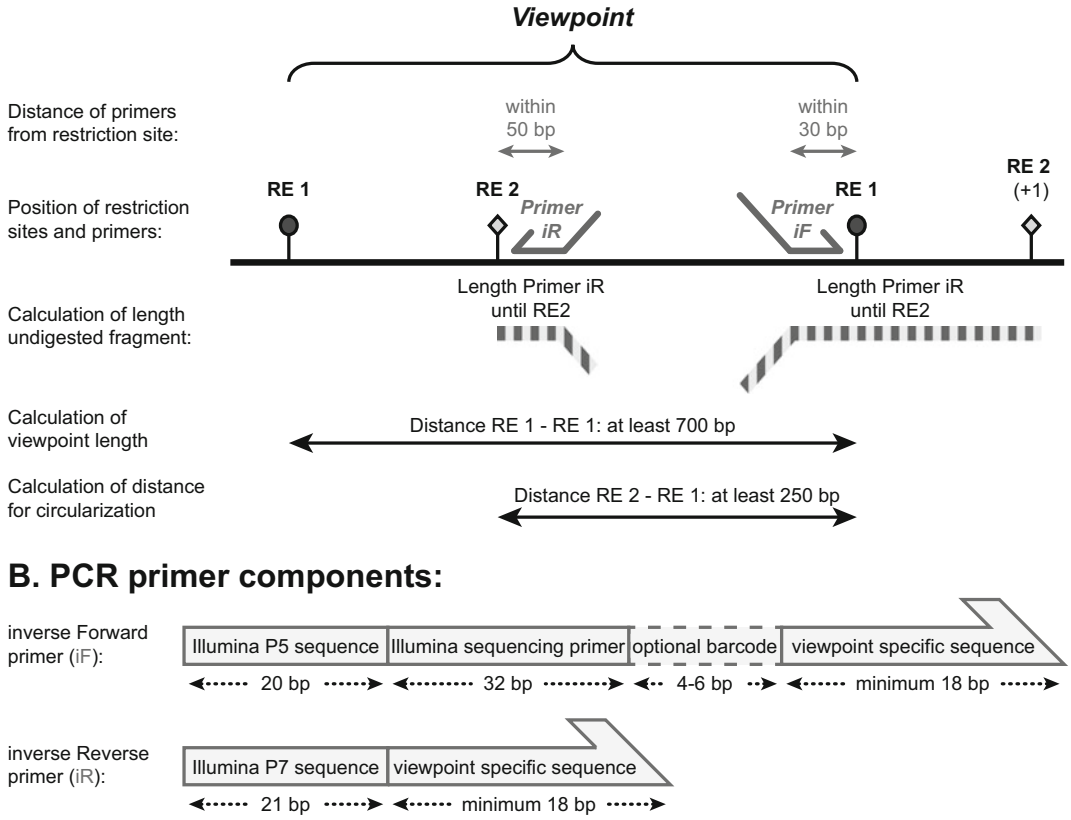


Fig. 3 Design considerations for 4C-seq experiments and PCR primers. (**a**) Scheme with design considerations for 4C-seq experiments. A useful viewpoint should be of sufficient length (distance between the cut sites of the primary restriction enzyme RE 1) and should have sufficient distance between the cut sites of the primary (RE 1) and secondary (RE 2) restriction enzyme to allow circularization of the 4C-library. The inverse primers (iF and iR) should be located within the indicated distance next to the cut sites of the restriction enzymes. The length of the undigested fragment (*see* **Note 4**) can be calculated by adding up the length of both the iF and iR primers, including adapter sequences, up to their respective cut sites + the distance between RE 1 and the first downstream cut site of the secondary restriction enzyme RE 2 (+1) (total length of the striped bars). (**b**) Components of the inverse Forward (iF) and inverse Reverse (iR) primers. We standardly design the iF primer next to RE 1 and the iR primer next to RE 2

restriction buffer to the combined samples, pool the samples together in a single 1.5 ml plastic micro tube, centrifuge the nuclei for 1 min at 230 × *g* at 4 °C and remove the supernatant. Next, continue with **step 1** of this subheading.

1. Take up nuclei in 500 µl 1.2× restriction buffer, place at 55 °C in a shaker at 750 rpm and immediately add 7.5 µl of a 20% SDS solution (final concentration 0.3%). Incubate for no more than 10 min at 60 °C, followed by another 50 min at 37 °C in a shaker at 750 rpm. *See* **Note 7** for the appearance of the solution.

2. Sequester the SDS in the solution by adding 50 µl of a 20% Triton X-100 solution (final concentration 2%) and incubate for 1 h at 37 °C in a shaker at 750 rpm. Optionally, a 5 µl aliquot can be taken at this point as "undigested control". See in **step 4** of this subheading how to revert cross-links and visualize this control.
3. Digest the cross-linked DNA by adding 400 Units of the selected restriction enzyme and incubate for 4–6 h at 37 °C in a shaker at 750 rpm. Add another 400 Units of the restriction enzyme and incubate overnight at 37 °C in a shaker at 750 rpm.
4. Prior to continuation, verify the efficiency of the DNA digestion. Take a 5 µl aliquot as "digested control". The optional "undigested control" from **step 2** should be added here as well. The remainder of the digested samples can be stored at 4 °C before continuing at **step 5**. Add 90 µl ultrapure water and 5 µl of a 5 M NaCl solution to the control(s) and incubate for 2 h at 65 °C in a shaker at 750 rpm. Lower the temperature to 45 °C, add 2 µl of a 1 M Tris–HCl, pH 7.5 solution, 2 µl of a 0.5 M EDTA solution and 2 µl of a Proteinase K solution and incubate for 2 h at 45 °C in a shaker at 750 rpm. Lower the temperature to 37 °C, add 2 µl of an RNaseA solution and incubate for 30 min at 37 °C in a shaker at 750 rpm. Transfer sample to a safety cabinet, add 120 µl of a phenol–chloroform–IAA solution and shake vigorously. Centrifuge the sample for 15 min at maximum speed (typically around 20,000 × *g*) at 20–25 °C. Transfer aqueous phase to a new plastic micro tube, add 12 µl of a 2 M NaAc solution, add 300 µl of 100% ethanol and centrifuge for 30 min at maximum speed at 4 °C. Remove the supernatant, add 200 µl of a 70% ethanol solution and centrifuge for 10 min at maximum speed at 4 °C. Briefly air-dry sample and dissolve the sample in 20 µl ultrapure water. Visualize the sample on a 1.5% agarose gel. The sample should run as a large smear with highest intensity between 500 and 1000 bp (*see* Fig. 4a). If the chromatin shows good digestion, the remaining sample can be further processed (continue to **step 5**). If chromatin is partially digested, showing a smear with highest intensity in the range of 1–5 kb, **step 3** of this subheading should be repeated with the remaining sample. If no or minimal digestion has occurred, tissue preparation, cell lysis or the choice of restriction enzyme should be optimized (*see* **Notes 1**, **3** and **6**).
5. Take the remainder of the digested nuclei (between 500 and 600 µl) and inactivate the remaining restriction enzyme by adding 40 µl of a 20% SDS solution (final concentration around 1.5%) and an incubation of no more than 20 min at 65 °C.
6. Transfer the sample to a 50 ml conical tube and transfer to a 37 °C water bath. Add 6.13 ml 1.15× ligation buffer (premixed: 710 µl 10× ligation buffer + 5.42 ml ultrapure water),

add 375 µl of a 20% Triton X-100 solution to sequester the SDS and incubate for 1 h at 37 °C, with occasional shaking.

7. Transfer the sample to a 16 °C water bath and incubate for 15 min. Ligate the diluted cross-linked DNA fragments by adding 100 Units of T4 DNA ligase and incubation at 16 °C for 4 h, followed by 30 min at 20–25 °C.
8. De-cross-link the sample by adding 15 µl of a Proteinase K solution and overnight incubation in a water bath at 65 °C.
9. Transfer the sample to a 37 °C water bath, add 30 µl of RNaseA solution and incubate for 45 min at 37 °C. Transfer the sample to a safety cabinet, add 7 ml of a phenol–chloroform–IAA solution, shake vigorously and centrifuge for 15 min at 3200 × *g* at 20–25 °C. Transfer aqueous phase to a new 50 ml conical tube and add 7 ml of ultrapure water, 1.5 ml of a 2 M NaAc solution, and 35 ml of 100% ethanol. Mix gently and place at −80 °C for at least 2 h. Centrifuge for 45 min at 3200 × *g* at 4 °C. A relatively large pellet will be visible that consists mainly of salts from the ligation buffer (*see* **Note 8**). Remove the supernatant, add 25 ml of a 70% ethanol solution and centrifuge for 15 min at 3200 × *g* at 4 °C. Remove the supernatant, dry for 1 h at 55 °C, transfer the sample to a 37 °C water bath and dissolve the pellet by adding 150 µl of a 10 mM Tris–HCl, pH 7.5 solution and 1 h incubation combined with occasional tapping of the tube. After a 1-h incubation, the pellet should have fully dissolved.
10. Transfer the solution to a 1.5 ml plastic micro tube. Measure 2 µl of the solution using a dye-incorporation based method (e.g., Qubit, *see* **Note 8**). When starting with 1×10^7 cells, the remaining amount of DNA should typically be around 40 µg. Run 500 ng of the sample on a 1.5% gel to confirm the efficiency of ligations (*see* Fig. 4b). The sample may be stored at −20 °C for several weeks at this stage.
11. Digest the sample with the second restriction enzyme (*see* **Note 3**) by diluting the DNA to a concentration of 100 ng/µl in the appropriate 1× restriction buffer (supplemented with BSA, if required) and add 1 Unit of restriction enzyme/µg of DNA. Digest the DNA overnight at 37 °C in a shaker at 750 rpm.
12. Inactivate the restriction enzyme by 20 min incubation at 65 °C. Add an equal volume of a phenol–chloroform–IAA solution, shake vigorously and centrifuge the sample for 15 min at maximum speed (typically around 20,000 × *g*) at 20–25 °C. Transfer aqueous phase to a new plastic micro tube, add 1/10 volume of a 2 M NaAc solution, add 2.5 volumes of 100% ethanol and place tube for 30 min at −80 °C. Centrifuge for 30 min at maximum speed at 4 °C, remove the supernatant,

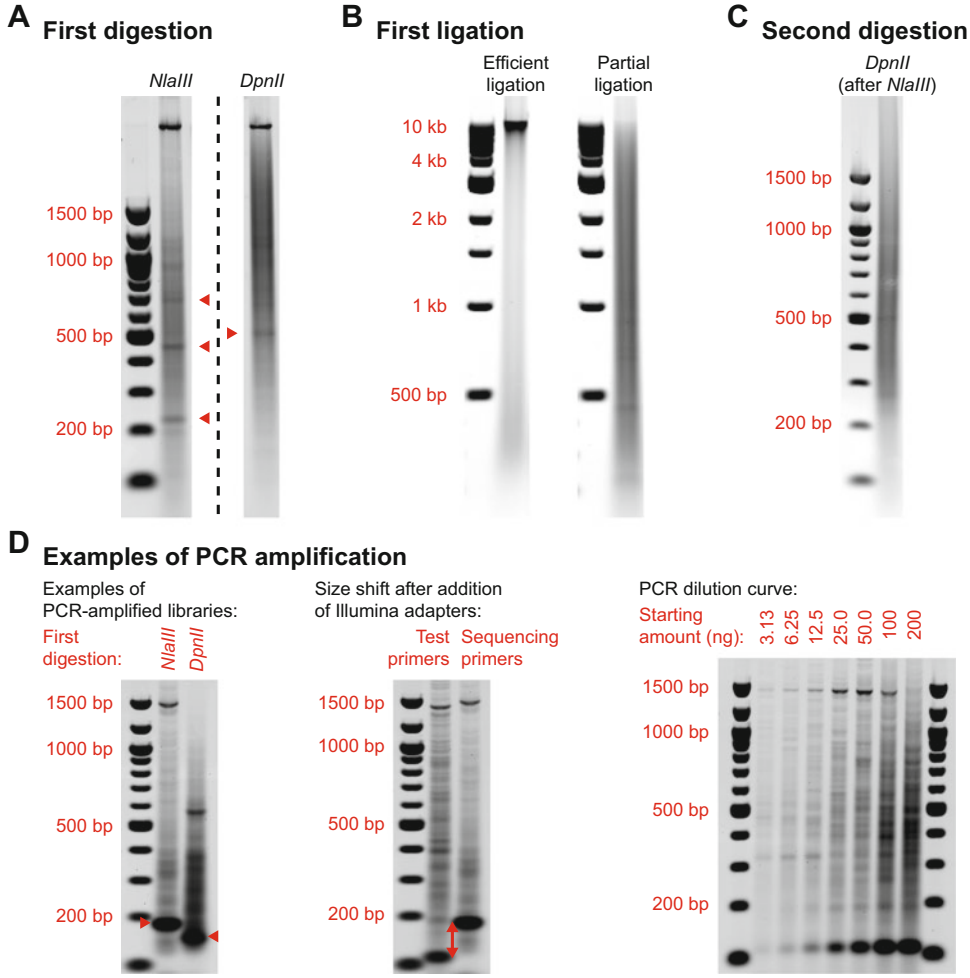


Fig. 4 Visualization of cross-linked chromatin and DNA at different stages of the 4C-seq procedure. (**a**) Digestion of cross-linked chromatin (2 % formaldehyde, 10 min) with *NlaIII* (*left*) and *DpnII* (*right*). Digestion of cross-linked material is not 100 % efficient, which results in a smear with an average length that is larger than randomly predicted. In sufficiently digested samples, within the large smear several specific products that stem from repeated sequences can be observed (*red arrow* heads for mouse genomic DNA). (**b**) Ligation of cross-linked chromatin (initially digested with *NlaIII*). *Left*: highly efficient ligation; *right*: partial ligation. Notice that specific bands that were visible after digestion have disappeared. In our hands, both efficiently and partially ligated samples can give good downstream results. (**c**) Second digestion of DNA, after de-crosslinking. A smear should be visible with a large distance distribution and no specific products. (**d**) PCR amplification of 4C-seq library with locus specific primers. After PCR amplification, a smear should be visible with fragment lengths that extend beyond 1 kb. *Left*: examples of PCR-amplified 4C-seq libraries first digested by *NlaIII* and *DpnII* (iF and iR sequencing primers). Undigested bands, which are specific to each primer set, are indicated by arrowheads. *Middle*: shift in the size of the undigested fragment upon addition of Illumina adapters (iF and IR test primers versus iF and iR sequencing primers, *see* **Note 4**). *Right*: titration to find the optimal concentrations for PCR amplification (iF and IR test primers). Too low concentrations result in poor yield, whereas overloading of the PCR results in an unwanted shift towards smaller fragments. Linearity here is observed in lanes 1–5 (3.13–50 ng per PCR reaction), after which the reaction becomes overloaded. The optimal concentration in this example is therefore 50 ng per PCR reaction. All samples were visualized on 1.5 % agarose gels with a 100 bp or 1 kb ladder as reference (sizes of bands indicated in *red*). Primer sequences in **Note 4**

add 1 ml of a 70% ethanol solution and centrifuge for 10 min at maximum speed at 4 °C. Briefly air-dry sample and dissolve the sample in 100 µl ultrapure water. Verify the efficiency of the digestion by visualizing 2.5 µl of the sample on a 1.5% agarose gel (*see* Fig. 4c).

13. Subsequently, the final circularized 3C-library is generated by re-ligation under diluted conditions. Transfer the sample to a 50 ml conical tube, add 12.5 ml ultrapure water, add 1.4 ml 10× ligation buffer and transfer the tube to a 16 °C water bath. Add 200 Units of T4 DNA ligase and incubate for 4 h at 16 °C followed by 30 min at 20–25 °C.
14. Transfer the sample to a safety cabinet, add 14 ml of a phenol–chloroform–IAA solution, shake vigorously and centrifuge for 15 min at 3200 × *g* at 20–25 °C. Transfer the aqueous phase to a new 50 ml conical tube, add 14 ml of ultrapure water, 2.8 ml of a 2 M NaAc solution, mix gently, and divide the solution equally over two 50 ml conical tubes. Add 35 ml of 100% ethanol to each tube, mix gently and place at –80 °C for at least 2 h (or overnight). Centrifuge for 45 min at 3200 × *g* at 4 °C. In both tubes, a relatively large pellet will be visible (*see* **Note 8**). Remove the supernatant, add 25 ml of a 70% ethanol solution to both tubes and centrifuge for 15 min at 3200 × *g* at 4 °C. Remove as much as possible of the supernatant, dry for 1 h at 55 °C, transfer the tubes to a 37 °C water bath and dissolve the pellets by adding 200 µl of a 10 mM Tris–HCl, pH 7.5 solution to each tube and 1 h incubation combined with occasional tapping of the tube. After a 1-h incubation, the pellet should have fully dissolved.
15. Clean the library by addition of 1 ml of Qiagen PB loading buffer to each tube and mix by pipetting up and down a few times. Load 4 QIAquick Spin Columns with 600 µl each of the solution and centrifuge for 1 min at maximum speed (typically around 20,000 × *g*) at 20–25 °C. Remove the flow through, add 600 µl wash buffer PE and centrifuge for 1 min at maximum speed at 20–25 °C. Remove the last traces of wash buffer by transferring the spin columns to new plastic 1.5 ml micro tubes and 1 min centrifugation at maximum speed at 20–25 °C. Elute the samples by transferring the spin columns to new plastic 1.5 ml micro tubes, addition of 40 µl of a 10 mM Tris–HCl solution to each Spin column, 1 min incubation at 20–25 °C and 1 min centrifugation at maximum speed at 20–25 °C. Pool all samples in a single 1.5 ml plastic micro tube and measure 2 µl of the solution using a dye-incorporation based method (e.g., Qubit, *see* **Note 8**). When starting with 1×10^7 cells, the DNA concentration should typically be around 100 ng/µl. At this stage, the sample can be stored long-term at -20 °C.

3.3 Optimization of PCR Conditions and Preparation of 4C-seq Libraries

1. First, the optimal amount of the circularized 3C library for PCR amplification needs to be determined. This amount depends on the concentration of the circular DNA and the amount of salt contamination in the library. To determine the optimal concentration, a dilution curve with the following amounts of DNA/50 μl PCR reaction should be run: 12.5, 25, 50 and 100 ng. The total volume of the circularized 3C library that is added to each PCR reaction should not exceed 2 μl. If possible, use a confirmed primer set (*see* **Note 4**). Primers with added Illumina adapter sequences are not required at thmis step.
   Components of PCR reaction:
   - 5 μl 10× Buffer 1 (Roche Expand Long Template PCR System)
   - 1 μl 10 mM dNTP solution
   - 0.5 μl 100 μM iF primer (50 pmol final concentration)
   - 0.5 μl 100 μM iR primer (50 pmol final concentration)
   - Variable amount of circularized 3C library
   - Ultrapure water up to 49.25 μl
   - 0.75 μl DNA Polymerase mix (e.g., 5 U/μl, Roche Expand Long Template PCR System)

   PCR program:
   - 1× (2 min at 94 °C)
   - 30× (15 s at 94 °C, 1 min at 55 °C; 3 min at 68 °C)
   - 1× (7 min at 68 °C)

   After PCR, visualize 20 μl of the material on a 1.5 % agarose gel. The optimal concentration is the sample where the intensity of the smear is still linearly increased, where the size range in PCR products is maximal and where minimal primer dimers are detected (*see* Fig. 4d, right).
2. Prepare as many PCR reactions as necessary to amplify a total amount of 1 μg circularized 3C library, containing iF and iR primers with Illumina sequences (*see* **Note 4** for design strategy). Use the same PCR program as mentioned in **step 1**.
3. After PCR, pool products from the same viewpoint, add five volumes of Qiagen PB loading buffer and load two QIAquick Spin Columns with the PCR product (columns may need to be loaded multiple times). After each loading step, centrifuge for 1 min at maximum speed (typically around 20,000 × *g*) at 20–25 °C and remove the flow through. After loading, add 600 μl wash buffer PE and centrifuge for 1 min at maximum speed at 20–25 °C. Remove the last traces of wash buffer by transferring the spin columns to new plastic 1.5 ml micro tubes and 1 min centrifugation at maximum speed at 20–25 °C. Elute

the samples by transferring the spin columns to new plastic 1.5 ml micro tubes, add 50 µl of ultrapure water to each Spin column, incubate for 1 min at 20–25 °C and 1 min centrifugation at maximum speed at 20–25 °C. Load the Spin columns a second time with 50 µl ultrapure water and centrifuge for 1 min at maximum speed at 20–25 °C.

4. Pool the eluate of both columns and measure 2 µl of the solution using a dye-incorporation based method (e.g., Qubit).
5. Verify that the material has amplified well and that most unincorporated primers have been removed by visualizing the material on a 1.5% agarose gel (*see* Fig. 4d).
6. Pool PCR products of up to 20 viewpoints together (or barcoded PCR products, if the same viewpoint from different samples should be pooled, *see* **Note 4**) at a final concentration of 1.625 µg/µl. To avoid imbalances within the first six bases that are used for base calibration, particularly when few different viewpoints or barcodes are used, we advise mixing in 25% PhiX balancer DNA. Samples can be sequenced on the Illumina HiSeq system using at least 75 bp read length without further processing. Due to the wide length distribution of the PCR amplified material, we load the flow cell at moderately reduced cluster density (around 50–75%).

3.4 Data Analysis

After Illumina sequencing, bioinformatics tools should be used to generate a density profile of (averaged) DNA contacts for each viewpoint. In recent years, several 4C-seq data analysis pipelines have been developed. Publicly available examples are indicated in Subheading 2.4.

All abovementioned pipelines essentially follow the same data flow, as briefly outlined below (discussed in more detail in ref. 21). Most pipelines were developed with a specific biological question in mind, therefore using somewhat different strategies to identify significant interactions and/or changes in 3D genome organization. The choice of pipeline should be made with the biological question and availability of bioinformatics infrastructure in mind.

Summary of common data analysis steps:

1. Demultiplexing of data using the sequence specific iF sequence (and optional barcode) to identify the viewpoint-origin. Followed by removal of iF primer and barcode sequences.
2. Mapping of the remainder of reads to the reference genome followed by exclusion of reads that do not map next to restriction sites. This mapping serves as the quantitation of proximity ligation events between the viewpoint and other restriction fragments, thereby providing the snapshot of average genome-wide 3D chromatin interactions. Depending on the pipeline, the reads can further be translated into restriction fragments

and can be fitted to polymer-based models that compensate for reduced contact frequencies at increasing distance from the viewpoint.

3. In the final step, significant interactions within data sets or significant changes between data sets are determined. Depending on the pipeline, different approaches may be taken:
   - Running mean approaches with fixed window-size and thresholding to identify distant interacting regions.
   - Domainogram approaches [22], using multi-scale clustering in windows of variable size to determine local interaction trends or to identify significant distant interacting regions and determine their approximate size.
   - Statistical approaches, using variance-stabilizing transformation, to reliably determine differences in 3D organization between different samples at different length scales.

4 Notes

1. This protocol is optimized for mammalian tissue samples. With minor modifications it may be used for circulating cells, cell lines and tissues samples from other organisms as well. In these cases, the cell lysis may need to be optimized (*see* **Note 6**). The collagenase treatment can be omitted for circulating cells and for cell lines that do not produce collagen. If required for cell lines, we advise to do trypsin treatment prior to the start of the described protocol.
2. This protocol is optimized for 10 million mammalian cells (1×10^7 cells), which corresponds to about 60 μg of chromatin. If different starting amounts are used, all volumes should be scaled accordingly. Doing the experiment on fewer than 1 million cells (1×10^6 cells) is not advised as loss of material during the procedure will generally result in insufficient material for the final PCR amplification.
3. High-resolution 4C-seq uses two sequential digestions with restriction enzymes that have a 4 bp recognition site. Depending on the species and the choice of restriction enzymes, the distance between informative restriction fragments is in the order of 1–1.5 kb. The choice of restriction enzymes should depend on the following parameters:
   - The primary and secondary restriction enzyme should be insensitive to the methylation state of the DNA (e.g., CpG methylation in mammalian cells).
   - The primary restriction enzyme should be capable of digesting cross-linked DNA with high efficiency at 37 °C. For mammalian cells, results with only four restriction enzymes

have been published so far: *DpnII* and its isoschizomer *MboI* (recognizing GATC), *NlaIII* (recognizing CATG), and *Csp6I* (recognizing GTAC) [10, 18, 23].

- At the intended viewpoint, digestion with the primary restriction enzyme should generate a restriction fragment that is at least 700 bp long. Though we and other groups have used smaller fragments (up to 500 bp, [23]), the success rate of smaller fragments in our hands is less reliable. *See also* Fig. 3a.
- Several secondary restriction enzymes have been successfully used: *DpnII* (GATC), *NlaIII* (CATG), *Csp6I* (GTAC), *TaqαI* (TCGA, 65 °C), and *MseI* (TTAA) [10, 18, 24–26]. At the intended viewpoint, the secondary restriction enzyme should cut at least 250 bp away from the primary restriction enzyme to allow for efficient circularization [27]. *See also* Fig. 3a.

4. Good quality primers at the intended viewpoint are essential for obtaining high-quality 4C-seq data. The design strategy we present here is optimized for the Illumina HiSeq system with single-end (SE) sequencing and 100 bp read length. The use of paired-end (PE) reads or other Illumina sequencing platforms requires a modified Illumina P7 sequence.
   Guidelines for primer design (*see also* Fig. 3b):
   - Inverse Forward (iF) and inverse Reverse (iR) primers are designed with Primer3 (version 4.0.0; http://bioinfo.ut.ee/primer3/) using standard settings, except for Primer size (min 18, opt 20, max 27) and Primer Tm (min 54.0, opt 55.0, max 57.0). iF and iR primers are designed independently. To avoid nonspecific amplification, primers should not map to repeated sequences.
   - Illumina sequencing will be directed from the iF primer, whose sequence up to the restriction site will be included into the Illumina read. We design the iF primer within the first 30 bp from the primary restriction site to keep sufficient read length for the identification of interacting fragments (RE 1 in Fig. 3a).
   - The optimal length of PCR products for Illumina sequencing, without added Illumina adapters is around 150 bp (supplemental Fig. 5 in [10]). To promote the efficiency of sequencing, we keep the average fragment length as short as possible by designing the iR primer within the first 50 bp from the secondary restriction site (RE 2 in Fig. 3a).
   - After primer design, we test the quality of the iF and iR primers, without Illumina adapters, by PCR amplification on a circularized 3C template followed by gel electrophoresis (*see* Fig. 4d, left). The resulting PCR products should

run as a large smear with maximum size between 1 and 1.5 kb. In many cases, an undigested fragment can be observed with a predictable length (*see also* Fig. 3a). Few other discrete bands may be visible as well. Primers are rejected if (1) a very weak smear is detected, (2) a smear with a maximum below 1 kb is detected, (3) a very faint smear with a very strong undigested band is detected, (4) a faint smear with multiple strong (over 5) discrete bands is detected or (5) mainly primer-dimers or other fragments below 100 bp are detected.

– After verification of the primers, Illumina adapter sequences are added (*see* Fig. 3b):

  iF: AATGATACGGCGACCACCGA (Illumina P5 sequence)—ACACTCTTTCCCTACACGACGCTCTTCCGATCT (Illumina sequencing primer)—optional barcode—viewpoint specific primer sequence.

  iR: CAAGCAGAAGACGGCATACGA (short Illumina P7 sequence)—viewpoint specific primer sequence.

  Primers should be PAGE purified. Similar to the short primers, the quality should first be tested by PCR amplification on a circularized 3C template followed by gel electrophoresis (*see* Fig. 4d, middle). An optional 4–6 bp barcode can be added to the iF primer if the same viewpoint from different samples is sequenced in the same Illumina lane. We advise to use base-balanced and varied barcodes that can tolerate one or two sequencing errors (e.g., similar to the Illumina TruSeq LT Kit barcodes, as provided in the TruSeq Sample Preparation Pooling Guide).

– Primers used for Fig. 4d are as follows:

  Mouse, primary enzyme: *NlaIII*, secondary enzyme: *DpnII,* viewpoint: *Hoxd13* gene [10], iF test primer: AAAATCCTAGACCTGGTCATG, iF sequencing primer: AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAATCCTAGACCTGGTCATG, iR test primer: GGCCGATGGTGCTGTATAGG, iR sequencing primer: CAAGCAGAAGACGGCATACGAGGCCGATGGTGCTGTATAGG

  Mouse, primary enzyme: *DpnII*, secondary enzyme: *NlaIII,* viewpoint: *Amn* gene, iF sequencing primer: ATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCAGGCGCACTCTTAGCTG, iR sequencing primer: CAAGCAGAAGACGGCATACGATATGGTAAGGCTCGGGGCTG

5. If less than 1×10^7 cells can be isolated at once, the procedure should be scaled down. Multiple smaller samples of cross-linked nuclei can be pooled prior to continuing **step 3.2**.
6. Depending on the type of cells, the cell lysis step may need to be optimized. The following cell lysis buffers have been successfully used as well:
   Primary mammalian cells and cell lines, insect cell lines: 10 mM Tris–HCl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitor cocktail, EDTA free [12, 21] or 10 mM Tris–HCl, pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 1× complete protease inhibitor cocktail, EDTA free [27].
   Primary insect cells: 15 mM Tris–HCl, pH 7.4, 0.34 M sucrose, 15 mM NaCl, 60 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 1× cOmplete protease inhibitor cocktail, EDTA free [28, 29].
7. Upon addition of the restriction buffer, the nuclei may visibly aggregate, which negatively influences the efficiency of the restriction enzyme reaction. A short incubation of the nuclei at higher temperature (60 °C) in the presence of 0.3% SDS will dissociate the aggregates, resulting in a somewhat milky solution without visible aggregates. The duration of the exposure to high temperature should be minimized though, as it negatively influences the integrity of formaldehyde cross-links.
8. The large volumes of ligation buffer result in large quantities of salt in the DNA solution, which are incompletely removed in the following cleanup steps (phenol–chloroform–IAA, ethanol, and QIAquick Spin Columns). Particularly DTT contamination results in a strong absorption at 260 nm, thereby interfering with the use of standard spectrophotometric determination of DNA concentration (e.g., using a NanoDrop). Dye incorporation based assays (e.g., Qubit or PicoGreen) are insensitive to the presence of salt contamination, and therefore provide a more reliable output. After PCR amplification, dye incorporation can discriminate between double stranded and single stranded DNA as well, thereby accurately quantifying the amounts of PCR-amplified DNA without measuring unincorporated primers.

Acknowledgements

We thank Céline Hernandez for comments on the manuscript. Work in the D.N. laboratory is supported by funds from the Fondation pour la Recherche Médicale (FRM—Amorçage de jeunes équipes 2014, grant AJE20140630069), the Biologie Intégrative des Génomes project funded by the Inititiative d'Excellence Paris-Saclay (ANR-11-IDEX-0003-02) and the Centre National de la Recherche Scientifique (CNRS).

References

1. de Laat W, Duboule D (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502(7472): 499–506. doi:10.1038/nature12753
2. Gorkin DU, Leung D, Ren B (2014) The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell 14(6):762–775. doi:10.1016/j.stem.2014.05.017
3. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10(6):1453–1465
4. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A 103(28):10684–10689. doi:10.1073/pnas.0600326103
5. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376–380. doi:10.1038/nature11082
6. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Bluthgen N, Dekker J, Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485(7398):381–385. doi:10.1038/nature11049
7. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153(6):1281–1295. doi:10.1016/j.cell.2013.04.053
8. Tiwari VK, McGarvey KM, Licchesi JD, Ohm JE, Herman JG, Schubeler D, Baylin SB (2008) PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol 6(12):2911–2927. doi:10.1371/journal.pbio.0060306
9. Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol 9(10):1167–1174. doi:10.1038/ncb1637
10. Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D (2011) The dynamic architecture of Hox gene clusters. Science 334(6053):222–225. doi:10.1126/science.1207194
11. Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F, Duboule D (2014) Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. Elife 3:e02557. doi:10.7554/eLife.02557
12. Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, Cavalli G (2011) Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell 144(2):214–226. doi:10.1016/j.cell.2010.12.026
13. Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M, Simonis M, de Laat W, van Lohuizen M, van Steensel B (2011) Interactions among Polycomb domains are guided by chromosome architecture. PLoS Genet 7(3):e1001343. doi:10.1371/journal.pgen.1001343
14. Denholtz M, Bonora G, Chronis C, Splinter E, de Laat W, Ernst J, Pellegrini M, Plath K (2013) Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell 13(5):602–616. doi:10.1016/j.stem.2013.08.013
15. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. doi:10.1126/science.1067799
16. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38(11):1348–1354. doi:10.1038/ng1896
17. David FP, Delafontaine J, Carat S, Ross FJ, Lefebvre G, Jarosz Y, Sinclair L, Noordermeer D, Rougemont J, Leleu M (2014) HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. PLoS One 9(1):e85879. doi:10.1371/journal.pone.0085879
18. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, Verstegen MJ, de Wit E, Tanay A, de Laat W (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods 9(10):969–972. doi:10.1038/nmeth.2173
19. Klein FA, Anders S, Pakozdi T, Ghavi-Helm Y, Furlong EEM, Huber W (2014) FourCSeq:

analysis of 4C sequencing data. bioRxiv. doi:10.1101/009548

20. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B (2013) r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res 41(13):e132. doi:10.1093/nar/gkt373
21. Gheldof N, Leleu M, Noordermeer D, Rougemont J, Reymond A (2012) Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. Methods Mol Biol 786:211–225. doi:10.1007/978-1-61779-292-2_13
22. de Wit E, Braunschweig U, Greil F, Bussemaker HJ, van Steensel B (2008) Global chromatin domain organization of the Drosophila genome. PLoS Genet 4(3):e1000045. doi:10.1371/journal.pgen.1000045
23. van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W (2012) 4C technology: protocols and data analysis. Methods Enzymol 513:89–112. doi:10.1016/B978-0-12-391938-0.00004-5
24. Kernohan KD, Vernimmen D, Gloor GB, Berube NG (2014) Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. Nucleic Acids Res 42(13):8356–8368. doi:10.1093/nar/gku564
25. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung HK, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui CC, Gomez-Skarmeta JL, Nobrega MA (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507(7492):371–375. doi:10.1038/nature13138
26. Woltering JM, Noordermeer D, Leleu M, Duboule D (2014) Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. PLoS Biol 12(1):e1001773. doi:10.1371/journal.pbio.1001773
27. Simonis M, Kooren J, de Laat W (2007) An evaluation of 3C-based methods to capture DNA interactions. Nat Methods 4(11):895–901. doi:10.1038/nmeth1114
28. Bonn S, Zinzen RP, Perez-Gonzalez A, Riddell A, Gavin AC, Furlong EE (2012) Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. Nat Protoc 7(5):978–994. doi:10.1038/nprot.2012.049
29. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE (2014) Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512(7512):96–100. doi:10.1038/nature13417

Chapter 21

Chromosome Conformation Capture on Chip (4C): Data Processing

Benjamin Leblanc*, Itys Comet, Frédéric Bantignies, and Giacomo Cavalli

Abstract

4C methods are useful to investigate dependencies between regulatory mechanisms and chromatin structures by revealing the frequency of chromatin contacts between a locus of interest and remote sequences on the chromosome. In this chapter we describe a protocol for the data analysis of microarray-based 4C experiments, presenting updated versions of the methods we used in a previous study of the large-scale chromatin interaction profile of a Polycomb response element in *Drosophila*. The protocol covers data preparation, normalization, microarray probe selection, and the multi-resolution detection of regions with enriched chromatin contacts. A reanalysis of two independent mouse datasets illustrates the versatility of this protocol and the importance of data processing in 4C. Methods were implemented in the R package MRA.TA (Multi-Resolution Analyses on Tiling Array data), and they can be used to analyze ChIP-on-chip data on broadly distributed chromatin components such as histone marks.

Key words Epigenetics, Chromatin, Polycomb, Chromosome Conformation Capture, 4C, Microarray, Bioinformatics, Normalization, Multi-resolution statistics

1 Introduction

The Chromosome Conformation Capture (3C) [1] and derived genomic methods are key molecular approaches to uncover the links between the linear organization of genetic information on the genome sequence, its functional regulation, and its three-dimensional structures [2, 3]. A remarkably effective principle is at the origin of these methods: in appropriate conditions, after cell fixation and controlled chromatin fragmentation, the ligation probability of two separate genomic sequences depends mainly on their spatial proximity within the nuclear space (Fig. 1a). Following this principle, 3C-based techniques convert genome structures into frequencies of ligation products among genomic sequences. The structural information represented by 3C ligation frequencies covers a wide range of molecular scales, from few kilobases to whole chromosomes. This encompasses chromatin structures

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_21,

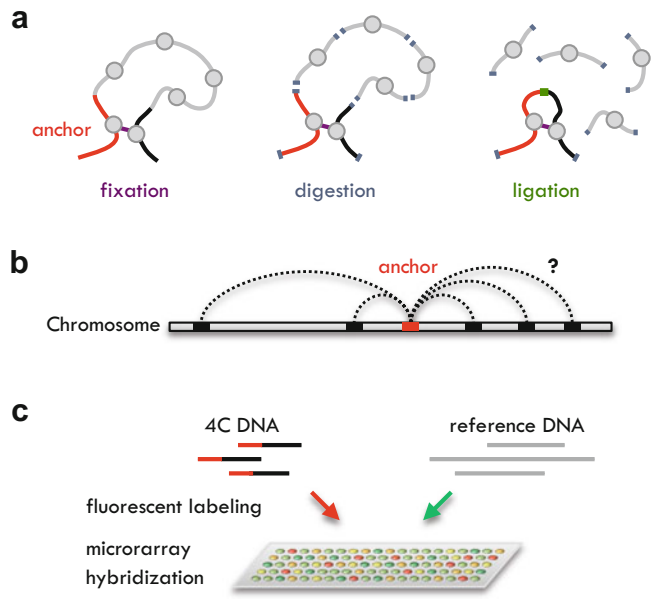


Fig. 1 Molecular analysis of chromatin conformation by 4C. (**a**) Chromosome Conformation Capture (3C) based experiments consist in three key steps allowing the transformation of chromatin contacts into DNA ligation events. (**b**) By performing a purification of 3C ligation products containing a specific sequence (*red*), the 4C approach interrogates the relative frequency of interactions between remote genomic sequences (*black*) and this specific DNA element, usually named anchor or bait fragment. (**c**) The DNA library of purified 3C ligation products containing the anchor sequence can be analyzed together with a reference library using dual channel microarray hybridization

known to be important for transcriptional regulation, such as enhancer–promoter interactions, which were difficult to address experimentally before the introduction of the 3C approach [4]. In the study of Polycomb group (PcG) protein activities, 3C-based methods allow to investigate chromatin contact frequencies between PcG-regulated genes and PcG-recruiting sequences [5, 6]. PcG-dependent maintenance of gene repression requires the posttranslational modification of histone H3 on lysine 27 (H3K27me3) which is a key activity of the Polycomb Repressive Complex 2 (PRC2). The H3K27me3 histone mark is broadly distributed on the genome and has been found to correlate with distinct topological domains [7, 8] which are chromosomal segments spanning approximately 200 kb to 1 Mb in mammals and characterized by high frequencies of internal chromatin contacts [9, 10].

Among 3C-derived methods, 4C designate a family of protocols designed to interrogate chromatin contact frequencies between a specific locus, named *anchor* or *bait*, and any remote sequence in the genome (Fig. 1b). This approach is particularly suited for the study of chromatin interactions involving gene and regulatory elements at a locus of interest. Technically, 4C has been implemented

with varying strategies. Depending on protocols 4C stands for "Circular Chromosome Conformation Capture" [11, 12] or "Chromosome Conformation Capture on Chip" [13]. In initial studies, plasmid circularization of DNA fragments is employed in order to identify 3C ligation products containing the anchor fragment [11, 13], whereas alternative strategies involve primer extension from the anchor sequence [14, 15]. As in 3C, fragmentation of chromatin is performed by restriction enzymes cutting DNA at 4 or 6 bp specific motifs, which define the theoretical resolution of the technique. When combined to chromatin immunoprecipitation (ChIP), the modified procedure allows to access chromatin contact frequencies at the subset of genomic loci that interact with a protein of interest [14]. The content of DNA libraries generated by the 4C procedure, which represent chromatin contact frequency information, can be analyzed either by sequencing (for instance refs. 11, 12, 16, 17) or by microarray hybridization (Fig. 1c) (for instance refs. 11, 13–15, 18, 19).

Prominent difficulties of microarray-based 4C data analysis are due to the limited sensitivity of the molecular approach combined with the limited range of measurement levels offered by microarrays in comparison to random chromatin contact probabilities that vary exponentially with genomic distances on the chromosome [20]. Sensitivity in 4C experiments is constrained by the use of an anchor fragment representing two copies in the genome of each diploid cell. In consequence, 4C protocols can reveal a maximum of four (in the G2 phase of the cell cycle) 3C ligation products per cell.

Here we describe data processing procedures for the visualization and localization of most frequent chromatin contacts from 4C combined to microarray hybridization, covering the normalization of raw data, the selection of relevant probes when using tiling array designs and the statistical analysis of 4C enrichments at multiple resolutions. The initial version of this protocol was used in our previous study investigating chromosome-wide interactions of the *Fab-7* Polycomb Response Element (PRE), a *cis*-regulatory element recruiting PcG proteins at the homeotic complex bithorax (BX-C) in *Drosophila* [15]. We present here an updated protocol with several improvements. We have introduced an algorithmic definition of the "background bias" concept used to balance the influence of the reference DNA library (detailed in Subheading 1.1), a more general approach to match and select probes according to restriction fragments of the 4C library (detailed in Subheading 1.2), and in complement with the multi-resolution representation of 4C enrichments, a segmentation strategy which localizes 4C-enriched domains at maximum resolution and maximum scale (detailed in Subheading 1.3). We have reanalyzed three independent datasets [13–15] to illustrate the effectiveness of this updated procedure on previously published data (Fig. 2, *see* Subheading 1.4).

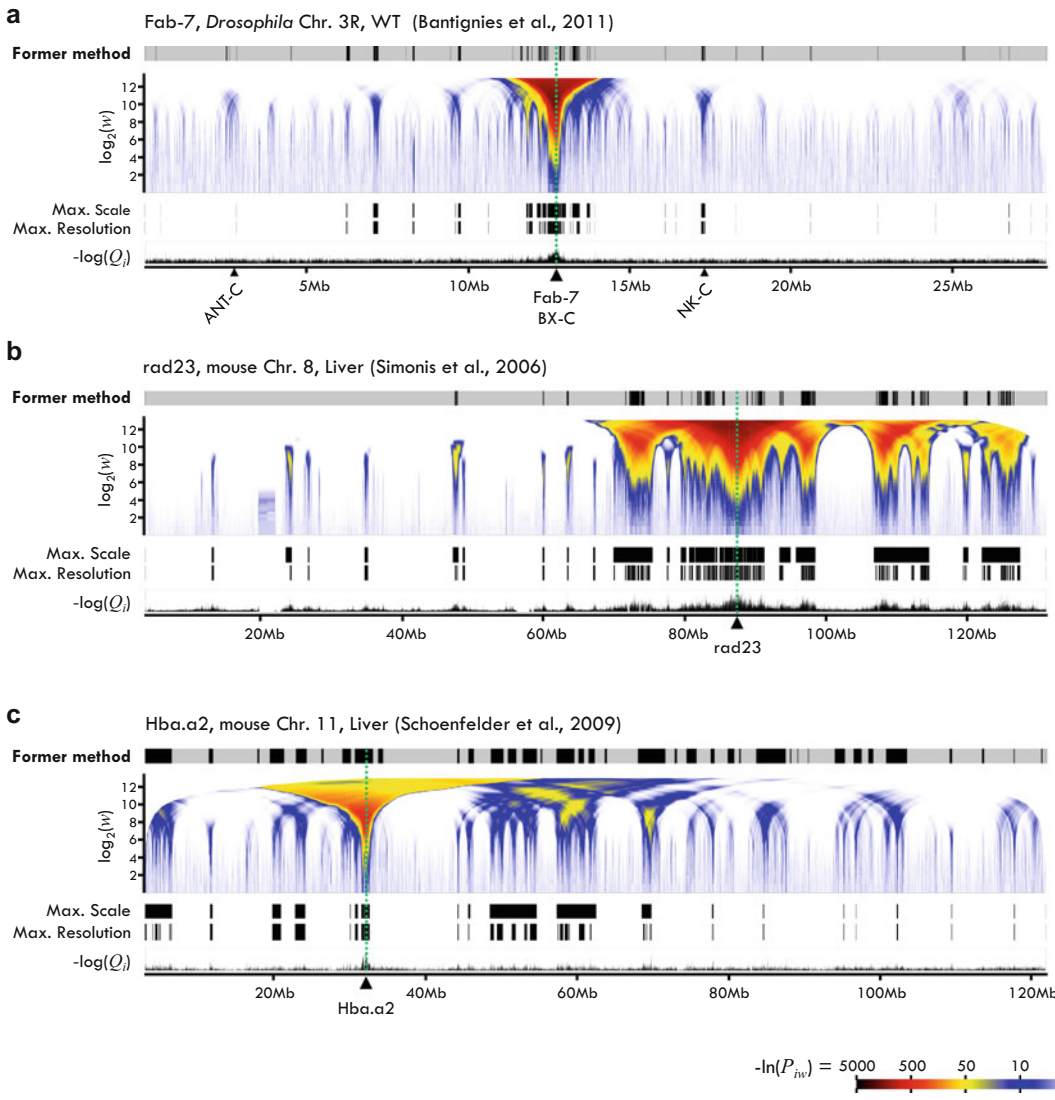


Fig. 2 Examples from the reanalysis of three datasets. (**a**) Representation of 4C anchored at the *Fab-7* element in the homeotic complex bithorax (BX-C) on chromosome 3R in *Drosophila* embryos [15], revealing large-scale enriched contact frequencies with other Polycomb-regulated gene complexes including antennapedia (ANT-C) and NK-C. (**b**) 4C anchored at the *rad23* locus on chromosome 8 in mouse liver tissues [13]. (**c**) 4C including a chromatin immunoprecipitation step targeting RNA polymerase II, anchored at the *Hba* locus on chromosome 11 in mouse liver tissues [14]. For each panel, the *x*-axis represents chromosome coordinates in Mb. On the domainograms (*color graph*), the *y*-axis represents window size (noted *w*) as the $\log_2$ of the number of microarray probes, and multi-resolution 4C enrichment scores (noted *Piw*) are represented by the color gradient. Anchor fragment locations are indicated by *green vertical dotted lines*. 4C-enriched domains identified by reanalyses of the published 4C data using our procedure are indicated in black on the Max. Scale and Max. Resolution tracks. *Bottom graphs* represent per-probe 4C enrichment scores, here noted –log (*Qi*). Former method in **a**: Originally identified maximum resolution domains (unpublished data). Former method in **b**: Published enriched interactions [13] with genomic coordinates updated from MM6 to MM9 mouse genome release. Former method in **c**: Published enriched interactions [14] with genomic coordinates updated from MM8 to MM9 mouse genome release. Updates of genomic coordinates were performed using the liftOver utility from UCSC (http://genome.ucsc.edu/cgi-bin/hgLiftOver)

The next sections describe the methods used for normalization, probes selection and localization of 4C-enriched chromatin domains (*see* Subheading 3).

1.1 Normalization

Microarray-based 4C experiments rely on the simultaneous hybridization of a 4C DNA library together with a reference DNA library, and the calculation of the $\log_2$ ratio between the 4C and reference channels as an indicator of 3C ligation frequencies for each probe, assuming that biases possibly introduced by the 4C protocol and the hybridization procedure should—ideally—be eliminated. However, reference libraries usually employed in 4C protocols, for instance based on purification and digestion of genomic DNA, do not represent a control of random 3C ligation products and generally imply that the structure of DNA sequences in the 4C and reference libraries have marked differences. Such differences can translate into a severe inconsistency of probe hybridization and fluorescence levels between the two microarray channels, lowering the relevance of a conventional $\log_2$ ratio between 4C and reference data as a good indicator of 3C ligation frequencies.

Our normalization strategy aims to balance the influence of the reference (or control) channel when computing the $\log_2$ ratio representing 4C enrichments. We refer to this strategy as background bias estimation and correction. We use the term background measurements (or signals) to designate measurements from the expected dense population of probes that should indicate no 4C enrichment and thus a $\log_2$ ratio distribution centered on zero independently of technical parameters. We first calculate A (average $\log_2$ intensity in both channels) and M ($\log_2$ ratio of 4C over control intensities) values as conventionally employed in microarray analysis [21]. We observe that these values are linked to the raw 4C and reference data by scaling factors and by a –45° rotation in the (A,M) plane (Fig. 3a). We define the background bias as the slope of the densest sub-population of probe measurements in the rescaled (A,M) plane, and apply a "background traversal" algorithm to provide an estimation of this slope (Fig. 3b, c). The bias correction consists in rotating the rescaled A and M values in order to eliminate the estimated slope (Fig. 3d). In this approach, if background measurements are distributed along the horizontal axis as expected, the estimated slope is zero degrees and the corrected M values, reflecting 3C ligation frequencies, are equal to the rescaled $\log_2$ ratio. In contrast, when the slope of background measurements deviates from the horizontal axis, the impact of the reference channel on corrected M values is proportionally reduced. In the extreme case of an estimated bias attaining the limit of 45° (Fig. 3a), indicating a severe inconsistency in the raw data, the influence of the reference channel is eliminated and the corrected M values are equal to measurements from the 4C channel only (Fig. 3a). When tested with naive simulations, bias estimations provided by the background traversal algorithm have an accuracy of ±2.5° in 95% of the cases.

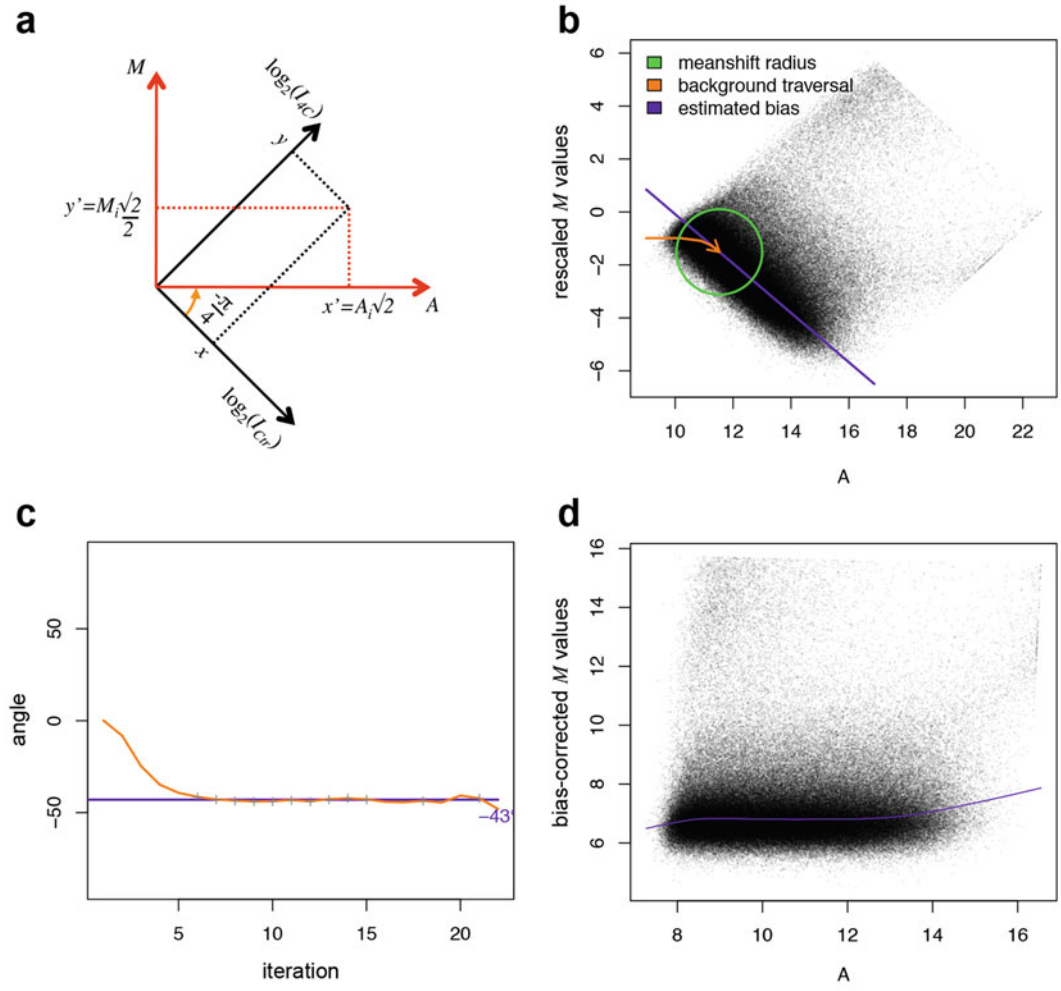


Fig. 3 Background bias estimation and correction in 4C data. (**a**) This scheme indicates the scaling factors and −45° rotation angle in the (*A*,*M*) plane that link *A* (average of $\log_2$ intensities) and *M* ($\log_2$ ratio) values to *I4C* and *ICtr* values which represent raw microarray data expressing the fluorescence intensity in 4C and reference DNA channels respectively. (**b**) Example of rescaled MA plot on raw 4C data [15] showing a strong inconsistency between 4C and reference channels. To quantify and balance such inconsistency we define the background bias as the slope (*purple line*) of the densest population of probes in the rescaled (*A*,*M*) plane. Estimation of this slope is derived from angles of displacement along the "background traversal" curve (*orange*). The curve is determined by iterative mode seeking (or meanshift [42]) within a defined radius (*green circle*), starting at probes with minimal *A* values and converging to a maximum density of the (*A*,*M*) values distribution (*arrow* of the *orange curve*). (**c**) Angle value (*orange curve*) at each iteration of the background traversal shown in panel **b**. In this example the estimated bias angle is 43° (*purple line*). (**d**) MA plot of the 4C data shown in panel **b** after correction of the estimated bias. The *purple curve* represents a LOWESS regression on these corrected data

Additionally, bias-corrected *A* and *M* values are normalized using the standard LOWESS [21, 22] method in order to compensate for nonlinear differences between fluorescence responses of the microarray in each channel, and to center the distribution of background level *M* values on zero.

1.2 Probes Filtering

Measurement of enriched sequences in 4C libraries can be achieved with restriction enzyme-specific arrays (for instance refs. 13, 14, 23, 24) or tiling arrays (for instance refs. 15, 25–27). By design, enzyme-specific arrays contain probes that only map to extremities of restriction fragments whereas tiling arrays map the genome at nearly constant resolution (Fig. 4a). Depending on the restriction fragment sizes and the probes spacing on the genome, tiling arrays

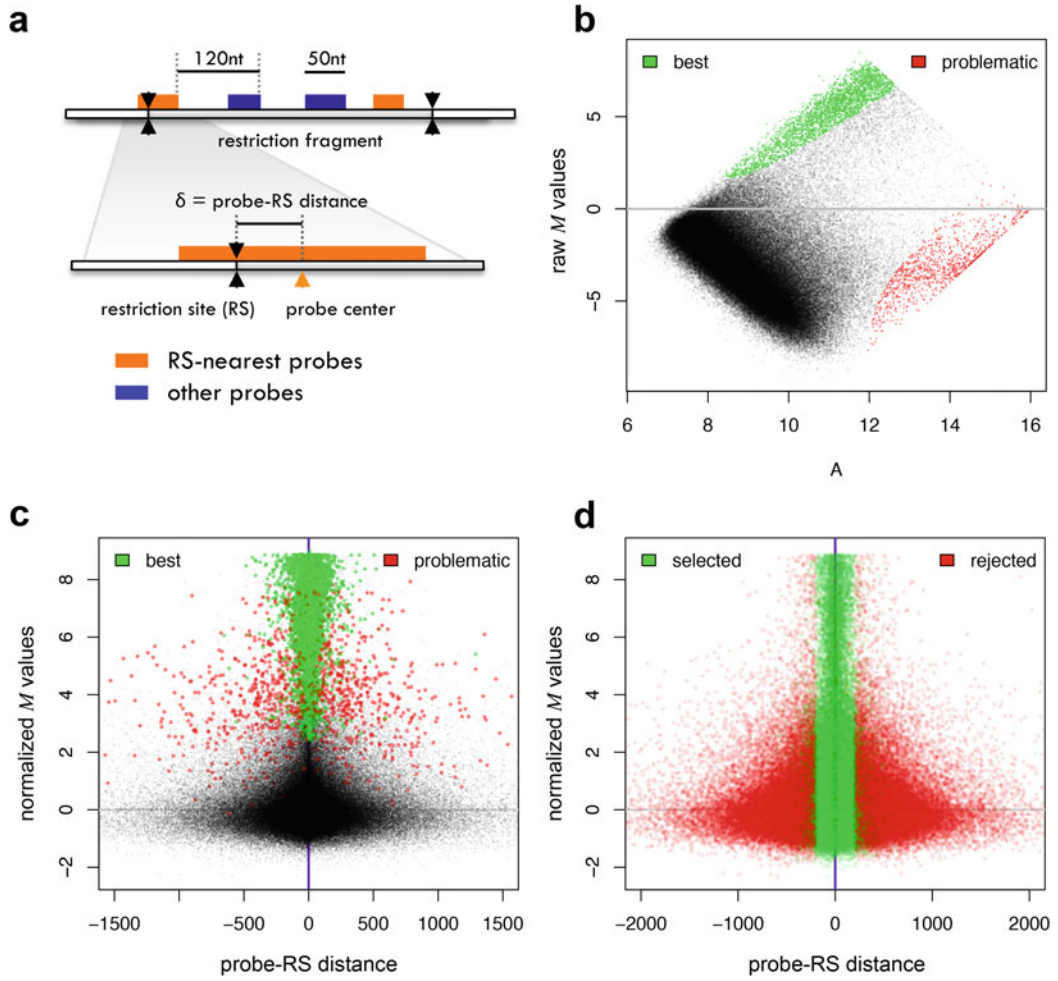


Fig. 4 Filtering of tiling array probes. (**a**) Illustration of probe and restriction site locations on a fragment of the chromosome with the tiling array design used for our original 4C study [15]. *Black arrows* represent restriction sites (RS). *Orange boxes* represent microarray probes that are nearest to restriction sites (ranked position = 1). *Blue boxes* represent other probes (ranked position ≥ 2). Below is represented a closer view with the definition of the distance (δ) between the probe center and the nearest restriction site. (**b**) MA plot of raw 4C data [15] where probes associated with best and problematic 4C enrichments are highlighted in *green* and *red* respectively. (**c**) Control graph on the 4C data shown in panel b. *x*-axis indicates probe distance to restriction site; *y*-axis represents the normalized *M* values reflecting 4C enrichments. Probes associated with best 4C enrichments are close to restriction sites and in majority correspond to a ranked position of 1 or 2 (nearest or second nearest probe). (**d**) Same control graph as in panel c showing selected (*green*) and rejected (*red*) probes

may provide replicate measurement of 4C enrichments via multiple probes mapping the same restriction fragment. However, 4C protocols which use primer extended fragments—or which linearize circular DNA by utilizing a secondary restriction enzyme—do not preserve 3C-ligated fragments in their full length, and only the subset of probes that map to the vicinity of primary restriction sites will hybridize with DNA sequences present in the 4C library. As a consequence, analyzing 4C data based on tiling arrays implies to define how probes are taken into account in the computation of 4C enrichments for each extremity of the restriction fragments.

Our procedure relies on two quality controls to determine a subset of tiling array probes that are the most relevant for the computation of 4C enrichments. A preliminary control consists in categorizing probes that are likely associated with best and problematic 4C enrichments (Fig. 4b). We define these two populations of probes with a rank-based scoring of the raw data. Briefly, our scoring assumes that the best 4C enrichments are expected near maximal $\log_2$ ratios (Fig. 4b, green), and that probes are problematic when presenting neutral or low $\log_2$ ratios coupled with maximal average levels in the two channels (Fig. 4b, red), as these probes can represent very high corrected *M* values in case of a pronounced background bias (Fig. 3b, d).

The main control consists in representing the distribution of normalized *M* values as a function of probe distance to the nearest restriction site (Fig. 4a). As 4C library preparation do not preserve the full-length of 3C-ligated restriction fragments, this representation is expected to show a decreasing trend on normalized *M* values as probe distances increase. Consistently, the identified best and problematic 4C enrichments should also reflect differences in probe distances to the nearest restriction site (Fig. 4c, green and red). Practically, these two controls help defining the set of probes accepted for the computation of 4C enrichments (Fig. 4d) by choosing a probe to restriction site distance cutoff that is relevant for the 4C protocol (*see* examples given in the Subheading 3).

1.3 4C-Enriched Chromatin Domains

Enriched sequences in the 4C libraries reflect the frequency of 3C ligation products between the anchor fragment and remote restriction fragments on the chromosome. But due to experimental and technical limitations, 4C enrichment profiles on the genome are difficult to interpret at the full resolution of the restriction map (Fig. 2). Locally, these profiles exhibit a poor correlation of 4C enrichments at both ends of the same fragment or between consecutive fragments and can be very noisy (Fig. 2a). Nevertheless, initial studies successfully revealed regions of enriched chromatin interactions by applying statistical methods based on fixed-size sliding windows to the 4C enrichment profiles [13, 14, 23]. A limitation in such methods is the choice of the window size, which

can impact both sensitivity and resolution when identifying 4C-enriched domains. This limitation was elegantly addressed by de Wit and colleagues in their study of the global chromatin domain organization in *Drosophila* [28]. Technically, the approach consists in generalized nonparametric windowed statistics allowing a multi-resolution representation of the genomic profile—the domainogram (*see* examples in Fig. 2)—and associated with a segmentation algorithm that identifies an optimal set of enriched domains without a priori knowledge of the relevant window sizes. This approach has found successful application in different microarray-based 4C studies [15, 24, 29].

Our procedure employs the same nonparametric windowed statistics (Fig. 5a, b) and domainogram representation introduced by de Wit et al. [28], coupled to a variant technique for the segmentation, and followed by a hierarchical filtering algorithm which determines the localization of 4C-enriched chromatin domains at maximum resolution and maximal scale. First, the segmentation algorithm computes the complete set of arbitrary sized genomic windows presenting locally optimal 4C enrichment scores (Fig. 5c, d). Then, the hierarchical filtering algorithm constructs a multi-resolution tree from resulting genomic segments (Fig. 5e, left). In this process, the genomic coordinates of each segment are updated according to logical constraints determining the internal consistency of the multi-resolution tree, and the redundant segments are eliminated (Fig. 5e, right). Once the tree construction is completed, the algorithm extracts the localization of maximum resolution and maximal scale domains (Fig. 5e, right; *see* examples in Fig. 2).

1.4 Discussion and Conclusion

Methods presented here address basic issues of 4C data processing, as illustrated by a reanalysis of our own data and of two independent datasets in mouse (Fig. 2). Briefly, our implementation of the 4C approach combined short size chromatin fragments (200–300 bp on average) generated by a 4 bp motif restriction enzyme (DpnII) with a primer extension strategy and tiling array hybridization [15]. The original 4C protocol introduced by Simonis et al. was based on long chromatin fragments (>1 kb) generated by a 6 bp motif restriction enzyme, DNA circularization and enzyme-specific arrays [13]. The 4C protocol implementation in Schoenfelder et al. included a ChIP procedure targeting RNA polymerase II and used long chromatin fragments and enzyme-specific arrays [14]. In our original study, we observed an extreme bias on raw data and we used transformed $\log_2$ ratio of 4C enrichments nearly eliminating the influence of the reference library. The background bias definition introduced in the updated normalization procedure strongly supports this choice with bias estimations ranging between 40° and 43° on these data. It is noteworthy that the bias estimation method consistently matches the exchange of fluorescence channels (dye swapping) between replicates on 4C data

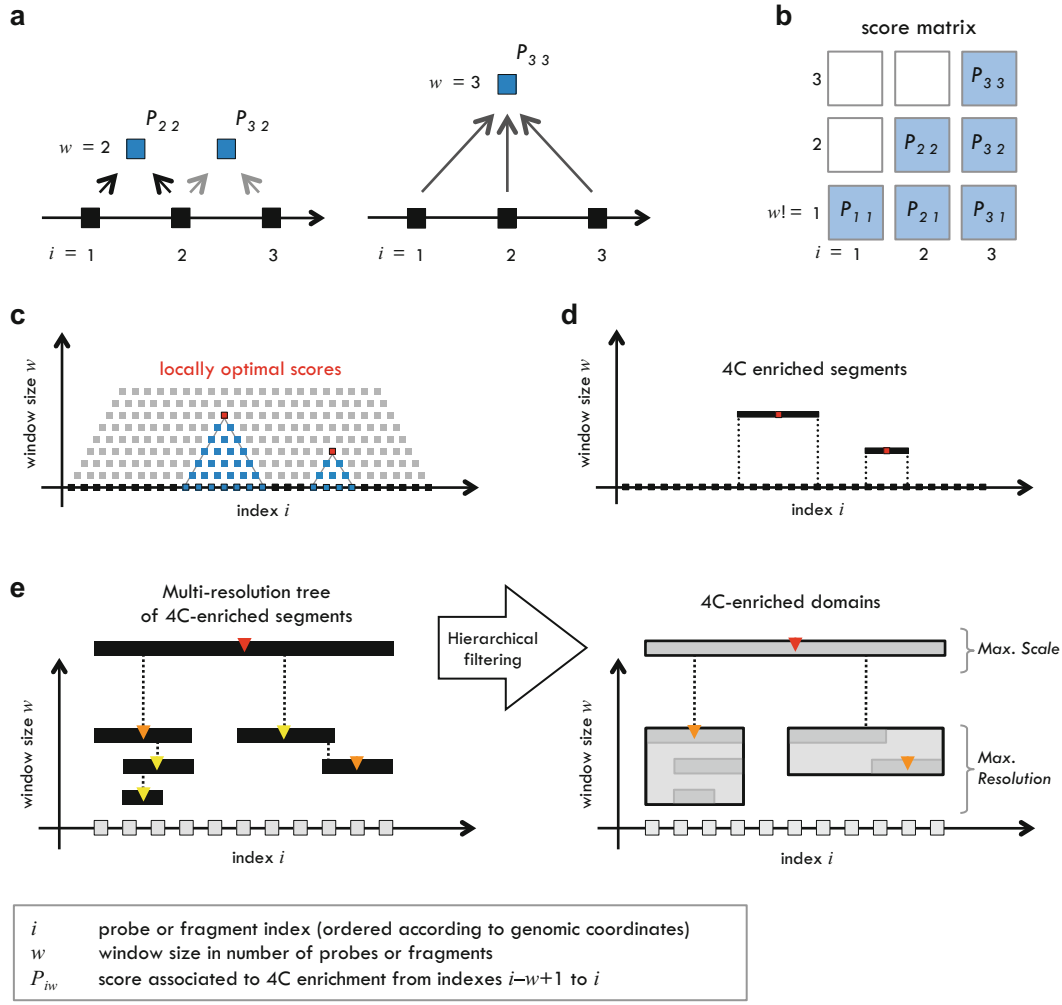


Fig. 5 Localization of 4C-enriched domains. (**a**). Multi-resolution statistics principle applied to three consecutive measurements. Black boxes represent probe or fragment locations on the chromosome. Blue boxes represent windows of two consecutive measurements (*left*) and three consecutive measurements (*right*). *Arrows* indicate measurements taken into account to calculate the 4C-enrichment score for each window. (**b**) Complete matrix of 4C-enrichment scores (noted *Piw*) representing multi-resolution statistics for three measurements. (**c**) The 4C-enrichment score *Piw* associated to a genomic window is locally optimal (*red*) when it is stronger than the scores associated to any sub-window (*blue*). (**d**) By definition, a locally optimal *Piw* value corresponds to a precisely localized 4C-enriched segment. (**e**) Left scheme illustrates the result of a 4C-enrichment profile segmentation based on locally optimal scores. Right scheme represents the corresponding 4C-enriched domains determined by the hierarchical filtering algorithm. On the *x*-axis, *grey boxes* represent probe or restriction fragments indexed along the chromosome. *Black boxes* represent 4C-enriched segments in the *x*= *genomic location*, *y* = *window size* plane. *Colored arrows* represent locally optimal 4C-enrichment scores (*yellow* = strong, *orange* = stronger, *red* = strongest)

from Simonis et al., with estimated biases ranging from approximately 16° to 13° before and 0° to −5° after dye swap. Together with bias estimations ranging from 22° to 36° on data from Schoenfelder et al. these reanalyses indicate that our bias estimation procedure can reflect technical specificities of different 4C protocols. Results

of the multi-resolution segmentation on re-normalized 4C data from the two independent studies globally overlap with regions originally identified with fixed resolution analyses (Fig. 2b, c). However these reanalyses illustrate the direct influence of different data analysis methods on the number and precise localization of segmented 4C-enriched domains. Our normalization and multi-resolution approach can lead to the distinction of additional variations of chromatin contacts at short distances from the anchor locus and to increased detection of long-distance chromatin contacts (Fig. 2b). These observations do not constitute a quantitative assessment of the present protocol, but they show its relevance to analyze data generated by different 4C techniques.

A main limitation of this protocol is that the multi-resolution analysis does not provide false positive rate estimations on 4C-enriched domains. This can be partially overcome by the use of random permutations in order to empirically estimate significance cutoffs [15, 28]. It should also be noted that the protocol does not treat separately probes targeting repeat sequences and does not take into account possible clusters of missing data in case of poor combination of the restriction map, microarray design and presence of repetitive elements. Another limitation is the absence of correction for the exponential decrease of random chromatin interactions between the anchor and surrounding sequences at increasing genomic distances. Addressing this issue with 4C data acquired via microarray hybridization is beyond the scope of our methods. However a workaround has been proposed by Tolhuis and colleagues in their study of chromatin interactions among Polycomb domains [24] with an associated 4C data processing protocol that is also based on the domainogram approach [19].

Since the early 4C studies using microarray hybridization to reveal the frequency of 3C ligations among genomic fragments, sequencing has become the prevalent method to analyze the content of genomic libraries. In the case of 4C and all 3C-derived approaches, sequencing has major advantages compared to microarrays. The range of investigated 3C ligation frequencies is only limited by sequencing depth and sequencing offers the possibility to control and correct experimental biases with a far greater level of detail, revealing for instance that the length of restriction fragments actually influences 3C ligation frequencies [30]. In recent years, methods have been developed specifically for the analysis of 4C interaction profiles based on high-throughput sequencing data [31–36], including adaptations and extensions of the domainogram representation and underlying multi-resolution statistics [31, 32, 37].

Prior to these developments, the initial version of the microarray-based 4C data analysis protocol presented here allowed us to establish that, at the molecular level, a well characterized PRE in *Drosophila* exhibits its most frequent chromatin contacts with other Polycomb-regulated chromatin domains, even over

large (up to 10 Mb) genomic distance. A similar observation was reported in the study of Tolhuis and colleagues [24] and further confirmed within a complete 3C sequencing (Hi-C) study [7].

2 Materials

All presented data analysis methods were implemented in the R package MRA.TA (Multi-Resolution Analyses on Tiling Array data) which is available online (http://purl.org/NET/MiMB.4C) together with additional scripts for basic processing of raw microarray data and computation of restriction maps (*see* Subheading 3). Multi-resolution methods in the MRA.TA package are also relevant for the analysis of ChIP-on-chip data and using our protocol should only require minor adaptations when processing data generated with the Roche Nimblegen or Agilent Technologies platforms.

The following procedure should be compatible with standard GNU/Linux distributions and the R statistical environment [38] version 3.2 or higher. As a prerequisite it needs R packages devtools, stringr, getopt, plotrix, and the Bioconductor [39] packages Biostrings and GenomicRanges to be installed. Performing an update of microarray probe coordinates to a recent genome assembly additionally requires installation of the short read aligner bowtie [40] version 1.1.x.

3 Methods

1. Workflow

3.1 Overview

Practically the protocol covers operations starting from raw data to the production of graphs representing 4C-enrichment profiles at multiple resolutions, and tables with genomic coordinates of chromatin domains presenting the strongest 4C-enrichments. The workflow is divided in three main steps. The first step is the retrieval and preparation of data, including the update of microarray probe coordinates and computation of the restriction map on the latest genome assembly. The second step consists in the normalization of raw 4C data, matching microarray probes to restriction fragments, and filtering probes based on quality controls. The third step allows the visualization and localization of chromatin domains presenting the strongest 4C enrichments.

Data preparation operations are executed in the standard GNU/Linux terminal whereas 4C data analysis tasks are performed within the R statistical environment. Excluding data preparation, running the analyses on microarray data addressing 10^5 chromosomal fragments (for example, corresponding to DpnII

restriction sites of chromosome arm 3R in *Drosophila*) requires approximately 7 min with an up-to-date desktop computer.

In the following, text boxes represent examples of source code and commands to execute, including comments highlighted in blue. For the main text, we use a monospace style to highlight words that designate identifiers, for instance file names, variable names, and so on.

3.2 Installation and Documentation

Download and decompress the MiMB.4C archive from http://purl.org/NET/MiMB.4C. A brief description of the archive content and the organizing principle of included tools are given by the README file. Detailed information on the usage and parameters of presented R functions is available as built-in documentation of the MRA.TA package.

Open the R environment to install the MRA.TA package:

```
library("devtools")
install_github("benja0x40/MRA.TA")
```

The following sections describe the main operations of our protocol. A complete sequence of operations is recapitulated in the dataPreparation.sh and enrichmentAnalysis.R scripts included in the MiMB.4C archive. Both scripts are executable for demonstration. In the terminal:

```
./daraPreparation.sh # download ≈400 MB, execution ≈8 minutes with 6 CPU cores
Rscript enrichmentAnalysis.R
```

4 Data Preparation

4.1 Importation of Genome Sequence and 4C Data

Download the dm6 assembly of the *Drosophila* genome sequence from UCSC and create bowtie indexes for alignments. Download the 4C dataset of our original 4C study from the Gene Expression Omnibus database [41]. In the terminal:

```
./importGenome.sh"dm6"
./importRawData.sh "GSE23887"
```

4.2 Updating Array Design Data

Align microarray probe sequences to the genome and update array design with the resulting genomic coordinates. In the terminal:

```
Rscript updateDesignData.R -i "dm6" -p "Genome_Data/UCSC_dm6/Indexes/bowtie/genome" -o "Raw_Data/GSE23887_RAW/GPL10867.ndf.gz"
```

Filter out probes that serve as internal controls for hybridization and fluorescence imaging.

4.3 Computation of the Restriction Map

Compute restriction fragments for the DpnII enzyme, which cuts DNA sequences on GATC motifs. In the terminal:

```
Rscript computeRestrictionMap.R -i "dm6" -n
"DpnII" -m "GATC" -s
"Genome_Data/UCSC_dm6/genome.fa.gz"
```

5 Processing of Raw 4C Data

5.1 Loading R Functions and raw 4C Data

Open the R environment and load the `MRA.TA` package. The experimental setup corresponding to the downloaded 4C dataset is documented by the `Experiment_Design/GSE23887_dm6.txt` table, which is used to load raw 4C data from each sample. In the R environment:

```
library("MRA.TA")
# Read predefined experimental setup
experiment.design <- read.delim
("Experiment_Design/GSE23887_dm6.txt",
stringsAsFactors=F)
# Read raw data
array.data.format <- file.formats$nimblegen.pair
r1.ct   <-   readData(experiment.design$SAMPLE_PATH[1],
array.data.format)
r1.4C   <-   readData(experiment.design$SAMPLE_PATH[2],
array.data.format)
```

Match raw 4C data to the updated array design information using probe identifiers.

5.2 Matching Array Probes to Restriction Fragments

Read precomputed restriction fragments and define the start and end positions depending on the enzyme cut site within each motif (*see* **Note 1**). Make genomic intervals `probes.grg` and `fragments.grg` with the constructor function `GRanges`, using probe and restriction fragment identifiers and genomic coordinates respectively. Match probes to fragments using these genomic intervals.

```
probes.grg <- matchProbesToFragments(probes.grg, frag-
ments.grg)
```

Reorder data consistently with matched probes.

5.3 Normalization with Background Bias Estimation and Correction

Compute raw A and M values. Apply background bias estimation and correction followed by LOWESS normalization (*see* **Note 2**).

```
r1.A <- (log2(r1.4C$PM) + log2(r1.ct$PM))/2
r1.M <- (log2(r1.4C$PM) - log2(r1.ct$PM))
res <- normalizeArrayData(r1.A, r1.M, name="r1_norm ",
plots=TRUE)
r1.A <- res$A; r1.M <- res$M
```

5.4 Tiling Array Probes Filtering

Identify probes associated with best and problematic 4C enrichment scores.

```
r1.EQ <- enrichmentQuality(r1.ct$PM, r1.4C$PM, plots=T)
# Replicate 1
r2.EQ <- enrichmentQuality(r2.ct$PM, r2.4C$PM, plots=T)
# Replicate 2
```

Make control plots of normalized 4C enrichment versus probe distance to restriction site. Compute the set of rejected probes based on enrichment scores, length of restriction fragments and relative positions to restriction sites (*see* **Note 3**).

```
# a) Reject tiny fragments OR no fragment match OR
excessive distance to RS
reject <- with(probes.grg, RF_LEN < 50 | is.na(RF_ID) |
RF_DIST>250)
# b) Reject tiny fragments OR no fragment match OR
excessive rank to RS
reject <- with(probes.grg, RF_LEN < 50 | is.na(RF_ID) |
RF_RANK>2)
# Additionally reject probes that appear "problematic"
in both replicates
reject <- reject | (r1.EQ$is.worst & r2.EQ$is.worst)
```

Update `probes.grg` and normalized A and M values to retain accepted probes only.

6 Visualization and Localization of 4C-Enriched Chromatin Domains

6.1 Computing 4C Enrichment Scores per Restriction Fragment

Calculate the average 4C enrichment for each half-fragment (5′ and 3′ ends) and the associated statistical scores used for multiresolution analysis in next sections (*see* **Note 4**). Pool replicates by combining these statistical scores using R.A. Fisher's combined probability method (*see* **Note 5**).

```
y1 <- combineByFragments(r1.M, probes.grg, FUN=mean) #
Replicate 1
y2 <- combineByFragments(r2.M, probes.grg, FUN=mean) #
Replicate 2
y1$SCORE <- enrichmentScore(y1$VALUE)
y2$SCORE <- enrichmentScore(y2$VALUE)
Yi <- y1$SCORE + y2$SCORE
Yi <- sapply(-2*Yi, pchisq, df=2*2, lower.tail=FALSE, log.
p=TRUE)
```

6.2 Domainogram Visualization

Compute genomic coordinates for domainogram visualization. Define the range of genomic locations and window sizes to be represented.

```
vis.coords <-
visualizationCoordinates(y1$SITE, y1$SITE)
xlim <- c(min(vis.coords$start), max(vis.coords$end)) #
Whole chromosome
wlim <- c(1, 2^16); w2y <- wSize2yAxis(wmax=wlim[2],
logscale=T)
```

Plot the domainogram (see documentation provided in the MRA.TA package).

```
domainogram(Yi, vis.coords$start,
vis.coords$end, w2y, xlim, wlim)
```

6.3 Multi-resolution Segmentation

Run the multi-resolution segmentation and hierarchical filtering algorithms (*see* **Note 6**).

```
sgm <- segmentation(
Yi,    name="4C_segmentation",    wmax=2^12,    wmin=5,
gamma=1E-4, Tw=-10
)
```

Load results. Visualize enriched segments and filtered domains.

```
opts <- read.delim(sgm$file.segments,
stringsAsFactors=F, skip=1)
doms <- read.delim(sgm$file.domains,
stringsAsFactors=F)
doms.mr <- read.delim(sgm$file.maxresolution,
stringsAsFactors=F)
doms.ms <- read.delim(sgm$file.maxscale,
stringsAsFactors=F)
# Visualize segmentation and hierarchical filtering results
plotOptimalSegments(opts, vis.coords$start,
vis.coords$end, w2y, xlim, wlim)
plotDomains(doms, vis.coords$start,
vis.coords$end, w2y, xlim, wlim)
```

Retrieve genomic coordinates of 4C-enriched chromatin domains at maximum resolution and maximal scale.

```
doms.mr$RS_START <- y1$SITE[doms.mr$start] # Max. resolu-
tion: start
doms.mr$RS_END <- y1$SITE[doms.mr$end] # Max. resolution:
end
doms.ms$RS_START <- y1$SITE[doms.ms$start] # Max. scale:
start
doms.ms$RS_END <- y1$SITE[doms.ms$end] # Max. scale:
end
```

7 Notes

1. Restriction enzymes employed in 4C protocols cut DNA on well-defined 4 or 6 bp sequence motifs, for instance GATC in the case of DpnII. But depending on the enzyme used, the precise location of the cutting point on the top and bottom

strands of DNA is not necessarily at the extremities of the sequence motif. This detail should not make any sensible difference with the data analysis methods we propose. However, the cut site location has to be defined precisely because the provided R functions do not assume a default and possibly incorrect location.

2. Depending on the distribution of raw *A* and *M* values, the background bias estimation with default parameter values may fail to execute or lead to unrealistic results. In that case, set the `plots` option to `TRUE` and try to slightly increase or decrease the following parameters: `smoothness` coefficient controlling the meanshift radius (Fig. 3a), `nsteps` imposing a minimum number of meanshift iterations, `epsilon` defining the convergence limit and corresponding to the minimum displacement required per meanshift iteration in the rescaled (*A*,*M*) plane. When observing the control graphs, reliable bias estimations converge rapidly to an approximately constant series of angles (Fig. 3c).
3. Accepting probes with attribute `RF_RANK` = 1 means accepting RS-nearest probes only (Fig. 4a). Practically, this selects tiling array probes that are as equivalent as possible as enzyme-specific array probes.
4. With high-resolution tiling arrays, when combining 4C enrichments from several probes for each extremity of the restriction fragments, the `FUN` parameter of `combineByFragments` may be used to apply alternative calculations (for instance a median or weighted mean of normalized *M* values).
5. The same reproducibility issue is observed in the three reanalyzed datasets. Strong 4C enrichments reveal two populations in probes targeting the most abundant 3C-ligation products: a population presenting a good correlation between replicate experiments, and a population showing high levels in one replicate but background level the other replicate. In this context, we use Fisher's combined probability method to combine 4C enrichment scores from paired replicates. We recommend considering alternative options when combining more than two replicates (for instance the average or median of *M* values).
6. The segmentation algorithm is controlled by three parameters: `wmin` and `Tw` determine the minimum window sizes and the enrichment score threshold for accepted segments, respectively. The third parameter `gamma` is a scoring penalty introduced with the original segmentation approach[28] which practically influences both the minimum size and enrichment score of accepted segments. A valid value for this parameter should be within $0 < \texttt{gamma} \leq 1$, the value of 1 corresponding to an absence of penalty. The value of the score threshold should be $\texttt{Tw} \leq 0$, a value of 0 corresponding to an absence of threshold.

Acknowledgements

We thank Elzo de Wit and Bas Tolhuis for their help sharing source code and biological data, and Tom Sexton for feedback and discussions. B.L. was supported by the Ministère de l'Enseignement Supérieur et de la Recherche and the European Research Council (ERC-2008-AdG No 232947), I.C. was supported by the Ministère de l'Enseignement Supérieur et de la Recherche and by the Ligue contre le Cancer. F.B. and G.C. are supported by the CNRS. G.C. research was supported by grants from the European Research Council (ERC-2008-AdG No 232947), the CNRS, the European Network of Excellence EpiGeneSys, the Agence Nationale de la Recherche, and the Association pour la Recherche sur le Cancer.

References

1. Dekker J, Rippe K, Dekker M et al ((2002) Capturing chromosome conformation. Science 295:1306–1311
2. de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26:11–24
3. Sexton T, Cavalli G (2015) The role of chromosome domains in shaping the functional genome. Cell 160:1049–1059
4. Marti-Renom MA, Mirny LA (2011) Bridging the resolution gap in structural modeling of 3D genome organization. PLoS Comput Biol 7:e1002125
5. Comet I, Savitskaya E, Schuettengruber B et al (2006) PRE-mediated bypass of two Su(Hw) insulators targets PcG proteins to a downstream promoter. Dev Cell 11:117–124
6. Lanzuolo C, Roure V, Dekker J et al (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol 9:1167–1174
7. Sexton T, Yaffe E, Kenigsberg E et al (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148(3):458–472
8. Rao SSP, Huntley MH, Durand NC et al (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159:1665–1680
9. Dixon JR, Selvaraj S, Yue F et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:1–5
10. Nora EP, Lajoie BR, Schulz EG et al (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485:381–385
11. Zhao Z, Tavoosidana G, Sjölinder M et al (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38:1341–1347
12. Göndör A, Rougier C, Ohlsson R (2008) High-resolution circular chromosome conformation capture assay. Nat Protoc 3:303–313
13. Simonis M, Klous P, Splinter E et al (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38:1348–1354
14. Schoenfelder S, Sexton T, Chakalova L et al (2009) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 42:53–61
15. Bantignies F, Roure V, Comet I et al (2011) Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell 144:214–226
16. Lomvardas S, Barnea G, Pisapia DJ et al (2006) Interchromosomal interactions and olfactory receptor choice. Cell 126:403–413
17. Ling JQ, Li T, Hu JF et al (2006) CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312:269–272
18. Sexton T, Kurukuti S, Mitchell JA et al (2012) Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. Nat Protoc 7:1335–1350
19. Tolhuis B, Blom M, van Lohuizen M (2012) Chromosome conformation capture on chip in single Drosophila melanogaster tissues. Methods 58:231–242

20. Lieberman-Aiden E, van Berkum NL, Williams L et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293
21. Yang YH, Dudoit S, Luu P et al (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30:e15
22. Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836
23. Palstra R-J, Simonis M, Klous P et al (2008) Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. PloS One 3:e1661
24. Tolhuis B, Blom M, Kerkhoven RM et al (2011) Interactions among Polycomb domains are guided by chromosome architecture. PLoS Genet 7:e1001343
25. Montavon T, Soshnikova N, Mascrez B et al (2011) A regulatory archipelago controls hox genes transcription in digits. Cell 147:1132–1145
26. Lonfat N, Montavon T, Jebb D et al (2013) Transgene- and locus-dependent imprinting reveals allele-specific chromosome conformations. Proc Natl Acad Sci U S A 110:11946–11951
27. Ong CT, Van Bortle K, Ramos E et al (2013) XPoly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in drosophila. Cell 155:148–159
28. de Wit E, Braunschweig U, Greil F et al (2008) Global chromatin domain organization of the Drosophila genome. PLoS Genet 4:e1000045
29. Noordermeer D, de Wit E, Klous P et al (2011) Variegated gene expression caused by cell-specific long-range DNA interactions. Nat Cell Biol 13:1–10
30. Yaffe E, Tanay A (2011) Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet 43:1059–1065
31. van de Werken HJG, Landan G, Holwerda SJB et al (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods 9:969–972
32. van de Werken HJG, de Vree PJP, Splinter E et al (2012) 4C technology: protocols and data analysis. Methods Enzymol 513:89–112
33. Splinter E, de Wit E, van de Werken HJG et al (2012) Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58:221–230
34. Thongjuea S, Stadhouders R, Grosveld FG et al (2013) R3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res 41:1–12
35. Walter C, Schuetzmann D, Rosenbauer F et al (2014) Basic4Cseq: an R/Bioconductor package for analyzing 4C-seq data. Bioinformatics 30:3268–3269
36. Williams RL, Starmer J, Mugford JW et al (2014) FourSig: a method for determining chromosomal interactions in 4C-Seq data. Nucleic Acids Res 42:1–16
37. Splinter E, de Wit E, Nora EP et al (2011) The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev 25:1371–1383
38. R Core Team (2015) R: a language and environment for statistical computing. http://www.r-project.org/
39. Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80
40. Langmead B, Trapnell C, Pop M et al (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
41. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210
42. Fukunaga K, Hostetler L (1975) The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inf Theory 21:32–40

Part V

Chapter 22

In Vivo Models to Address the Function of Polycomb Group Proteins

Frédéric Bantignies*

Abstract

Initially discovered as repressors of homeotic gene expression in *Drosophila*, Polycomb group (PcG) proteins have now been shown to be involved in a plethora of biological processes. Indeed, by repressing a large number of target genes, including specific lineage genes, these chromatin factors play major roles in a multitude of cellular functions, such as pluripotency, differentiation, reprogramming, tissue regeneration, and nuclear organization. In this book chapter are presented in vivo approaches and technologies, which have been used in both mammalian and *Drosophila* systems to study the cellular functions of Polycomb group proteins.

Key words Polycomb, Differentiation, Myogenesis, Reprogramming, Heterokaryons, Transdetermination, TALEN technology, Nuclear organization

Initially discovered as repressors of homeotic gene expression in *Drosophila*, the contribution of Polycomb group (PcG) proteins in various biological processes has literally exploded over the past decade. Indeed, by regulating a large number of target genes in the genome, including specific lineage genes, these chromatin factors play major roles in a multitude of biological phenomena, such as maintenance of pluripotency, proliferation, differentiation, and reprogramming.

In embryonic stem cells (ESC), PcG proteins repress nearly all genes involved in differentiation [1, 2], suggesting that PcG proteins are playing pivotal roles to control cell lineage commitment. Among the different differentiation processes governed by PcG proteins, myogenesis has been one of the best-characterized, most probably because of the possibility to culture myogenic cells in vitro from immortalized myoblast cell lines or from the isolation of muscle stem cells (MuSCs) [3, 4]. In this chapter, Mozzetta describes two methods that can be used to isolate and culture skeletal MuSCs from bulk muscles (using fluorescence activated cell sorting (FACS)) and from single myofibers. Through these

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_22,

techniques, the different stages of myogenesis can be recapitulated and studied in vitro.

PcG proteins seem also to have an important role in the maintenance of pluripotent ESC as well as in their establishment. Indeed, the activities of PcG proteins seem required for pluripotent reprogramming and knowledge in this area are particularly important for cancer research and regenerative medicine. The involvement of PcG proteins in pluripotent reprogramming was first demonstrated using an original system, which relies on the capacity of mouse ESC to directly reprogram differentiated human somatic cells towards pluripotency by the formation of heterokaryons [5]. Here, Malinowski and Fisher describe polyethylene glycol (PEG)-mediated cell fusion protocol for the generation and analysis of interspecies heterokaryons (between human B lymphocytes and mouse ES cells) and intraspecies hybrids (between mouse B lymphocytes and mouse ES cells). These methods represent alternative approaches to other reprogramming strategies, such as induction of pluripotent stem cells (iPS) and somatic cell nuclear transfer.

Another aspect of reprogramming in which PcG proteins are involved occurs during tissue regeneration, which can provide many new insights to understand the processes of wound healing. In *Drosophila*, transplantation of larval imaginal disks following disk dissection and fragmentation is an attractive in vivo model to study tissue regeneration and reprogramming phenomena like transdetermination, where cellular identities can occasionally switch fates. During the regeneration process, PcG-mediated gene silencing was found down-regulated, which may lead to increase cellular plasticity and occasional changes in cell fates [6, 7]. Imaginal disks can be easily manipulated, and Katsuyama and Paro describe how to transplant them into a recipient organism (larval or adult abdomen) and recover the implants to analyze regenerative properties. Moreover, the disk transplantation technique can be applied to the study of tumor formation properties and mechanisms in *Drosophila* [8, 9].

In the last few years, genomic editing has become increasingly popular and more easily accessible with the recent developments of the artificial transcription activator-like effector nucleases (TALENs) system or the clustered regularly interspaced short palindromic repeats (CRISPRs) technology. These technologies provide invaluable tools for targeted genome editing in many species and allow functional genomic applications. Akmammedov and colleagues describe a rapid TALEN assembly protocol called easyT together with a strategy for genome engineering via homologous recombination, which allow deleting specific regulatory sequences into the *Drosophila* genome [10]. This can have multitude applications, including here the dissection of molecular mechanisms of tissue regeneration and transdetermination.

Finally, in both *Drosophila* and mammals, PcG proteins form microscopically visible structures called "Polycomb bodies". These structures are known to represent the sites of PcG target gene silencing and clustering in the nucleus [11]. In order to address the dynamic of these nuclear compartments during development, Cheutin and Cavalli present an in vivo method to follow the motion of these structures in living *Drosophila* embryos containing fused GFP-PcG proteins. Importantly, this method can be easily applied to observe the movement of any fluorescently labeled proteins forming microscopically visible structures in vivo.

References

1. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA et al (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441:349–353
2. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS et al (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125:301–313
3. Palacios D, Mozzetta C, Consalvi S, Caretti G, Saccone V et al (2010) TNF/p38alpha/polycomb signaling to Pax7 locus in satellite cells links inflammation to the epigenetic control of muscle regeneration. Cell Stem Cell 7:455–469
4. Stojic L, Jasencakova Z, Prezioso C, Stutzer A, Bodega B et al (2011) Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. Epigenetics Chromatin 4:16
5. Pereira CF, Piccolo FM, Tsubouchi T, Sauer S, Ryan NK et al (2010) ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. Cell Stem Cell 6:547–556
6. Klebes A, Sustar A, Kechris K, Li H, Schubiger G et al (2005) Regulation of cellular plasticity in Drosophila imaginal disc cells by the Polycomb group, trithorax group and lama genes. Development 132:3753–3765
7. Lee N, Maurange C, Ringrose L, Paro R (2005) Suppression of Polycomb group proteins by JNK signalling induces transdetermination in Drosophila imaginal discs. Nature 438:234–237
8. Martinez AM, Schuettengruber B, Sakr S, Janic A, Gonzalez C et al (2009) Polyhomeotic has a tumor suppressor activity mediated by repression of Notch signaling. Nat Genet 41:1076–1082
9. Sievers C, Comoglio F, Seimiya M, Merdes G, Paro R (2014) A deterministic analysis of genome integrity during neoplastic growth in Drosophila. PLoS One 9:e87090
10. Katsuyama T, Akmammedov A, Seimiya M, Hess SC, Sievers C et al (2013) An efficient strategy for TALEN-mediated genome engineering in Drosophila. Nucleic Acids Res 41:e163
11. Bantignies F, Cavalli G (2011) Polycomb group proteins: repression in 3D. Trends Genet 27:454–464

Chapter 23

A Rapid TALEN Assembly Protocol

Arslan Akmammedov*, Tomonori Katsuyama, and Renato Paro

Abstract

Owing to their modular and highly specific DNA recognition mode, transcription activator-like effector nucleases (TALENs) have been rapidly adopted by the scientific community for the purpose of generating site-specific double-strand breaks (DSBs) on a DNA molecule. A pair of TALENs can be used to produce random insertions or deletions of various lengths via nonhomologous end-joining or together with a homologous donor DNA to induce precise sequence alterations by homologous recombination (HR). Here, we describe a method for TALEN assembly (easyT) and a strategy for genome engineering via HR.

Key words TALEN, Genome engineering, Site-specific gene integration

1 Introduction

TALENs are fusion proteins containing a transcription activator-like effector (TALE) domain and a FokI nuclease domain [1]. FokI is a nonspecific nuclease that can generate DSBs in a DNA molecule upon homodimerization. The need for two FokI domains to dimerize requires that TALENs be used as a pair (left and right), binding two sequences separated by 14–32 bp and in a nuclease-to-nuclease orientation. Our TALE domain is a truncated derivative of the AvrXa7-FN TALEN [2] that retains 207 N-terminal amino acids and 63 C-terminal amino acids. TALE repeats within the TALE domain are responsible for the recognition of a specific DNA sequence—one repeat recognizing a single nucleotide. Repeats within the repeat domain contain 34 amino acids (the last "half-repeat" contains 17 amino acids and can also recognize a single nucleotide) that differ only in amino acids at positions 12 and 13. These two amino acids, also known as repeat variable diresidue (RVD), confer the sequence preference to each repeat by following the simple RVD code: diresidue HD recognizes cytosine, NI recognizes adenine, NG recognizes thymine, and NK recognizes guanine [3, 4]. By following the RVD code, a pair of TALENs can be designed to recognize a target DNA sequence and introduce a DSB at that site.

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_23,

In our lab, we are interested in studying downregulation of silencing established by Polycomb group (PcG) proteins observed in regenerating *Drosophila* imaginal discs following disc fragmentation [5]. For regeneration of a damaged imaginal disc, cells at the site of injury need to activate specific genes that are normally under PcG-mediated silencing. It is thought that Jun N-terminal kinase (JNK) signaling pathway is responsible for specific PcG target gene reactivation in these cells [6]. At the molecular lever, however, it is still not clear how this reactivation is accomplished. One attractive hypothesis is that transcription factor (TF) AP-1, which acts downstream of JNK, binds to the regulatory elements of these specific PcG target genes and either modulates PcG function or reactivates gene expression regardless of PcG silencing.

Thus, in order to investigate whether AP-1-binding sites found within the regulatory DNA regions of *unpaired* (*upd*) and *wingless* (*wg*) genes, two PcG target genes, are necessary for their reactivation during imaginal disc regeneration, we synthesized two TALEN pairs with the easyT assembly method and used them together with a homologous donor DNA to delete two AP-1-binding sites and insert a selectable marker gene to identify these mutants among F1 adults [7]. We used green fluorescent protein under the control of 3xP3 promoter to identify mutants by fluorescence in the eye.

2 Materials

2.1 easyT Unit Library Preparation

1. easyT kit (22 plasmids): Temp_type-a-NI, Temp_type-a-NG, Temp_type-a-NK,Temp_type-a-HD,Temp_type-b-NI,Temp_type-b-NG, Temp_type-b-NK, Temp_type-b-HD, Temp_type-c-NI, Temp_type-c-NG, Temp_type-c-NK, Temp_type-c-HD, Temp_type-d-NI, Temp_type-d-NG, Temp_type-d-NK, Temp_type-d-HD, Temp_unit1(a′)-NI, Temp_unit1(a′)-NG, Temp_unit1(a′)-NK, Temp_unit1(a′)-HD, Temp_tail unit, TALEN backbone plasmid. The easyT kit is available from our lab upon request.
2. Phusion® High-Fidelity DNA Polymerase.
3. QIAquick PCR Purification Kit.
4. Agarose.
5. Tris-borate-EDTA buffer: 89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3.

2.2 TALEN Synthesis with easyT

1. Restriction enzymes: BaeI.
2. T4 DNA ligase.
3. QIAquick Gel Extraction Kit.
4. DNA Clean & Concentrator™-5.
5. *Taq* DNA Polymerase with ThermoPol® Buffer.

6. GenElute™ Plasmid Miniprep Kit.
7. DH5α competent *E. coli*.

2.3 Drosophila Embryo Microinjection

1. QIAGEN Plasmid Midi Kit.
2. Fly lines: w^{1118} and w^{1118} $lig4^{169}$.
3. UV microscope.
4. QIAamp DNA Micro Kit.

3 Methods

The method described in this chapter can be divided into three parts: easyT unit library preparation, TALEN synthesis with easyT, and homologous donor DNA cloning and *Drosophila* embryo microinjection.

3.1 easyT Unit Library Preparation and TALEN Backbone Plasmid Pre-digestion

1. Use Phusion® High-Fidelity DNA Polymerase and set up 48 PCR reactions following the manufacturer's recommendations. Use plasmids from easyT kit as a template for PCR reactions. Consult Tables 1 and 2 for plasmid template and primer pair information. Use 5 ng of plasmid template per each 50 μL PCR reaction.
2. Run 5 μL of each PCR reaction on 0.8% agarose gel to confirm a single band of the correct size in each reaction (*see* **Note 1**). Amplified unit sizes: unit 1—137 bp, unit 2, 6, 10, 14, 18—158 bp, unit 3, 7, 11, 15, 19—157 bp, unit 4—142 bp, unit 5—148 bp, unit 8—145 bp, unit 9—144 bp, unit 12—146 bp, unit 13—143 bp, unit 16—148 bp, unit 17—141 bp, tail unit 17—136 bp, tail unit 18—143 bp, tail unit 19—139 bp, tail unit 16 and 20—136 bp.
3. Purify DNA from PCR reactions using QIAquick PCR Purification Kit and dilute DNA to 5 ng/μL. Label tubes with nucleotide recognized by the RVD and unit number. For example, unit 1 generated from template Temp_uit1(a′)-NI is labeled as A1 (tail unit is labeled as tail 16, tail 17, etc.). Units are ready to be used for TALEN assembly.
4. Digest TALEN backbone plasmid (easyT kit) with BaeI. Use 4 μL of 10× T4 DNA ligase buffer, 0.4 μL of 100× BSA, 2.5 μL of 0.32 mM *S*-adenosylmethionine (SAM), 4 μL of BaeI, and 1 μg of TALEN backbone plasmid, and dilute with sterile water to 40 μL. Incubate at 25 °C for 2 h.
5. Run digestion reaction on 0.8% agarose gel. Excise the 5.2 kb band and extract DNA using QIAquick Gel Extraction Kit. Pre-digested TALEN backbone plasmid is ready to be used for TALEN assembly (*see* **Note 2**).

Table 1
Primers for easyT TALEN assembly method and genomic DNA PCR

Primers for easyT unit library	
TALunit1-F(MluI)	AAAAGTACCAATATGATTCACGCGTGGCGCAATGCACTGA
TALagcct-R	GATAGGTACTAAGGTAGCCTGTACTAGGCTTGCTTGCCGCC
TALagcct-F	CTTGACTTTAGTACCTAAGCACAGCCTTGGAGACGGTGCAGC
TALcttga-R	GATAGGTACTAAGGTAGCCTGTACTTCAAGAGCTTGCTTGCCGCC
TALcttga-F	CTTGACTTTAGTACCTAAGCACCTTGAGACGGTACAGCGG
TALgaaac-R	GATAGGTACTAAGGTAGCCTGTACTGTTTCCAGCGCTTGCTTGCCGCC
TALgaaac-F	CTTGACTTTAGTACCTAAGCACGAAACGGTACAGCGGCTG
TALunit4-R(HindIII)	TACGGGGGTATAGTGTACTAAGCTTGCTTGCCGCC
TALunit5-F(HindIII)	AAAAGTACCAATATGAAGCTTTGGAGACAGTACAGCGG
TALunit8-R(XbaI)	TACGGGGGTATAGTGTACTTCTAGAGCTTGCTTGCCGCC
TALunit9-F(XbaI)	AAAAGTACCAATATGTCTAGAGACGGTACAGCGG
TALunit12-R(XhoI)	TACGGGGGTATAGTGTACTCTCGAGCGCTTGCTTGCCGCC
TALunit13-F(XhoI)	AAAAGTACCAATATGCTCGAGACGGTACAGCGG
TALunit16-R(BsrGI)	TACGGGGGTATAGTGTACAGTTTCCAGCGCTTGCTTGCCGCC
TALunit17-F(BsrGI)	AAAAGTACCAATATGAGAAACTGTACAGCGGCTGTTG
TALtail17-F(BsrGI)	AAAAACAAAAGTACCAATATGAGAAACTGTACATGCCCAGTTATCTCGCCCTGA

TALtail18agcct-F	CTTGACTTTAGTACCTAAGCACAGCCTTGGAGACGATTGTTGCCCAGTTATC
TALtail19cttga-F	CTTGACTTTAGTACCTAAGCACCTTGAGACGATTGTTGCCCAGTTATC
TALtail20gaaac-F	CTTGACTTTAGTACCTAAGCACGAAACGATTGTTGCCCAGTTATC
TALtail-R(AatII)	TACGGGGGTATAGTGTACTGGACGTCCGCCGAGG
Primers for 4-mer amplification	
TAL4mer-F	AAAAACAAAAGTACCAATATG
TAL4mer-R	GGGGGGTACGGGGGTATAGTGTAC
Primers for colony PCR and sequencing	
TALrepeat-F	GTTACAGTTGGACACAGG
TALrepeat-R	ACGTGCGTTCGGGAATG
TALrepeat-F2	ACTGACGGGTGCCCCCCT
TALrepeat-R2	GGGCGAGATAACTGGGC
Primers for genomic PCR of modified wg locus	
wg-F(outside)	CAACCAAATAGTACTCCTCCCTTTCTCCTG
wg-R(on EGFP)	GTTCTTCTGCTTGTCGGCCATGATATAG

Table 2
Templates and primer pairs for easyT unit library

	NI (A)		NG (T)		NK (G)		HD (C)	
Unit	**Template**	**Primer pair**	**Template**	**Primer pair**	**Template**	**Primer pair**	**Template**	**Primer pair**
Unit 1	Temp_unit1(a′)-NI	TALunit1-F(MluI) TALagcct-R	Temp_unit1(a′)-NG	TALunit1-F(MluI) TALagcct-R	Temp_unit1(a′)-NK	TALunit1-F(MluI) TALagcct-R	Temp_unit1(a′)-HD	TALunit1-F(MluI) TALagcct-R
Unit 2, 6, 10, 14, 18	Temp_type-b-NI	TALagcct-F TALcttga-R	Temp_type-b-NG	TALagcct-F TALcttga-R	Temp_type-b-NK	TALagcct-F TALcttga-R	Temp_type-b-HD	TALagcct-F TALcttga-R
Unit 3, 7, 11, 15, 19	Temp_type-c-NI	TALcttga-F TALgaaac-R	Temp_type-c-NG	TALcttga-F TALgaaac-R	Temp_type-c-NK	TALcttga-F TALgaaac-R	Temp_type-c-HD	TALcttga-F TALgaaac-R
Unit 4	Temp_type-d-NI	TALgaaac-F TALunit4-R(HindIII)	Temp_type-d-NG	TALgaaac-F TALunit4-R(HindIII)	Temp_type-d-NK	TALgaaac-F TALunit4-R(HindIII)	Temp_type-d-HD	TALgaaac-F TALunit4-R(HindIII)
Unit 5	Temp_type-a-NI	TALunit5-F(HindIII) TALagcct-R	Temp_type-a-NG	TALunit5-F(HindIII) TALagcct-R	Temp_type-a-NK	TALunit5-F(HindIII) TALagcct-R	Temp_type-a-HD	TALunit5-F(HindIII) TALagcct-R
Unit 8	Temp_type-d-NI	TALgaaac-F TALunit8-R(XbaI)	Temp_type-d-NG	TALgaaac-F TALunit8-R(XbaI)	Temp_type-d-NK	TALgaaac-F TALunit8-R(XbaI)	Temp_type-d-HD	TALgaaac-F TALunit8-R(XbaI)
Unit 9	Temp_type-a-NI	TALunit9-F(XbaI) TALagcct-R	Temp_type-a-NG	TALunit9-F(XbaI) TALagcct-R	Temp_type-a-NK	TALunit9-F(XbaI) TALagcct-R	Temp_type-a-HD	TALunit9-F(XbaI) TALagcct-R
Unit 12	Temp_type-d-NI	TALgaaac-F TALunit12-R(XhoI)	Temp_type-d-NG	TALgaaac-F TALunit12-R(XhoI)	Temp_type-d-NK	TALgaaac-F TALunit12-R(XhoI)	Temp_type-d-HD	TALgaaac-F TALunit12-R(XhoI)

Unit 13	Temp_type-a-NI	TALunit13-F(XhoI)	Temp_type-a-NG	TALunit13-F(XhoI)	Temp_type-a-NK	TALunit13-F(XhoI)	Temp_type-a-HD	TALunit13-F(XhoI)
		TALagcct-R		TALagcct-R		TALagcct-R		TALagcct-R
Unit 16	Temp_type-d-NI	TALgaaac-F	Temp_type-d-NG	TALgaaac-F	Temp_type-d-NK	TALgaaac-F	Temp_type-d-HD	TALgaaac-F
		TALunit16-R(BsrGI)		TALunit16-R(BsrGI)		TALunit16-R(BsrGI)		TALunit16-R(BsrGI)
Unit 17	Temp_type-a-NI	TALunit17-F(BsrGI)	Temp_type-a-NG	TALunit17-F(BsrGI)	Temp_type-a-NK	TALunit17-F(BsrGI)	Temp_type-a-HD	TALunit17-F(BsrGI)
		TALagcct-R		TALagcct-R		TALagcct-R		TALagcct-R
Tail unit	Template	Primer pair						
Tail_unit17	Temp_tail unit	TALtail17-F(BsrGI)						
		TALtail-R(AatII)						
Tail_unit18	Temp_tail unit	TALtail18agcct-F						
		TALtail-R(AatII)						
Tail_unit19	Temp_tail unit	TALtail19cttga-F						
		TALtail-R(AatII)						
Tail_unit20 (Tail_unit16)	Temp_tail unit	TALtail20gaaac-F						
		TALtail-R(AatII)						

3.2 TALEN Synthesis with easyT

Here we use synthesis of a TALEN pair targeting the *wg* locus as an example to describe TALEN assembly (*see* **Note 3**). These TALENs are 20-mers, i.e., containing 20 units, and recognize 19 bp sequences (tail unit is not involved in DNA recognition). *See* Fig. 1 for a schematic overview of easyT TALEN assembly.

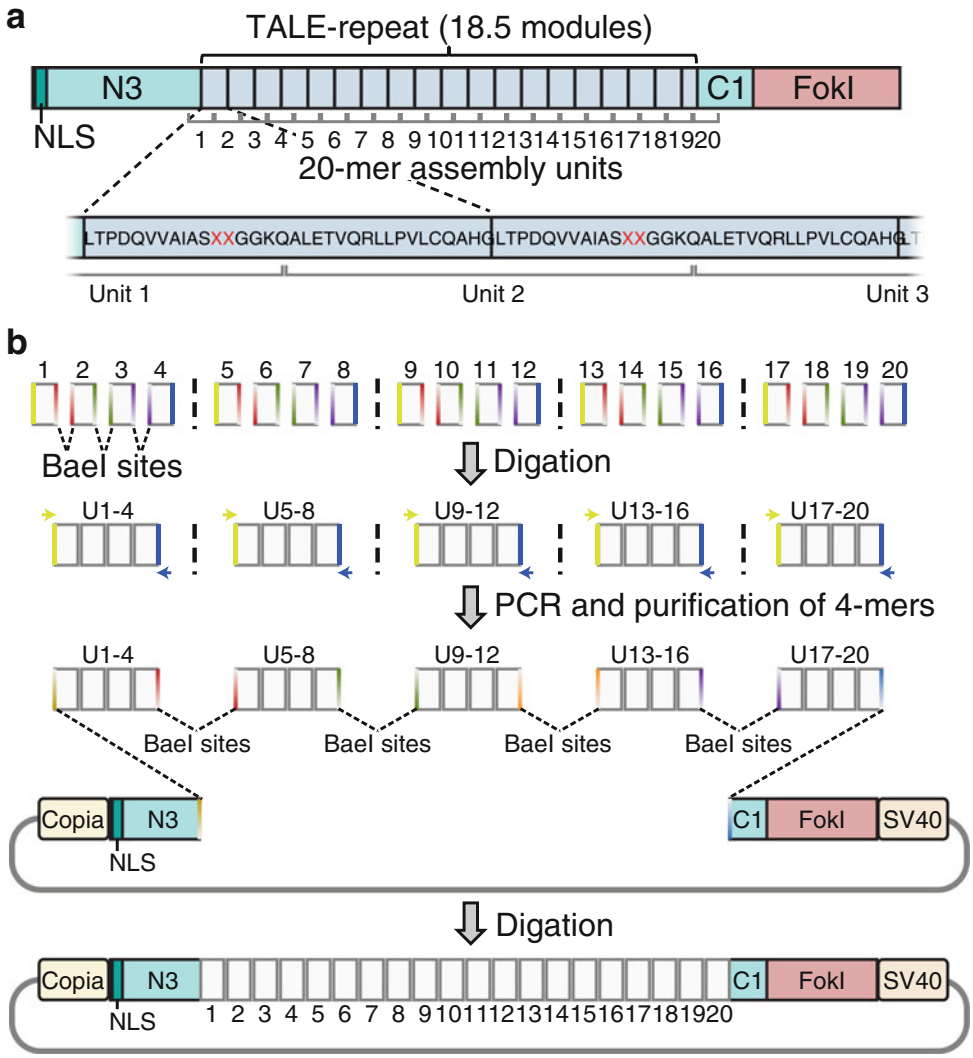


Fig. 1 Construction of TALENs with the easyT protocol. (**a**) A schematic representation of a TALEN with a TALE-repeat length of 18.5 modules. The TALE-repeat is assembled from 20 monomer units. The boundaries of monomer units were shifted from those of the TALE-repeat modules. (**b**) Overview of TALEN cloning. In the first step, four units are assembled into 4-mers in a "digation" reaction. In the second step, 4-mers are PCR-amplified, run on an agarose gel, gel-extracted, and concentrated. Finally, 4-mers were assembled into the TALEN backbone plasmid in the second digation reaction. *Yellow* and *blue arrows* indicate primers used for 4-mer amplification. Reproduced from Katsuyama, T. et al., An efficient strategy for TALEN-mediated genome engineering in Drosophila, Nucleic Acids Research, 2013, 41, 17, e163-e163, by permission of Oxford University Press

1. Use the following guidelines to select TALEN target site: TF-binding site to be deleted should be positioned in the spacer region between left and right TALEN-binding sequences, TALEN-binding sequences should be preceded by the nucleotide thymine, and the spacer region between left and right TALEN should be about 20 bp. According to these guidelines, the sequence 5′-tcatctgatgcttcacagaatcagtagc**tgactca**ctcc-gattcagtttcaggaattca-3′ was selected for the deletion of a putative AP-1-binding site within the *wg* locus (underlined—TALEN-binding sequences, in bold—the putative AP-1-binding site).
2. Setting up ten digation reactions. A digation reaction contains both BaeI restriction enzyme and T4 DNA ligase, and hence both DNA digestion and ligation are taking place. Thus, if digation reaction contains units 1–4, at the end of digation the reaction will contain monomers of units 1–4; 2-mers of units 1–2, 2–3, and 3–4; 3-mers of units 1–2–3 and 2–3–4; and 4-mers of units 1–2–3–4. Take ten PCR tubes (five per each TALEN) and add the following in each tube: 2 μL of 10× T4 DNA ligase buffer, 0.2 μL of 100× BSA, 1.25 μL of 0.32 mM SAM, 2 μL of 2.5 mM dNTPs, 2 μL of BaeI, 1 μL of T4 DNA ligase, and 7.55 μL of sterile water. Label PCR tubes L1 through L5 for left TALEN and R1 through R5 for right TALEN. In tube L1 add 1 μL of units C1, A2, T3, and C4; in tube L2 add 1 μL of units T5, G2, A3, and T8; in tube L3 add 1 μL of units G9, C2, T3, and T12; in tube L4 add 1 μL of units C13, A2, C3, and A16; and in tube L5 add 1 μL of units G17, A2, A3, and tail 20. Correspondingly, in tube R1 add 1 μL of units G1, A2, A3, and T4; in tube R2 add 1 μL of units T5, C2, C3, and T8; in tube R3 add 1 μL of units G9, A2, A3, and A12; in tube R4 add 1 μL of units C13, T2, G3, and A16; and in tube R5 add 1 μL of units A17, T2, C3, and tail 20. Mix by pipetting up and down several times and incubate at 25 °C for 1 h.
3. Purify DNA from digation reaction using DNA Clean & Concentrator™-5 kit and elute in 6 μL of sterile water.
4. Set up ten 50 μL PCR reactions using Phusion® High-Fidelity DNA Polymerase and primers TAL4mer-F and TAL4mer-R (Table 1). Use 1 μL of eluate from previous step to amplify 4-mers L1, L2, L3, L4, L5, R1, R2, R3, R4, and R5.

PCR program:					
98 °C	<98 °C	55 °C	72 °C>	72 °C	4 °C
3 min	10 s	10 s	15 s	7 min	∞
<32 cycles>					

5. Run PCR reactions on 2.0% agarose gel until 4-mers are separated from 3-mers. Cut the bands corresponding to 4-mers and purify DNA using QIAquick Gel Extraction Kit (*see* **Note 4**).
6. Check concentration of 4-mers. Use DNA Clean & Concentrator™-5 kit to concentrate DNA if concentration of 4-mers is below 15 ng/μL.
7. Setting up second digation reaction: In the second digation reaction, purified 4-mers from the previous step are digated into pre-digested TALEN backbone plasmid prepared in Subheading 3.1, **step 5**. Take two PCR tubes and label them R and L for left and right TALEN. In each tube add 2 μL of 10× T4 DNA ligase buffer, 0.2 μL of 100× BSA, 1.25 μL of 0.32 mM SAM, 2 μL of 2.5 mM dNTPs, 2 μL of BaeI, 1 μL of T4 DNA ligase, and 30 ng of pre-digested TALEN backbone plasmid. In tube L add 30 ng of each of the 4-mers L1, L2, L3, L4, and L5 from Subheading 3.2, **step 5**. In tube R add 30 ng of each of the 4-mers R1, R2, R3, R4, and R5 from Subheading 3.2, **step 5**. Mix by pipetting up and down several times and incubate at 25 °C for 1 h.
8. Use 10 μL of second digation reactions to transform competent *E. coli* cells. We use DH5α-competent *E. coli*. Spread transformed *E. coli* cells on agar plates selecting for ampicillin resistance.
9. Identify colonies containing full-length TALENs by colony PCR. We recommend picking at least ten individual colonies per TALEN. Set up 20 PCR reactions. In each PCR tube add 2.5 μL of 10× ThermoPol buffer, 2 μL of 2.5 mM dNTPs, 1 μL of 10 mM primer TALrepeat-F, 1 μL of 10 mM primer TALrepeat-R, 0.1 μL of Taq DNA polymerase, 8.4 μL of sterile water, and a single *E. coli* colony dissolved in 10 μL sterile water.

PCR program:					
94 °C	<94 °C	55 °C	72 °C>	72 °C	4 °C
5 min	30 s	30 s	3 min	7 min	∞
<25 cycles>					

10. Run completed colony PCR reactions on 0.8% agarose gel. For ease of identification of positive colonies, we recommend running on a gel a PCR reaction with a previously confirmed TALEN (control) as a template. Figure 2 is an example of a gel picture with 1 kb DNA ladder, control TALEN PCR reaction, and eight colony PCR reactions (lanes 1–10, respectively). The band pattern of a control 20-mer TALEN shows a characteristic five-band pattern, with bands corresponding to 4-mer, 8-mer, 12-mer, 16-mer, and 20-mer. Colonies with this five-band pattern are positive colonies containing full-length TALENs. Culture positive colonies overnight in LB medium supplemented with ampicillin.

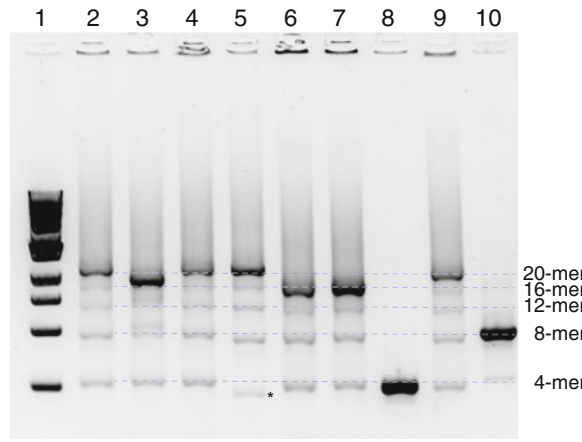


Fig. 2 Identifying positive clones by the pattern of PCR bands. The amplicons of colony PCR were run on a 0.8 % agarose gel. 1 kb DNA ladder (NEB) was loaded on lane 1. Positive clone having full 20-mer units (lane 2) shows a strong 20-mer band around 2 kb and additional four weak bands at the size of 4-, 8-, 12-, and 16-mer. Based on the pattern of these bands, we can estimate the number of 4-mers and identify possible positive clones: lanes 4 and 9. While lane 5 shows a strong band at 20-mer size, it is a negative because the size of the lowermost band (*) is obviously different. Reproduced from Katsuyama, T. et al., An efficient strategy for TALEN-mediated genome engineering in Drosophila, Nucleic Acids Research, 2013, 41, 17, e163-e163, by permission of Oxford University Press

11. Purify plasmid DNA from overnight cultures of positive colonies using GenElute™ Plasmid Miniprep Kit. Sequence plasmids using primers TALrepeat-F and TALrepeat-R and discard clones containing PCR-induced mutations (*see* **Note 5**).

3.3 Homologous Donor DNA and Drosophila *Embryo Microinjection*

1. 2 kb of genomic DNA sequences upstream and downstream of the putative AP-1-binding site targeted for deletion were amplified and cloned upstream and downstream of 3xP3-EGFP expression cassette. Thus, a selectable marker gene replaces AP-1-binding site on the donor DNA.
2. Next, donor DNA plasmid and two TALEN plasmids were midi-preped with QIAGEN Plasmid Midi Kit and injected into embryos of w^{1118} $lig4^{169}$ strain following standard embryo microinjection protocol (*see* **Note 6**). As a starting point, we recommend injecting 50 ng/μL per each TALEN plasmid and 100 ng/μL for donor DNA plasmid (*see* **Note 7**).
3. The adults of microinjected embryos were crossed to w^{1118}. Progeny of this cross was scored via green fluorescence in the eye for mutants that have integrated donor DNA.
4. Isolate genomic DNA of mutant flies using QIAamp DNA Micro Kit.

5. PCR reaction on genomic DNA of mutant flies using primers aligning on EGFP and outside of homology arms can be performed to check for a band of expected size and to confirm targeted integration (*see* **Note 8**).

4 Notes

1. If more than one band appear on the gel, use QIAquick Gel Extraction Kit to excise the band of the correct size and proceed to Subheading 3.1, **step 3**.
2. In our experience, complete digestion and clean extraction of the TALEN backbone plasmid are crucial to increase frequency of full-length TALENs in Subheading 3.2, **step 10**. To test whether TALEN backbone plasmid is completely digested, a ligation reaction containing only pre-digested TALEN backbone plasmid can be carried out. Completely digested and properly extracted TALEN backbone plasmid would not form self-ligation products and would not produce ampicillin-resistant *E. coli* colonies after transformation.
3. Our computation analysis indicated that sequences as short as 13 bp could be sufficient to specify uniquely a position within non-repetitive euchromatic regions of the *Drosophila* genome [7]. We also tested TALEN pairs recognizing 15 and 19 bp sequences and found no evidence that longer sequences result in higher frequency of targeted genome modification. Additionally, synthesis of TALENs recognizing 15 bp sequences is more efficient than synthesis of TALENs recognizing 19 bp sequences. Thus, if there are no additional restrictions/considerations on TALEN size or recognition sequence length, we recommend synthesizing TALENs recognizing 15 bp target sequences (i.e., 16-mers).
4. It is important to separate 4-mers from 3-mers on the gel. Presence of 3-mers in the second digation reaction (Subheading 3.2, **step 7**) would reduce the frequency of full-length TALENs in Subheading 3.2, **step 10**.
5. The sequencing company that performed TALEN sequencing for us can generate sequence read lengths of up to 1100 bp. Thus, 20-mer TALEN is the largest TALEN size we can generate and sequence through all units using forward and reverse primers aligning on TALEN backbone plasmid (outside of unit sequences). If the average sequence read length produced by your sequencing service provider is shorter than 1100 bp, consider synthesizing TALENs of smaller size (i.e., 16-mer, 17-mer).
6. We used homozygous lig4^{169} mutant flies as recipients of microinjections. *w*^{1118} *lig4*^{169} strain had been previously shown to increase donor DNA integration via HR [8].

7. We believe that optimal concentration of TALEN plasmids and donor DNA plasmid is a function of affinity of TALEN pair to their target site, TALEN toxicity due to off-target binding, toxicity associated with genomic regions on the donor DNA, and toxicity due to high concentration of plasmid DNA. As a result, optimal concentration of TALEN plasmids and donor DNA plasmid has to be determined empirically for each new experimental setup.
8. We perform genomic DNA PCR with primers aligning within EGFP and outside of homology arm sequences present on donor DNA to confirm targeted integration of 3xP3-EGFP and deletion of AP-1-binding site. In our experience, 2 kb regions can be easily amplified from genomic DNA. Consequently, the length of the homology arm sequences is constrained by the ability to amplify this region from genomic DNA.

Acknowledgments

We thank Makiko Seimiya for help with embryo microinjections. Work of T.K. and R.P. was supported by a grant from the Deutsche Forschungsgemeinschaft SPP1356—"Pluripotency and Cellular Reprogramming." Work in the R.P. laboratory is supported by ETH Zürich.

References

1. Mussolino C, Cathomen T (2012) TALE nucleases: tailored genome engineering made easy. Curr Opin Biotechnol 23(5):644–650. doi:10.1016/j.copbio.2012.01.013
2. Li T, Huang S, Zhao X, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res 39(14):6315–6325. doi:10.1093/nar/gkr188
3. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326(5959):1509–1512. doi:10.1126/science.1178811
4. Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326(5959):1501. doi:10.1126/science.1178817
5. Lee N, Maurange C, Ringrose L, Paro R (2005) Suppression of Polycomb group proteins by JNK signalling induces transdetermination in Drosophila imaginal discs. Nature 438(7065):234–237. doi:10.1038/nature04120
6. Katsuyama T, Paro R (2011) Epigenetic reprogramming during tissue regeneration. FEBS Lett 585(11):1617–1624. doi:10.1016/j.febslet.2011.05.010
7. Katsuyama T, Akmammedov A, Seimiya M, Hess SC, Sievers C, Paro R (2013) An efficient strategy for TALEN-mediated genome engineering in Drosophila. Nucleic Acids Res 41(17):e163. doi:10.1093/nar/gkt638
8. McVey M, Radut D, Sekelsky JJ (2004) End-joining repair of double-strand breaks in Drosophila melanogaster is largely DNA ligase IV independent. Genetics 168(4):2067–2076. doi:10.1534/genetics.104.033902

Chapter 24

Following the Motion of Polycomb Bodies in Living Drosophila Embryos

Thierry Cheutin* **and Giacomo Cavalli**

Abstract

During the last two decades, observation of cell nuclei in live microscopy evidences motion of nuclear compartments. Drosophila embryos constitute a good model to study nuclear dynamic during cell differentiation because they can easily be observed in live microscopy. Inside the cell nucleus, Polycomb group proteins accumulate in foci named Pc bodies. Here, we describe a method to visualize and analyze the motion of these nuclear compartments inside cell nuclei of Drosophila embryos.

Key words Polycomb bodies, Live microscopy, Drosophila embryo, Nucleus dynamics

1 Introduction

Before the discovery of green fluorescence protein (GFP), description of nuclear organization mainly relies on "in situ" observations made in electronic and optical microscopies. The main problem of sample preparation was to preserve cellular structure before observation. In optical microscopy, this problem is solved because GFP allows direct observation of living cells without any fixation. Moreover, live imaging shows that nuclear organization is much less static than expected since nuclear compartments and chromatin loci are mobile inside cell nuclei [1, 2]. Drosophila embryos represent a good cellular model to study the mobility of nuclear compartments during cell differentiation [3]. In transgenic fly lines, Polycomb-GFP (Pc-GFP) [4] or Polyhomeotic-GFP (Ph-GFP) [5] localizes in Pc bodies, similarly to endogenous proteins [6, 7]. How we can monitor and characterize the mobility of Pc bodies at several embryonic developmental stages will be presented in this protocol. This protocol is divided into three parts: preparation of Drosophila embryos for live microscopy, image acquisition, and analysis of time-lapse experiments to calculate mean square displacement (M.S.D.) and mean square change (M.S.C.).

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_24,

2 Materials

1. Grape juice agar plate.
2. Oil voltalef 3S (CTL Scientific Supply. Item# 24626.185).
3. FluoroDish (World Precision Instruments. Item#: FD35-100).
4. Confocal microscope.
5. Computer running Volocity (Improvision) and/or Imaris (Bitplane).

3 Methods

3.1 Sample Preparation

In order to observe Drosophila embryos in live microscopy, we need to collect and prepare them with minimal perturbations.

1. Flies lay eggs on grape juice agar plate during 2–14 h (adjust the duration of eggs collection to the developmental stage that you want to observe).
2. Transfer embryos on tape with a pencil and then manually dechorionate them with forceps.
3. Put 3–5 embryos on a slide next to a cover slip, add a drop of oil voltalef 3S, and cover them with a second cover slip. This setup is appropriated for observation that should not exceed 30–40 min.
4. For longer acquisition, dechorionated embryos are placed in FluoroDishes optimized for live microscopy and then cover with PBS to prevent embryo dehydration. (If embryos are properly dechorionated and transferred into dishes, they will hatch after 1 day.)

3.2 Image Acquisition

During embryonic Drosophila development, Pc bodies are highly mobile and therefore rapidly get out of focus. On one hand, fast and short time-lapse experiments are performed by acquiring 2D images during time. On the other hand, longer experiments required the acquisition of 3D volumes during time. The time resolution of 4D time-lapse experiments is worst than 2D movies because acquiring a volume (3D stack) takes much longer than collecting one image, but Pc bodies can be tracked for longer time.

1. We looked at epidermal cells (the most external layer of cells) by using a 60× N.A. 1.4 objective. The optical quality of the objective will decrease if we observe deeper cells in embryos.
2. Use confocal microscope to acquire time-lapse experiments (*see* **Note 1**).
3. Adjust laser power, pixel size, speed of acquisition, and number of sections per volume (for 4D time-lapse experiments) in order to collect images with good signal-to-noise ratio (*see* **Notes 2** and **3**).

3.3 Time-Lapse Image Processing

The most difficult and the most time-consuming step of this protocol is to process images of time-lapse experiments in order to quantify the mobility of Pc bodies. Firstly, image segmentation is required to define the coordinates (x, y, z) of Pc bodies. Secondly, we calculated M.S.D. and M.S.C by using their coordinates at different time points (*see* **Note 4**).

3.3.1 Image Segmentation

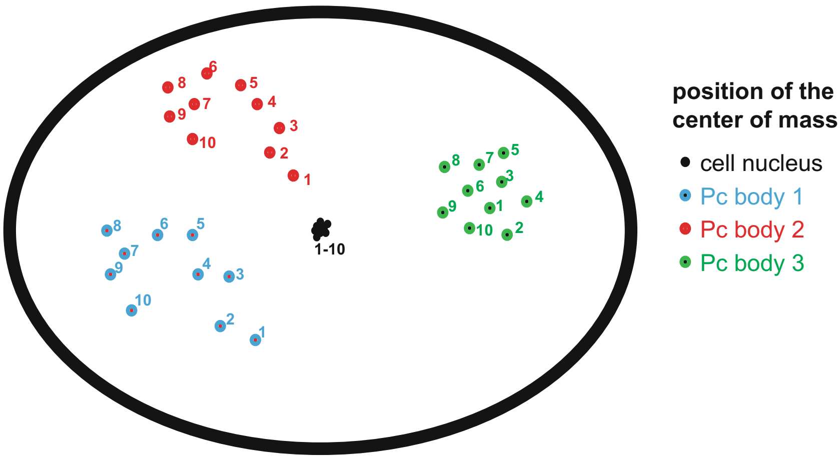
To segment our images, we use Volicity (Improvision) or Imaris (Bitplane).

1. To reduce the noise within images, we apply a mild median filter (3×3 for 2D images or $3 \times 3 \times 3$ for 3D images). Low-pass or Gaussian filters give similar results.
2. We crop one time-lapse experiment in order to only focus on one cell nucleus.
3. To define the Pc body that we wanted to track, we use threshold-based segmentation. We use the same approach to segment the cell nucleus.
4. We need to adjust the threshold several times in order to efficiently segment the Pc body during the entire experiment, because of the bleaching occurring during time-lapse experiments.
5. We manually remove all the objects which do not belong to the track that we want to analyze.
6. One important step is to verify that we always follow the same Pc body during the entire time-lapse by looking at the movie of segmented objects.
7. We extract the x, y, and z coordinates by using the center of mass of the segmented objects in each time point of the experiment.
8. We exported them in .txt files.

3.3.2 Track Analysis

We need to monitor the tracks of several Pc bodies per cell nucleus because they do not always move independently.

1. M.S.D. (Fig. 1): This measurement characterizes the motion of each object compared to the nucleus center, but do not take into account cell deformation.
2. M.S.C. (Fig. 1): Instead of looking at the movement of one Pc body between two time points (M.S.D.), you measure the change of distance between two Pc bodies during two time points (M.S.C.).
3. Scatterplot of M.S.D. vs. M.S.C at 1 s (or longer time): An absence of correlation between these two measurements evidences coordinated motion of Pc bodies. Indeed, if two Pc bodies move together, we would observe low M.S.C. and high M.S.D values.



Coordinate

time	nucleus	Pc body1	Pc body2	Pc body3			
t1	x1, y1	x1, y1	x1, y1	x1, y1	X1=x1-x1;Y1=y1-y1	X1=x1-x1;Y1=y1-y1	X1=x1-x1;Y1=y1-y1
t2	x2, y2	x2, y2	x2, y2	x2, y2	X2=x2-x2;Y2=y2-y2	X2=x2-x2;Y2=y2-y2	X2=x2-x2;Y2=y2-y2
t3	x3, y3	x3, y3	x3, y3	x3, y3	X3=x3-x3;Y3=y3-y3	X3=x3-x3;Y3=y3-y3	X3=x3-x3;Y3=y3-y3
⋮							
tn	xn, yn	xn, yn	xn, yn	xn, yn	Xn=xn-xn;Yn=yn-yn	Xn=xn-xn;Yn=yn-yn	Xn=xn-xn;Yn=yn-yn

M.S.D. calculation

MSDΔt1 = ((X2-X1)^2)+((Y2-Y1)^2)) + ((X3-X2)^2)+((Y3-Y2)^2))+...+((Xn-Xn-1)^2)+((Yn-Yn-1)^2)) / n-1

MSDΔt2 = ((X3-X1)^2)+((Y3-Y1)^2)) + ((X4-X2)^2)+((Y4-Y2)^2))+...+((Xn-Xn-2)^2)+((Yn-Yn-2)^2)) / n-2

MSDΔt3 = ((X4-X1)^2)+((Y4-Y1)^2)) + ((X4-X2)^2)+((Y4-Y2)^2))+...+((Xn-Xn-3)^2)+((Yn-Yn-3)^2)) / n-3

⋮

MSDΔt1 = ((X2-X1)^2)+((Y2-Y1)^2)) + ((X3-X2)^2)+((Y3-Y2)^2))+...+((Xn-Xn-1)^2)+((Yn-Yn-1)^2)) / n-1

⋮

MSDΔt1 = ((X2-X1)^2)+((Y2-Y1)^2)) + ((X3-X2)^2)+((Y3-Y2)^2))+...+((Xn-Xn-1)^2)+((Yn-Yn-1)^2)) / n-1

⋮

M.S.C. calculation

MSCΔt1 =[((x1-x1)^2+(y1-y1)^2)^0.5 - ((x2-x2)^2+(y2-y2)^2)^0.5]^2 + [((x2-x2)^2+(y2-y2)^2)^0.5 - ((x3-x3)^2+(y3-y3)^2)^0.5]^2 +...+[((xn-1-xn-1)^2+(yn-1-yn-1)^2)^0.5-((xn-xn)^2+(yn-yn)^2)^0.5]^2 / n-1

MSCΔt1 =[((x1-x1)^2+(y1-y1)^2)^0.5 - ((x2-x2)^2+(y2-y2)^2)^0.5]^2 + [((x2-x2)^2+(y2-y2)^2)^0.5 - ((x3-x3)^2+(y3-y3)^2)^0.5]^2 +...+[((xn-1-xn-1)^2+(yn-1-yn-1)^2)^0.5-((xn-xn)^2+(yn-yn)^2)^0.5]^2 / n-1

MSCΔt1 =[((x1-x1)^2+(y1-y1)^2)^0.5 - ((x2-x2)^2+(y2-y2)^2)^0.5]^2 + [((x2-x2)^2+(y2-y2)^2)^0.5 - ((x3-x3)^2+(y3-y3)^2)^0.5]^2 +...+[((xn-1-xn-1)^2+(yn-1-yn-1)^2)^0.5-((xn-xn)^2+(yn-yn)^2)^0.5]^2 / n-1

MSCΔt2 =[((x1-x1)^2+(y1-y1)^2)^0.5 - ((x3-x3)^2+(y3-y3)^2)^0.5]^2 + [((x2-x2)^2+(y2-y2)^2)^0.5 - ((x4-x4)^2+(y4-y4)^2)^0.5]^2 +...+[((xn-2-xn-2)^2+(yn-2-yn-2)^2)^0.5-((xn-xn)^2+(yn-yn)^2)^0.5]^2 / n-2

⋮

MSCΔt3 =[((x1-x1)^2+(y1-y1)^2)^0.5 - ((x4-x4)^2+(y4-y4)^2)^0.5]^2 + [((x2-x2)^2+(y2-y2)^2)^0.5 - ((x5-x5)^2+(y5-y5)^2)^0.5]^2 +...+[((xn-3-xn-3)^2+(yn-3-yn-3)^2)^0.5-((xn-xn)^2+(yn-yn)^2)^0.5]^2 / n-3

⋮

Fig. 1 Cartoon illustrating the motion of 3 PC bodies inside one cell nucleus. From time-lapse experiments, we compute the center of mass of one cell nucleus (black) and 3 PC bodies (blue, red and green) at different time points, which are used to calculate both M.S.D. and M.S.C

4 Notes

1. Confocal microscopy is the method of choice to perform live-cell imaging in Drosophila embryos. The light reflection induced by the vitteline membrane and the yolk autofluorescence are completely stopped by the pinhole of confocal microscopes, whereas they produce a relatively high background signal with regular epifluorescence microscopes. Although the cameras of the DeltaVision OMX (G.E. Healthcare Life Science) are more sensitive than the detectors of confocal microscopes, the quality of time-lapse experiments is better in confocal microscopy because of the high image contrast and the low image background.
2. To minimize phototoxicity and GFP bleaching and better follow the mobility of Pc bodies, you want to acquire images as fast as possible. However, a relatively good signal-to-noise (S/N) ratio is required to segment Pc bodies. The main difficulty is to find a good compromise between laser power, speed of acquisition, GFP bleaching, phototoxicity, and S/N ratio. We did not try image processing algorithms developed to improved S/N ratio, which could be useful for long time-lapse imaging.
3. Examples of time-lapse experiments:

   Imaging embryos expressing H2av-GFP is quite easy because the strong H2av-GFP signal gives images with good S/N ratio. In contrast, the signal of embryos expressing Ph-GFP is much weaker which prevents us to track Ph-GFP foci with the LSM 510 (Zeiss Imaging) because of bad S/N ratio. However, we were able to perform these time-lapse experiments by using the next generation of microscope (LSM 780: Zeiss Imaging) which is more sensitive.

   (a) Embryos expressing Pc-GFP (moderated signal): We used a Zeiss LSM 510 Meta with a pixel size of 70 nm and collected 2D images (512 × 128 pixels) every 250 ms for 15 s. For longer time-lapse experiments, we collected one volume (512 × 100 pixels: six optical sections with a Z-step of 500 nm) every 3 s during 3 min.

   (b) Embryos expressing H2av-GFP (strong signal): We used a Zeiss LSM 510 Meta with a pixel size of 48 nm and collected 2D images (512 × 128 pixels) every 250 ms for 15 s. (We can collect images with a smaller pixel size because of the stronger signal. Consequently, image segmentation will be more accurate.)

   (c) Embryos expressing Pc-GFP, Ph-GFP or H2av-GFP: We used a Zeiss LSM 780 with GaAsP detector and a pixel size of 66 nm to acquire 2D images (512 × 256 pixels) every

250 ms for 15 s. Long 4D tracking experiments were recorded with a pixel size of 88 nm, one volume (512×512 pixels; 12 sections with a z-step of 0.5 μm) every 10 s during 30 min.

4. After image segmentation, the precision of Pc bodies coordinates depends on the pixel size. We observed a bias when Pc bodies move along the z-axis because the z-step is much larger than the x–y pixel size. To prevent this, we did not take into account the z-coordinate to calculate M.S.D. and M.S.C. Another solution would be to increase the number of optical sections by using a z-step of 100 nm or less. However, image acquisition should be fast enough in order to neglect the mobility of Pc bodies between two optical sections.

References

1. Lanctot C, Cheutin T, Cremer M, Cavalli G, Cremer T (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat Rev Genet 8:104–115
2. Wachsmuth M, Caudron-Herger M, Rippe K (2008) Genome organization: balancing stability and plasticity. Biochim Biophys Acta 1783:2061–2079
3. Ficz G, Heintzmann R, Arndt-Jovin DJ (2005) Polycomb group protein complexes exchange rapidly in living Drosophila. Development 132:3963–3976
4. Dietzel S, Niemann H, Bruckner B, Maurange C, Paro R (1999) The nuclear distribution of Polycomb during Drosophila melanogaster development shown with a GFP fusion protein. Chromosoma 108:83–94
5. Netter S, Faucheux M, Theodore L (2001) Developmental dynamics of a polyhomeotic-EGFP fusion in vivo. DNA Cell Biol 20:483–492
6. Buchenau P, Hodgson J, Strutt H, Arndt-Jovin DJ (1998) The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. J Cell Biol 141:469–481
7. Cheutin T, Cavalli G (2012) Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. PLoS Genet 8:e1002465

Chapter 25

Reprogramming of Somatic Cells Towards Pluripotency by Cell Fusion

Andrzej R. Malinowski and Amanda G. Fisher*

Abstract

Pluripotent reprogramming can be dominantly induced in a somatic nucleus upon fusion with a pluripotent cell such as embryonic stem (ES) cell. Cell fusion between ES cells and somatic cells results in the formation of heterokaryons, in which the somatic nuclei begin to acquire features of the pluripotent partner. The generation of interspecies heterokaryons between mouse ES- and human somatic cells allows an experimenter to distinguish the nuclear events occurring specifically within the reprogrammed nucleus. Therefore, cell fusion provides a simple and rapid approach to look at the early nuclear events underlying pluripotent reprogramming. Here, we describe a polyethylene glycol (PEG)-mediated cell fusion protocol to generate interspecies heterokaryons and intraspecies hybrids between ES cells and B lymphocytes or fibroblasts.

Key words Reprogramming, Pluripotency, Cell fusion, Embryonic stem (ES) cell, Heterokaryon, Hybrid, Polyethylene glycol (PEG)

1 Introduction

The origins of experimental cell fusion date back nearly 50 years to experiments used to map biological properties to specific chromosomal locations and to generate monoclonal antibodies [1, 2]. More recently, cell fusion has been employed to investigate the reprogramming of differentiated cells towards pluripotency [3, 4]. In order to achieve this, somatic cells (such as B lymphocytes or fibroblasts) are fused with pluripotent cells (e.g., embryonic stem [ES] or embryonic germ [EG] cells). As a result, the somatic nucleus undergoes a series of reprogramming events associated with the acquisition of a pluripotent identity [3, 5–7].

The two other reprogramming strategies: induction of pluripotent stem (iPS) cells by transcription factor transduction and somatic cell nuclear transfer (SCNT), hold therapeutic potential, but are less accessible for detailed mechanistic studies. In contrast, cell fusion is a low-cost, fast, and simple approach that can unlock the mechanisms which are important for initiating pluripotent reprogramming [8].

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_25,

Cell fusion occurs in two sequential stages (Fig. 1a). First, a heterokaryon is formed, in which the two parental nuclei remain separate, but are contained within a single cell body. Second, nuclear fusion takes place and gives rise to a hybrid cell carrying a tetraploid

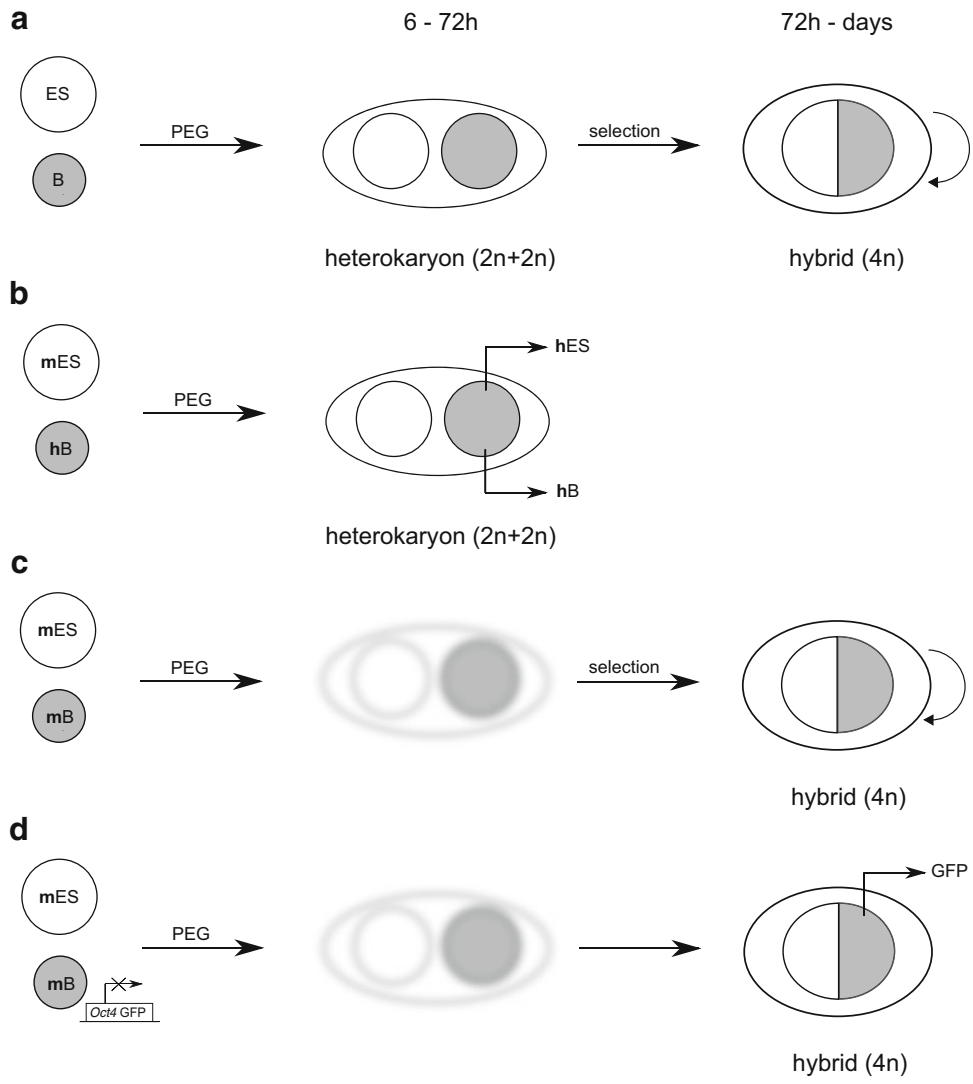


Fig. 1 Schematic representation of cell fusion-based pluripotent reprogramming (**a**) ES cells and B cells are fused using PEG to generate heterokaryons in which two nuclei share one cytoplasm, but remain discrete. This is followed by nuclear fusion and formation of self-renewing hybrid cells with tetraploid nuclei. (**b**) Experimental strategy to generate transient interspecies heterokaryons between mouse ES (mES) cells and human B (hB) lymphocytes. The resulting heterokaryons initiate a human ES-specific gene expression profile. (**c**) Experimental strategy to generate stable intraspecies hybrids between mES cells and mB lymphocytes. Intraspecies heterokaryons develop into stable hybrid cells which can be assessed for the pluripotent phenotype. Selection is used to eliminate unfused cells. (**d**) Experimental strategy to identify successfully reprogrammed hybrids using a somatic cell line with a silent pluripotent reporter. Intraspecies hybrids are generated by fusing mES cells with mB lymphocytes that carry a silent GFP transgene under the control of *Oct4* regulatory elements. Successfully reprogrammed hybrids reactivate GFP expression after 5–6 days

chromosome complement [5]. Heterokaryons can be generated by fusing cells from the same (intraspecies fusion) or different species (interspecies fusion). While intraspecies heterokaryons often progress to form karyotypically stable hybrids, interspecies heterokaryons are typically karyotypically unstable [5]. Interspecies heterokaryons, can however, be easily used to monitor nuclear events and transcriptional changes that are initiated within the reprogrammed nucleus (Fig. 1b) [9]. To this end, heterokaryon generation can be combined with a number of approaches, including species-specific gene expression analysis (quantitative RT-PCR or RNA-sequencing) and chromatin/DNA modifications profiling (e.g., immunofluorescence or bisulfite sequencing) [10]. Moreover, fusion experiments using genetically manipulated cells have been used to screen for specific factors such as Polycomb-group (PcG) proteins that are required for reprogramming [11]. The analysis of hybrid cells formed from heterokaryons can often be used to evaluate the acquisition of a stable pluripotent state (Fig. 1c). As only a small proportion of cells may fuse in experimental procedures described to date, heterokaryon and hybrid analysis often relies on strategies that select for fused cells, or against unfused cells in the population.

Here, we describe a cell fusion approach to reprogram somatic cells towards pluripotency using polyethylene glycol (PEG) as a fusing agent. The first section of the protocol details human–mouse heterokaryon formation and gene expression analysis by quantitative RT-PCR. The second section describes generation of mouse–mouse hybrids and provides an example of hybrid analysis (alkaline phosphatase staining).

2 Materials

2.1 Cell Culture Materials

1. Class II safety cabinet.
2. 50% (w/v) polyethylene glycol (PEG) 1500 in 75 nM HEPES pH 8.0 (Roche) (*see* **Note 1**).
3. Knockout Dulbecco's Modified Eagle's Medium (KO-DMEM) (Invitrogen) (*see* **Note 2**).
4. Calcium- and magnesium-free phosphate-buffered saline (CMF-PBS) (Invitrogen).
5. 0.05% (w/v) trypsin–EDTA (Invitrogen).
6. 0.1% (w/v) Gelatin: 25 ml liquefied 2% (v/v) gelatin (Sigma), 475 ml CMF-PBS. Filtered with a bottle-top filter (0.22-μm, Millipore).
7. Leukemia inhibitory factor (LIF) (Esgro, Merck Millipore).
8. Puromycin (Sigma) (*see* **Note 3**).
9. Liquid nitrogen.

10. 175 cm^2 tissue culture flasks.
11. 90 mm tissue culture dishes.
12. 37 °C water bath.
13. 37 °C, 5% CO_2 incubator.
14. Centrifuge.
15. 10 ml pipette.
16. 15 and 50 ml conical tubes (BD Falcon).
17. Conical 30 ml universal tubes (Sterilin).
18. 1.5 ml plastic tubes.
19. Hemacytometer.
20. Plastic Pasteur pipettes.

2.2 Cell Lines

1. Epstein-Barr virus (EBV) transformed human B (hB) cells carrying puromycin resistance cassette (*see* **Notes 4** and **5**).
2. Abelson virus transformed mouse pre-B (mB) cells carrying a puromycin resistance cassette.
3. Mouse embryonic stem (mES) cells cultured on gelatin-coated dishes.
4. Feeders: Primary mouse embryonic fibroblasts (MEF), mitotically inactivated with γ-irradiation. Expand MEFs for 2–3 passages, harvest confluent cells with trypsin and resuspend in MEF media w/o antibiotics in 50 ml conical tubes. Expose cells to 1.04 Gy/min for 100 s. Freeze in 10% (v/v) DMSO (Sigma), heat-inactivated fetal bovine serum (FBS).

2.3 Cell Media

1. hB cell medium: RPMI-1640 (Invitrogen), 10% (v/v) heat-inactivated FBS, 2 mM l-glutamine (Invitrogen), 10 μg/ml penicillin and streptomycin (Invitrogen).
2. mB cell medium: RPMI-1640 (Invitrogen), 20% (v/v) heat-inactivated FBS, 2 mM l-glutamine (Invitrogen), 10 μg/ml penicillin and streptomycin (Invitrogen), 1× nonessential amino acids (Invitrogen), 50 μM 2-mercaptoethanol (Invitrogen).
3. mES medium: KO-DMEM (Invitrogen), 10% (v/v) FBS batch tested for ES culture, 2 mM l-glutamine (Invitrogen), 10 μg/ml penicillin and streptomycin (Invitrogen), 1× nonessential amino acids (Invitrogen), 50 μM 2-mercaptoethanol (Invitrogen), 1000 U/ml of LIF—add fresh to medium.
4. MEF medium: DMEM (Invitrogen), 10% (v/v) heat-inactivated FBS, 2 mM l-glutamine (Invitrogen), 10 μg/ml penicillin and streptomycin (Invitrogen), 1× nonessential amino acids (Invitrogen), 50 μM 2-mercaptoethanol (Invitrogen).

2.4 Molecular Biology Materials

1. RNeasy Mini Kit (Invitrogen).
2. RNase-free water, e.g., diethylpyrocarbonate (DEPC)-treated water.

3. Turbo DNA-free kit (Ambion).
4. Superscript III First-Strand Synthesis System (Invitrogen).
5. SYBR green Master Mix for quantitative RT-PCR (Qiagen).
6. Human gene-specific primers (*see* **Note 6**).
7. NanoDrop ND-1000 spectrophotometer.
8. Thermal cycler.
9. 96-well plates for quantitative RT-PCR (Bio-Rad).
10. Quantitative RT-PCR engine.
11. Alkaline phosphatase kit (Sigma).

3 Methods

Steps 1–32 of the heterokaryon protocol and **steps 1–7** of the hybrid protocol are performed in a Class II safety cabinet using aseptic technique, sterile reagents and equipment.

3.1 Generation of Interspecies Heterokaryons Between Human B Lymphocytes and Mouse ES Cells Under Drug Selection

Cells Preparation

1. Prepare gelatinized 90 mm dishes. Add 5 ml 0.1% gelatin to cover the bottom of a dish. Incubate at 37 °C for at least 20 min.
2. Grow puromycin-resistant hB lymphocytes in suspension at $0.7–1.3 \times 10^6$ cell/ml using 175 cm^2 tissue culture flasks. Add fresh medium the day before fusion.
3. In parallel, grow mES cells on 90 mm gelatin-coated dishes. Plate mES cells 48 h before fusion so that the cells are 60–70% confluent on the day of fusion. Change medium the day before fusion.
4. Pre-warm PEG solution, KO-DMEM, and complete mES medium in a 37 °C water bath (*see* **Note 7**).
5. Harvest mES cells by adding 2 ml of 0.05% trypsin–EDTA to 90 mm culture dishes and incubate at 37 °C, 5% CO_2 for 5 min. Add 8 ml of mES medium to inactivate trypsin. Gently dissociate mES cell colonies to single cells by pipetting up and down (*see* **Note 8**).
6. Collect both cell types in conical tubes, centrifuge for 5 min at $290 \times g$, room temperature and aspirate the supernatant.
7. Wash both cell types twice in pre-warmed KO-DMEM and count cells. Typically, 1×10^7 mES cells are combined with 1×10^7 hB cells (*see* **Note 9**).
8. Resuspend each cell type population in 10 ml of pre-warmed KO-DMEM.

Fusion

9. Mix both cell types in a plastic universal tube in 1:1 ratio (*see* **Note 10**).

10. Centrifuge the cells for 5 min at 290 × *g*, room temperature and aspirate the supernatant.
11. Wash the cells in 10 ml of pre-warmed KO-DMEM and aspirate the supernatant completely (*see* **Note 11**).
12. Incubate the cell pellet in a 37 °C water bath for 60 s, gently swirling the tube.
13. Loosen the cell pellet by gently tapping the bottom of the tube.
14. Add 350 µl of PEG dropwise continuously over 60 s, directly to the pellet, simultaneously rocking/swirling the tube gently (*see* **Note 12**).
15. Incubate the cells with PEG in a 37 °C water bath for 90 s swirling the tube gently (*see* **Note 13**).
16. Add 350 µl (1 volume of PEG) of pre-warmed KO-DMEM dropwise over 60 s, rocking/swirling the tube gently (*see* **Note 14**).
17. Add 1.050 ml (three volumes of PEG) of pre-warmed KO-DMEM dropwise over 180 s, simultaneously rocking/swirling the tube gently.
18. Slowly add 10 ml of KO-DMEM.
19. Mix by gently inverting the tube 4–6 times.
20. Incubate in a 37 °C water bath for 3 min.
21. Centrifuge the cells for 5 min at 370 × *g*, 37 °C.
22. Aspirate the supernatant and carefully add 10 ml of pre-warmed mES medium without disrupting the pellet.
23. Incubate the cells at 37 °C for 3 min.
24. Carefully resuspend cells in the required volume of mES medium (*see* **Note 15**).
25. Remove an aliquot of fused cells (routinely $1–2 \times 10^6$) for Day 0 sample, centrifuge for 5 min at 290 × *g*, room temperature. Wash in CMF-PBS and aspirate the supernatant. Snap-freeze in liquid nitrogen and store at −80 °C.
26. Remove the gelatin solution from the dishes and plate the cells on a 90 mm dish in 10 ml of mES medium.
27. Culture the cells overnight at 37 °C, 5 % CO_2.

Culturing and Harvesting Heterokaryons

28. Add 1 µg/ml puromycin 8–12 h after fusion to select against unfused mES cells (*see* **Note 16**).
29. Starting from day 1 daily, gently wash the plates three times in CMF-PBS (*see* **Note 17**) and add new medium supplemented with puromycin.
30. On days 1, 2, and/or 3 as required, aspirate medium, wash three times with CMF-PBS and harvest heterokaryons by adding 2 ml of 0.05 % trypsin–EDTA to the culture dish. Incubate

at 37 °C, 5% CO_2 for 5 min. Add 8 ml of ES medium to inactivate trypsin and dissociate ES cell colonies to single cells by gently pipetting up and down.

31. Collect cells in 15 ml conical tubes and centrifuge 5 min at 290 × *g*, room temperature.
32. Wash the cells in CMF-PBS and centrifuge for 5 min at 290 × *g*, room temperature. Aspirate the supernatant and snap freeze the pellet in liquid nitrogen. Store at −80 °C.

Expression Analysis in Heterokaryons by Species-Specific Quantitative RT-PCR (See Note 18)

33. Extract RNA from frozen samples using RNeasy Mini Kit according to the manufacturer's instructions.
34. DNase-treat the samples to remove any contaminating DNA using Turbo DNA-free kit according to the manufacturer's instructions.
35. Measure the RNA concentration with a NanoDrop spectrophotometer.
36. Reverse transcribe 1–3 μg of RNA with the Superscript First-Strand Synthesis System according to the manufacturer's instructions using a thermal cycler.
37. Dilute cDNA in DEPC water.
38. Perform quantitative RT-PCR using SYBR green master mix and human gene-specific primers to determine the abundance of human-specific transcripts in the heterokaryons at each time point of the heterokaryon assay.
39. Normalize the amount of target transcripts to a human housekeeping gene, e.g., *GAPDH* (Fig. 2).

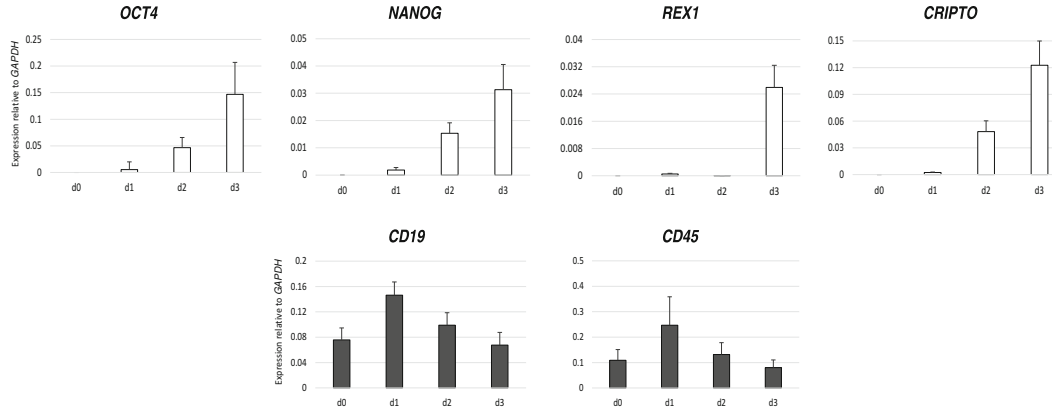


Fig. 2 Investigating the kinetics of human B (hB) lymphocyte reprogramming by mouse ES (mES) cells in heterokaryons. mES cells fused with hB lymphocytes to generate experimental heterokaryons were harvested at 1, 2, and 3 days after cell fusion. Human gene-specific quantitative RT-PCR was performed to assess upregulation of human ES-specific genes (*OCT4*, *NANOG*, *REX1*, *CRIPTO*) and downregulation of human lymphocyte-specific genes (*CD19*, *CD45*). Expression relative to human *GAPDH*. Error bars indicate SEM of three independent experiments

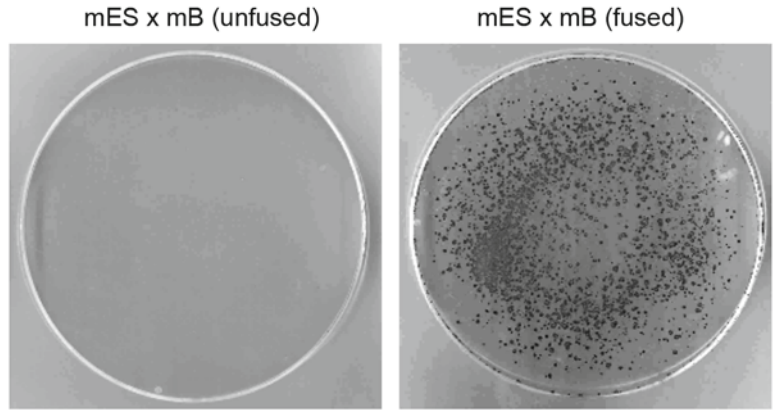


Fig. 3 Assessing the pluripotent reprogramming in hybrids by alkaline phosphatase (AP) staining. Puromycin resistant mouse B (mB) lymphocytes and mouse ES (mES) cells were fused, cultured for 10 days under puromycin selection and stained with alkaline phosphatase (*right*); unfused control (*left*)

3.2 Generation of Intraspecies Hybrids Between Mouse B Lymphocytes and Mouse ES Cells Under Drug Selection

1. Plate down 2×10^6 irradiated feeders per 90 mm dish in MEF media the day before fusion (*see* **Note 19**).
2. Grow puromycin-resistant mB lymphocytes in suspension at $0.7–1.3 \times 10^6$ cell/ml using 175 cm^2 tissue culture flasks. Add new medium the day before fusion.
3. In parallel, grow mouse ES cells on 90 mm gelatin-coated dishes. Plate ES cells 48 h before, so that the cells are 60–70% confluent on the day of fusion. Change medium the day before fusion.
4. Follow **steps 4–24** of the heterokaryon protocol.
5. Aspirate MEF medium and plate $0.5–2 \times 10^6$ fused cells per a 90 mm dish onto feeders in 10 ml of ES medium (*see* **Note 20**).
6. Add puromycin 16–18 h after fusion to select against unfused ES cells (*see* **Note 21**). Daily, gently but thoroughly wash the plates three times in CMF-PBS (*see* **Note 17**) and add new medium supplemented with puromycin.
7. On day 5–6 after fusion, individual colonies can be analyzed for re-activation of a pluripotent reporter (Fig. 1d) (*see* **Note 22**) and/or picked for single hybrid analysis (*see* **Note 23**).
8. Day 8–12 after fusion, perform alkaline phosphatase staining according to manufacturer's instructions (Fig. 3) (*see* **Note 24**).

4 Notes

1. PEG [$H(OCH_2CH_2)nOH$] acts as a dehydrating agent causing cell agglutination and increased cell-to-cell contacts, which promotes cell fusion [12]. We recommend using the following volumes of PEG per number of ES cells and B lymphocytes

fused: 10^8 of each cell type—1 ml PEG; 10^8—5×10^7—500 μl PEG; $<5 \times 10^7$—350 μl PEG.

2. KO-DMEM can be substituted by an alternative serum-free medium, e.g., DMEM or RPMI 1640.
3. Since only a proportion of cells forms heterokaryons, enriching for successfully fused cells facilitates gene expression analysis on the reprogrammed population. To this end, we recommend using drug selection/cell lines carrying drug resistance genes or selective markers. For each drug optimal working concentration should be first determined by titration on the drug-sensitive cells used for fusion.
4. This protocol uses B lymphocytes; however, different somatic cell types such as fibroblasts may be used. It must be noted that different somatic cell types may display different reprogramming efficiencies. The kinetics of reprogramming can also vary, according to cell type and/or cell cycle stage. For example, in our hands fibroblasts show a delayed reprogramming compared to B lymphocytes.
5. As an alternative to fusing adherent (mES) and non-adherent (B lymphocyte) cell types, cell fusion experiments can be performed between two adherent or two non-adherent cell types. In these cases a strategy for selecting against unfused cells of each cell type by using a combination of drugs/drug resistant cell lines or cell labeling/FACS sorting approach should be considered (*see also* **Note 16**).
6. Quantitative RT-PCR primers for the gene expression analysis in heterokaryons must be human-specific. This should be tested in silico PCR against mouse genes as well as in quantitative RT-PCR assay using mES cDNA template.
7. Optimal temperature for cells fusion is 37 °C. PEG, KO-DMEM, and complete ES medium should be pre-warmed at 37 °C prior to fusion.
8. It is important to obtain a single-cell suspension of ES cells at the end of this step as clusters of ES cells will promote ES–ES fusion.
9. In this protocol 1×10^7 mES cells are fused with 1×10^7 hB cells, which is sufficient for analyzing reprogramming at one time point (e.g., day 3) . For applications which require high numbers of fused cells such as gene expression kinetics in heterokaryons, we recommend scaling up fusion experiments.
10. Experimental cell fusion occurs with the efficiency of up to 11 % when using 1:1 ES cells–B lymphocytes ratio [10].
11. Removing all the residual supernatant from the tube prevents diluting PEG.
12. Slow addition of PEG to the cells minimizes the osmotic shock. Rocking/swirling the tube with the pellet promotes forming cell–cell interactions.

13. This step is crucial for initiating cell–cell fusions and should be performed at 37 °C.
14. PEG is gradually diluted in KO-DMEM in order to avoid disturbing inter-cell contacts. Dilution must be gentle (i.e., no vortexing) to avoid cell lysis due to osmotic pressure.
15. Fused cells should be handled gently as heterokaryon formation may continue up to 6 h after fusion.
16. As an alternative to drug selection, successfully fused cells can be enriched by fluorescent cell labeling followed by FACS sorting. In this case, each cell type is labeled with a different live cell fluorescent dye (e.g., Violet/CFSE CellTrace™; Life Technologies) directly before fusion. Double-labeled heterokaryons can be FACS sorted before, or, after culturing. In our hands sorting after culturing results in better yield and higher viability.
17. Extensive washes in PBS are designed to remove cell debris and unfused B cells.
18. As well as gene expression assays, heterokaryons can by analyzed by a number of experimental techniques including immunofluorescence, fluorescence in situ hybridization, bisulfite genomic sequencing, and FACS [7].
19. Hybrids are cultured on a feeder layer in order to increase their viability and survival. Alternatively, it is possible to culture hybrids on 0.1 % gelatin coated-dishes.
20. The density of plating depends on the specific assay for which hybrids are generated. For example, the density should be lower (0.5×10^6) when performing Alkaline Phosphatase assay in order to obtain discrete colonies or higher (2×10^6) when hybrids that reactivated a pluripotent reporter are isolated by FACS.
21. Efficient selection for heterokaryons/resulting hybrids is essential as contaminating unfused cells can affect the analysis of the hybrids. Elimination of unfused cells is usually achieved by using cell lines carrying drug resistance genes or selective markers. As an alternative, hypoxanthine–aminopterin–thymidine (HAT) selection can be used to eliminate cells deficient for hypoxanthine–guanine phosphoribosyl transferase (*Hprt*) gene (e.g., E14tg2a mES cells). Aminopterin blocks DNA synthesis by inhibiting nucleotide synthesis. Hypoxanthine and thymidine provide a salvage pathway, which can only be used by cells with HPRT enzyme activity. For each drug selection strategy working concentration/kinetics should be first determined by titration on the non-resistant fusion partner cell line. It is important to culture a mock fusion without PEG alongside the experimental fusion in order to monitor the selection. Analysis should be performed only after the selection is complete. Additionally, we advise to continue adding the selective drug throughout the entire course of the experiment to eliminate any hybrid cells suffering from chromosome/complement loss.

22. Successfully reprogrammed hybrids initiate expression of pluripotent markers such as Oct4, Nanog, or Sox2 from the somatic cell genome. Thus, by employing a somatic cell line harboring a silent pluripotent reporter, e.g., a mouse lymphocyte that carries a silent transgene (green fluorescent protein [GFP] under the control of Oct4 regulatory elements) can be used to easily identify, isolate, and analyze successfully reprogrammed hybrid cells.
23. One of the advantages of the hybrid approach is that each heterokaryon can give rise to unique self-renewing hybrid cells which can be followed in culture. Once proliferating, individual hybrid colonies may be isolated manually, cultured as clones and examined in respect to their genotype and phenotype. We recommend growing hybrid clones on feeders and in the presence of selective drug(s). As a proportion of hybrid clones may be affected by chromosome loss especially during extensive culture, their tetraploid status should be monitored by an assay determining the DNA content of cells, e.g., propidium iodide (PI) staining, or by conventional karyotype analysis.
24. Alkaline phosphatase, a marker of pluripotent cells [13], can be used to identify successfully reprogrammed hybrids.

References

1. Weiss MC, Green H (1967) Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. Proc Natl Acad Sci U S A 58(3):1104–1111
2. Kohler G, Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256(5517):495–497
3. Tada M et al (2001) Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr Biol 11(19):1553–1558
4. Cowan CA et al (2005) Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science 309(5739):1369–1373
5. Piccolo FM et al (2011) Using heterokaryons to understand pluripotency and reprogramming. Philos Trans R Soc Lond B Biol Sci 366(1575):2260–2265
6. Soza-Ried J, Fisher AG (2012) Reprogramming somatic cells towards pluripotency by cellular fusion. Curr Opin Genet Dev 22(5):459–465
7. Pereira CF et al (2008) Heterokaryon-based reprogramming of human B lymphocytes for pluripotency requires Oct4 but not Sox2. PLoS Genet 4(9):e1000170
8. Yamanaka S, Blau HM (2010) Nuclear reprogramming to a pluripotent state by three approaches. Nature 465(7299):704–712
9. Terranova R et al (2006) Acquisition and extinction of gene expression programs are separable events in heterokaryon reprogramming. J Cell Sci 119(Pt 10):2065–2072
10. Pereira CF, Fisher AG (2009) Heterokaryon-based reprogramming for pluripotency. Curr Protoc Stem Cell Biol Chapter 4:Unit 4B.1
11. Pereira CF et al (2010) ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. Cell Stem Cell 6(6):547–556
12. Hui SW, Stenger DA (1993) Electrofusion of cells: hybridoma production by electrofusion and polyethylene glycol. Methods Enzymol 220:212–227
13. O'Connor MD et al (2008) Alkaline phosphatase-positive colony formation is a sensitive, specific, and quantitative indicator of undifferentiated human embryonic stem cells. Stem Cells 26(5):1109–1116

Chapter 26

Imaginal Disc Transplantation in Drosophila

Tomonori Katsuyama* **and Renato Paro**

Abstract

Since Ephrussi and Beadle introduced imaginal disc transplantation to *Drosophila* research in 1936, the method played an important part towards a better understanding of disc patterning, tissue regeneration, and reprogramming phenomena like transdetermination. Despite increasing usage of high-throughput approaches towards solving biological problems this classical manual method is still in use for studying disc development in a semi-physiological context. Here we describe in detail a protocol and provide recommendations on the procedure in particular for analyzing the regenerative potential of imaginal disks. The steps consist of disc dissection and fragmentation, transplantation into the larval or adult abdomen, and the recovery of implants from the host abdomen. Additionally, we also describe how to make the special transplantation needle from a glass capillary.

Key words *Drosophila* imaginal discs, Polycomb, Transplantation, Epimorphic regeneration

1 Introduction

The imaginal disc transplantation method was introduced about 80 years ago to *Drosophila* research [1], and it has been employed to elucidate various questions of developmental biology. For example, the method played a significant role in demonstrating the maintenance of determined cellular states in imaginal discs [2, 3], which continues to remain a popular research topic as mechanisms of cellular memory are now being examined at the molecular level [4]. A donor larva is dissected and the isolated imaginal disc is injected with a transplantation needle (micro glass capillary) into the body cavity of a host larva or fly [5]. In a larval environment the transplanted disc fragment grows synchronously with host development and during metamorphosis and differentiates to a predetermined adult structure. In this way, comprehensive fate maps of each disc could be generated [6, 7]. When the imaginal disc or its fragment is cultured in an adult abdomen, the disc cells proliferate but remain undifferentiated. Upon appropriate cultivation time the fragments regenerate the lost part in an epimorphic

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_26,

manner and can reacquire the normal pattern. Hence, disc transplantation has been extensively used to study tissue regeneration (reviewed in ref. 8). Infrequently, regenerating discs can gain alternate identities of other discs by aberrant cellular programming, a phenomenon termed transdetermination [9]. Previous studies in our laboratory showed that disc regeneration and transdetermination are coupled to the regulation of the Polycomb group (PcG) proteins [10]. PcG proteins epigenetically maintain repressed gene expression states stable and heritable. PcG-mediated gene silencing was found to be downregulated in proliferative regenerating cells in transplanted discs, which may increase the cellular plasticity to reactivate the required developmental patterning and regeneration genes. Consistently, downregulation of PcG silencing in regenerated blastema was a prerequisite for transdetermination. The disc transplantation technique has also been applied to cancer research in *Drosophila* [11, 12]. The dysfunction of the PcG system in imaginal discs leads to neoplastic overgrowth [13, 14]. We applied ex vivo disc cultivation in adult abdomen to assess the genomic instability and accumulation of genetic mutations during long-term tumor progression [15].

Here we describe the technique for disc transplantation, which has been taught to us by the late Walter J. Gehring. The protocol begins with the preparation of hand-made transplantation needles, followed by disc dissection and fragmentation, transplantation into larval or adult abdomen, and the recovery of implants from the host abdomen. These procedures appear technically simple, yet require extensive practice to master the delicate manual operations.

2 Materials

2.1 Fly Lines

1. Donor: genotypes of interest.
2. Host flies and larvae: wild type, w^{1118}, or other genotypes of interest.

2.2 Reagents

1. Chan and Gehring's Balanced Saline (BSS): 3.2 g NaCl, 3.0 g KCl, 1.8 g $MgSO_4$-$7H_2O$, 0.69 g $CaCl_2$-$2H_2O$, 1.79 g tricin, 3.6 g glucose, and 17.1 g Sucrose per liter of water. Adjust to pH 6.95 with NaOH. Autoclave 121 °C/20 min. Make 10 mL aliquots in 15 mL tubes and store at −20 °C. Ringer's solution [16] can also be used instead of BSS.
2. BSS containing 0.1% BSA (BSSB). When BSA is added to BSS, do not autoclave but filter-sterilize with 0.2 μm filter. Make 10 mL aliquots in 15 mL centrifuge tubes and store at −20 °C.
3. Diethyl ether.
4. Needle cleaning solution (30 g potassium persulfate per liter of 80% sulfuric acid). Add 96% sulfuric acid gradually to

Milli-Q grade water on ice and make 80% solution. Add potassium persulfate.

5. Phosphate buffered saline (PBS, pH 7.0).
6. PBS containing 0.3% Triton X-100 (PBT).
7. 70% ethanol.

2.3 Materials and Equipment

1. Dissection microforceps (Fine Science Tools/Dumont #5).
2. Double-sided sticky tape 10 mm width.
3. Filter paper 2 × 5 cm, autoclaved.
4. Gelatin-coated glass slide.
5. Glass capillary (outer diameter (O.D.): 1 mm/inner diameter (I.D.): 0.7–0.8 mm; e.g., Drummond Scientific Co./Cat# 1-000-0300).
6. Glass slide uncoated.
7. Glass syringe of 2 mL with metal Luer lock tip.
8. Implement for ether anesthesia (*see* Fig. 4).
9. Microforge (e.g., Narishige MF-900).
10. Needle holder for 1.0 mm OD glass capillary.
11. Objective micrometer.
12. Plastic plate (e.g., 8 × 10 cm).
13. Silicon-coated glass slide.
14. Stereomicroscope.
15. Tungsten needle.
16. Vinyl tubing with O.D. 6 mm/I.D. 4 mm.

3 Methods

3.1 Manufacturing Transplantation Needles

1. Extend glass capillaries with a small flame (Fig. 1a). Hold both ends of a capillary and heat the center over a small flame of Bunsen burner for a few seconds (*see* **Note 1**). Pull the capillary ends straight after moving it away from the flame.
2. Cut the extended capillary at the center.
3. Break a tip of the clipped end with microforceps to give the shape of an injection needle (Fig. 1b, *see* **Note 2**).
4. Measure the inner diameter (I.D.) of sharpened end. Group the capillaries depending on their I.D. (*see* **Note 3**).
5. Create a constriction at the needle's end by using microforge (Fig. 1c, *see* **Note 4**). Place a needle tip to face to a microforge filament. Once the needle tip is bent around 30–45°, turn the needle 180° and bend it until it is getting strait. As an optional step, turn the needle around 90/270°, and repeat the bend and straighten to form an hourglass shape.

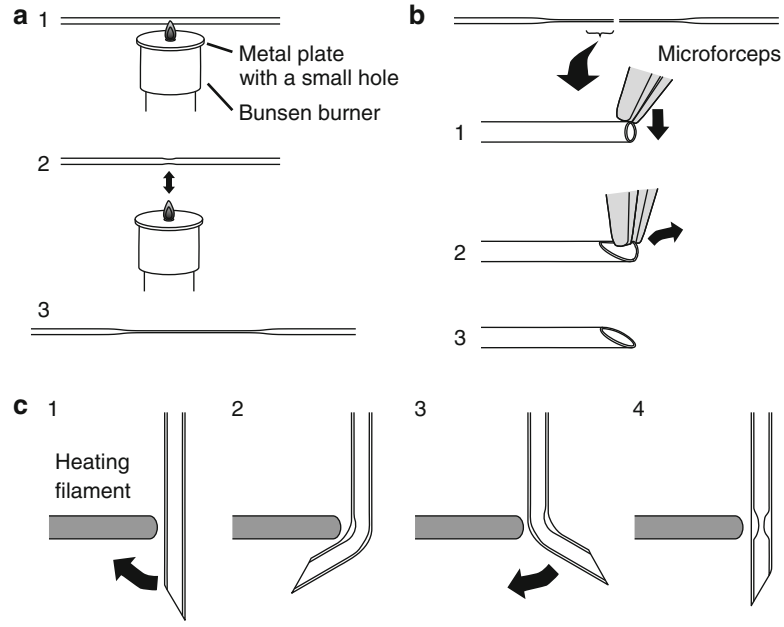


Fig. 1 How to make a transplantation needle. Chart depicting the manufacturing of a transplantation needle. (**a**) Pulling a glass capillary with a small flame. To make an appropriate flame, a metal plate with a small hole can be put on a Bunsen burner. (**b**) Shaping the needle tip. Make a crack at the extended capillary end by pushing a microforceps on it. If the end is chipped like in (**b***2*), break the unwanted edge outward with a microforceps. (**c**) Making a constriction with a microforge

3.2 Preparation of Donor Larvae and Host Flies/Larvae

1. For disc donors, we usually use larvae 100 h after egg laying at 25 °C. To obtain stage-synchronized cohorts of larvae, collect embryos in a 1-h window after two cycles of pre-collections and incubate them at 25 °C for exactly 100 h. If the discs of staged larvae are not required, use third-instar larvae still crawling around the food surface.
2. Host fly lines should be reared at 18 °C. We usually collect females hatched within 24 h and keep them in an empty small vial for 12–24 h at 18 °C for starvation (*see* **Note 5**). Newly emerged females without starvation can also be used.
3. Host larvae are reared for 72 h at 25 °C and then for 8 h at 18 °C. The stage-synchronized larvae should be prepared accordingly.

3.3 Transplantation of Imaginal Discs

1. Thaw an aliquot tube of BSS (10 mL). Spin down for removing insoluble material. Use a new BSS aliquot for each experiment.
2. Assemble the transplantation apparatus (Fig. 2). Fill the vinyl tubing connected to a glass syringe with distilled water. Set a plunger avoiding air bubbles inside. Connect a needle holder to the other side of the tube and take up approximately 1 mL

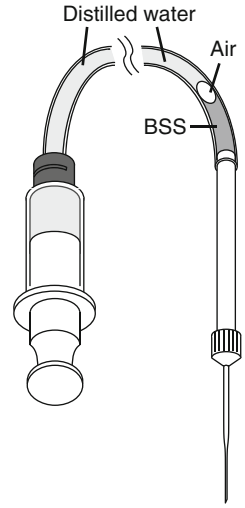


Fig. 2 Assembly of a transplantation apparatus. The plastic tubing is filled with water and BSS, which are separated by an air phase

BSS by pulling the syringe plunger. Air space will be created between the water and BSS, which allows separating them from each other. Connect a transplantation needle and press the plunger to fill BSS in (*see* **Note 6**).

3. Clamp the glass syringe to a metal stand and arrange it to the left of a stereomicroscope. Place the transplantation needle to the other side by passing the connecting tube behind the microscope. If the operator is left hander, the setup should be reversed.
4. Wash the donor larvae in distilled water, 70% ethanol (for 20–30 s), and BSS. Dissect the appropriate imaginal discs from the larvae with microforceps and tungsten needles (*see* **Note 7**). If one needs to inject fragments of discs, cut discs further at predetermined positions with a fine tungsten needle (Fig. 3, *see* **Note 8**). Collect dissected discs in fresh BSS (*see* **Note 9**).
5. Prepare host flies or larvae for transplantation.

   Host flies:

   Fill the device for anesthesia with diethyl ether and then anesthetize host flies for 20 s (Fig. 4a, *see* **Note 10**). To avoid extensive anesthesia, bring them quickly out onto a tray. Line up 15–20 flies by pasting their wings to the double-sided sticky tape on a glass slide.

   Host larvae:

   Wash 15–20 larvae in distilled water and subsequently in 70% ethanol (within 30 s), and dry them with a sterilized filter

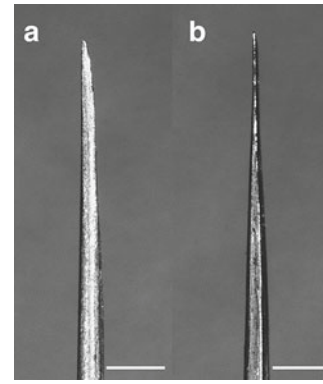


Fig. 3 The tip of tungsten needle for disc fragmentation. The *left* (**a**) and *right* (**b**) panels show the shapes from the lateral and edge side, respectively. The scale bar is 50 μm

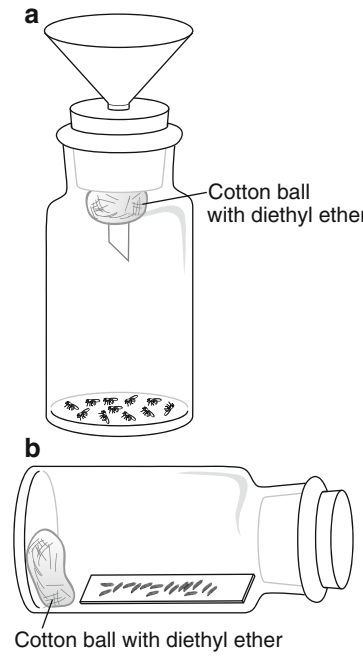


Fig. 4 Device for ether anesthesia. For flies (**a**) and larvae (**b**)

paper. Place them on a glass slide and treat with ether for 2 min (Fig. 4b). Remove the anesthetized larvae and immerse them still wet in a drop of BSS. Line them up on a gelatin-coated glass slide such that their dorsal sides are up and the heads pointing towards the operator's side. Blot excess liquid with a

sterilized filter paper. The larvae will remain stuck on the gelatinized glass slide.

6. Place the slides with the donors and hosts on a plastic plate on the stage of stereomicroscope (*see* **Note 11**).
7. Hold the transplantation needle with the right hand and take up a disc (or a disc fragment) by pulling the syringe plunger with the left hand.
8. While keeping the tip of the transplantation needle at the center of the visual field in the microscope, move the plastic plate with the left hand to find a host to be injected. The head of the host should point toward the operator. The best positions for injections are at the posterior half (abdominal segment A4–A6) of the fly abdomen (Fig. 5a) or one-third from the back (abdominal segment A6–A8) in the larva (Fig. 5b). The cut edge of the transplantation needle should face downward (Fig. 5c).
9. Next use a supporting needle (a sewing needle connected to a needle holder) with the left hand to immobilize the host when the transplantation needle is inserted. As soon as the transplantation needle penetrates the epithelium of the host,

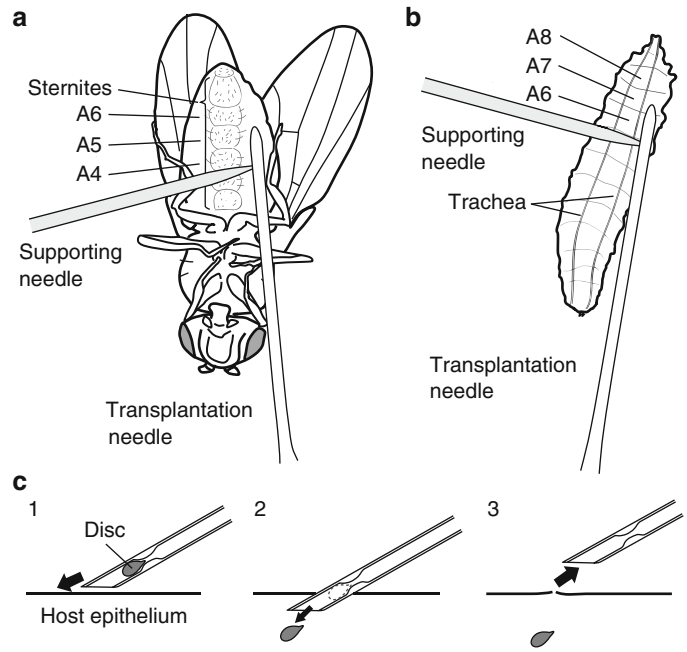


Fig. 5 Schematics of disc transplantation. (**a**, **b**) Injections of discs into adult or larval abdomen are shown. We usually target the injection spot aside of the sternites in the abdominal segments (A) A4–A6 of host flies (**a**) and at the outer side of tracheal tubes in A6–A8 of the larvae (**b**). (**c**) Inject the transplantation needle by turning the opening edge downward (**c***2*). Thereby the injected disc is more likely to be released into the body cavity (**c***2*, **c***3*)

remove the supporting needle and reach with the left hand the syringe plunger to inject the disc into the body cavity. Withdraw the transplantation needle quickly but carefully.

10. To confirm that the disc is injected successfully, expel a drop of BSS beside the injected host. This drop also helps to mark which hosts are injected.
11. Once all the hosts are injected return them to a vial with normal fly medium. For the recovery of host flies, remove each fly from the sticky tape by holding its wings with microforceps. For the larval host recovery, wet larvae with BSS and after a short while transfer them with a brush or microforceps.
12. The injected hosts will recover from anesthesia usually in 10–15 min. We usually keep the injected host female flies in the presence of one or two male flies. Incubating at higher temperature (e.g., 29 °C) requires more attentiveness to humidity.
13. Cleaning of the transplantation apparatus. Remove BSS from the transplantation needle and rinse extensively with Milli-Q grade water. If the needle has a clog at the constriction, a single hair of a brush can be used to remove it. From time to time the needle, holder, and tubing should be cleaned with needle cleaning solution (*see* Subheading 2). A cleaned needle is kept in Milli-Q water in a tube until the next experiment.

3.4 Recovering the Implants from the Host Adult Abdomen

1. Soak CO_2-anesthetized host flies in PBT and rinse in PBS.
2. Dissect a fly in a drop of BSSB on a silicon-coated glass slide. Separate the abdomen by cutting at the thorax–abdomen junction. Open the abdomen by peeling away the abdominal epithelium from the ventral midline at A6 or A7 position toward the anterior. Search the implant, which is usually floating in hemolymph but sometimes attached to other tissues such as gut. The discs transplanted in the host larvae can differentiate to corresponding adult structures together during the host's metamorphosis and stay in their abdomen.
3. Collect the implants (discs or differentiated adult structure) in BSSB and proceed to subsequent experiments. The protocols for immunohistochemical analysis or cuticle analysis, for example, are available from elsewhere [17, 18].

4 Notes

1. One can use a metal plate with a small hole and put it on the Bunsen burner to produce a small flame. Keep the capillary on the flame till the glass appears soft.
2. Crack the tip of the clipped end by pushing with the microforceps. In this case, a microforceps with a blunt tip is better. The cutting surface should be smooth to reduce damage on the host epidermis

when injecting. This procedure requires some practice till many of the manufactured tips fulfill the quality requirements.

3. The proper I.D. of transplantation needle is dependent on the size of the disc to be transplanted. We usually use one with 120–130 μm I.D. for transplanting a leg disc, about 150 μm for an eye-antennal disc, and 170–180 μm for a wing disc. A too narrow needle tip provokes unfavorable damages or stretching to the implant when taken up and released.
4. This constriction is required to keep an implant within the needle tip during the transplantation procedure. Also the diameter of constriction is an important determinant of the speed of liquid flow when a disc is transplanted. Heating up bends the needle toward the filament, hence, pay attention not to bring the needle into contact with the filament.
5. Starvation allows the host flies to have additional space in the abdomen for the injected saline containing the disc transplant.
6. Select a transplantation needle with an appropriate I.D. depending on what discs (or disc fragments) are injected. The constriction size is also important. Check that BSS solution is dripping (not continuously flowing) from the needle tip when the syringe pressure is raised.
7. For regeneration experiments, we usually accomplish disc dissection, fragmentation, and subsequent transplantation all within 1 h. Therefore the number of discs for a transplantation experiment may differ depending on one's skills. All the equipment for disc dissection, including microforceps, tungsten needle, and dissection glass slide, should be sterilized with 70% ethanol.
8. Accurate disc fragmentation in a drop of buffer requires considerable experience. For a beginner it is relatively easy to cut discs on uncoated glass slide compared to silicon-coated one; the thickness of the puddle is reduced on the uncoated glass surface so that one can bring a disc close to the puddle edge and keep it confined during fragmentation. The tungsten needle for disc fragmentation should be sharpened to a katana shape (*see* Fig. 3). This edge allows operator to exert cutting pressure more efficiently upon disc fragmentation compared to the electrically sharpened needle that usually shows a fencing-sword shape.
9. The imaginal discs are sticky to the plastic tips of Pipetteman. In order to avoid loosing them during disc transfer, we usually use a glass capillary with a hand-made connector to Pipetteman. However, the plastic tips can also be used upon lubrication by repeatedly sucking up and down larval fat bodies.
10. The time for ether anesthesia varies depending on the extent of the filled state. The excessive anesthesia will kill the host, while insufficient treatment will not allow keeping the hosts immobilized during a transplantation experiment. Therefore, in order

to accurately anesthetize the host in the same way each time, we recommend filling ether in advance (at least 5 min before use).

11. The plastic plate can facilitate the move between the donor and host slides under the microscope.

References

1. Ephrussi B, Beadle GW (1936) A technique of transplantation for Drosophila. Am Nat 70:218–225
2. Gehring WJ (1978) Imaginal discs: determination. In: Ashburner M, Wright TRF (eds) The genetics and biology of Drosophila, vol 2c. Academic Press, New York
3. Garcia-Bellido A, Nöthiger R (1976) Maintenance of determination by cells of imaginal discs of Drosophila after dissociation and culture in vivo. Wilhelm Roux's Arch Dev Biol 180(3):189–206
4. Steffen PA, Ringrose L (2014) What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat Rev Mol Cell Biol 15(5):340–356
5. Ursprung H (1967) In vivo culture of Drosophila imaginal discs. In: Wilt FH, Wessells NK (eds) Methods in developmental biology. Crowell, New York
6. Bryant PJ (1975) Pattern formation in the imaginal wing disc of Drosophila melanogaster: fate map, regeneration and duplication. J Exp Zool 193(1):49–77
7. Schubiger G (1968) Anlageplan, Deter minationszustand und Transdetermination sleistungen der miinnlichen Vorderbeinscheibe von Drosophila melanogaster. Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen 160(1):9–40
8. Worley MI, Setiawan L, Hariharan IK (2012) Regeneration and transdetermination in Drosophila imaginal discs. Annu Rev Genet 46:289–310
9. Hadorn E (1978) Imaginal discs: transdetermination. In: Ashburner M, Wright TRF (eds) The genetics and biology of Drosophila, vol 2c. Academic Press, New York
10. Lee N, Maurange C, Ringrose L, Paro R (2005) Suppression of Polycomb group proteins by JNK signalling induces transdetermination in Drosophila imaginal discs. Nature 438(7065):234–237
11. Srdic Z, Frei H (1980) Development capacities of mixed normal and neoplastic l(3)gl imaginal discs and neuroblast tissue in Drosophila hydei. Differentiation 17(3):187–192
12. Woodhouse E, Hersperger E, Shearn A (1998) Growth, metastasis, and invasiveness of Drosophila tumors caused by mutations in specific tumor suppressor genes. Dev Genes Evol 207(8):542–550
13. Classen AK, Bunker BD, Harvey KF, Vaccari T, Bilder D (2009) A tumor suppressor activity of Drosophila Polycomb genes mediated by JAK-STAT signaling. Nat Genet 41(10):1150–1155
14. Martinez AM et al (2009) Polyhomeotic has a tumor suppressor activity mediated by repression of Notch signaling. Nat Genet 41(10):1076–1082
15. Sievers C, Comoglio F, Seimiya M, Merdes G, Paro R (2014) A deterministic analysis of genome integrity during neoplastic growth in Drosophila. PLoS One 9(2):e87090
16. Schubiger G (1971) Regeneration, duplication and transdetermination in fragments of the leg disc of Drosophila melanogaster. Dev Biol 26(2):277–295
17. Katsuyama T, Paro R (2013) Innate immune cells are dispensable for regenerative growth of imaginal discs. Mech Dev 130(2–3):112–121
18. Schubiger G, Schubiger M, Sustar A (2012) The three leg imaginal discs of Drosophila: "Vive la difference". Dev Biol 369(1):76–90

Chapter 27

Isolation and Culture of Muscle Stem Cells

Chiara Mozzetta*

Abstract

Polycomb group (PcG) proteins are key epigenetic factors responsible for the proper spatiotemporal repression of defined transcriptional programs along the process of cell differentiation, including myogenesis. The discovery of the pivotal role played by PcG factors during myogenic differentiation relied on the possibility to culture myogenic cells in vitro. We describe here the methods currently used to isolate muscle stem cells (MuSCs) both from single myofibers and from bulk muscles by fluorescence-activated cell sorting (FACS), highlighting experimental details and critical steps. Through these techniques MuSCs can be efficiently isolated and cultured in vitro to recapitulate the different phases of myogenesis: activation, expansion, differentiation, and self-renewal.

Key words Polycomb, Satellite cells, Single myofibers, Muscle differentiation, Muscle stem cells

1 Introduction

Polycomb group (PcG) proteins are key epigenetic factors primarily implicated in the maintenance of gene expression patterns, thus ensuring the cell-specific expression of defined transcriptional programs during the process of cellular differentiation. Among the number of developmental lineages governed by PcG proteins, myogenesis has been one of the best characterized thus far. Pivotal works have demonstrated the crucial role played by Polycomb-repressive complex 2 (PRC2) in the coordination of proper spatiotemporal silencing of muscle-specific genes along myogenic differentiation [1–4], with PRC2 being important to maintain the silencing of muscle-specific genes in undifferentiated myoblasts [1, 4] while also repressing stemness factors (i.e., Pax7) in differentiating myotubes [2, 5]. Moreover, PcG proteins have been shown to preserve the transcriptional identity of skeletal muscle stem cells (MuSCs) by repressing alternative transcriptional programs [3, 6].

Beyond the biological importance of understanding the epigenetic mechanisms governing skeletal muscle differentiation, myogenesis has been undoubtedly one of the best characterized differentiation

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins*: *Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_27,

systems because of the possibility to easily model the different steps of myogenic differentiation in vitro. This has been possible thanks to the availability of immortalized myoblast cell lines (i.e. C2C12 [7, 8]) that can be simply maintained in culture as undifferentiated cells, in the presence of high serum, or induced to differentiate into multinucleated myofibers by serum deprivation, recapitulating what happens in vivo in MuSCs during skeletal muscle regeneration and differentiation. However, even though most of the data on the epigenetic regulation of myogenic differentiation were accumulated over the recent years using this nice and easy experimental model, the next challenge for muscle molecular biologists will be to apply global epigenomic approaches to primary MuSCs with the aim to understand how PcG proteins (and other chromatin modifiers) regulate skeletal muscle differentiation in vivo and if and how these processes are deregulated in pathological conditions (i.e. muscular degenerative disorders).

Here we describe two methods that are currently routinely used by muscle biologists to isolate and culture MuSCs in medium/large scale. The first method entails the isolation by fluorescence-activated cell sorting (FACS) of single cells released by enzymatic digestion from bulk muscles, taking advantage of the MuSC-specific expression of defined cell surface markers, as compared to other muscle-resident progenitor cells (Fig. 1a). In this procedure, MuSCs are labeled with antibodies against MuSC-specific cell surface markers (positive selection) and cell surface markers for non-MuSC populations (negative selection). The most used markers for positive selection comprise a7-integrin, cluster of differentiation protein CD34, transmembrane heparan sulfate proteoglycans syndecan-3 and syndecan-4, chemokine receptor CXCR4, caveolae-forming protein caveolin-1, and vascular cell adhesion molecule 1 VCAM-1 (for review ref. 9). Negative selection is instead most often performed using antibodies specific for other stem cells progenitors (i.e., stem cell antigen 1, Sca-1), hematopoietic (i.e., cluster of differentiation protein CD45), erythroid (i.e., Ter119), and endothelial (i.e., CD31) cells, and for monocytes and granulocytes (i.e., CD11b) (Fig. 1b). FACS-mediated procedure allows the isolation, in a rather abundant scale, of a pure population of MuSCs that can be either analyzed right after sorting or placed in culture for study of proliferation and/or differentiation potentials (Fig. 1c).

The second method is rather based on the enzymatic dissociation of intact muscles, followed by physical trituration and isolation of single myofibers with associated MuSCs [10, 11] (Fig. 2a). Through this technique it is possible to analyze ex vivo, in non-adherent culture, the behavior of single MuSCs and their progeny monitoring the different phases of MuSC-mediated myogenic differentiation: activation, proliferation, differentiation, and self-renewal [12, 13] (Fig. 2b). Single myofibers can also be attached to culture dish so that associated MuSCs can migrate from under the basal lamina and be grown in vitro to have large-scale culture of MuSC-derived myoblasts [14] (Fig. 2c).

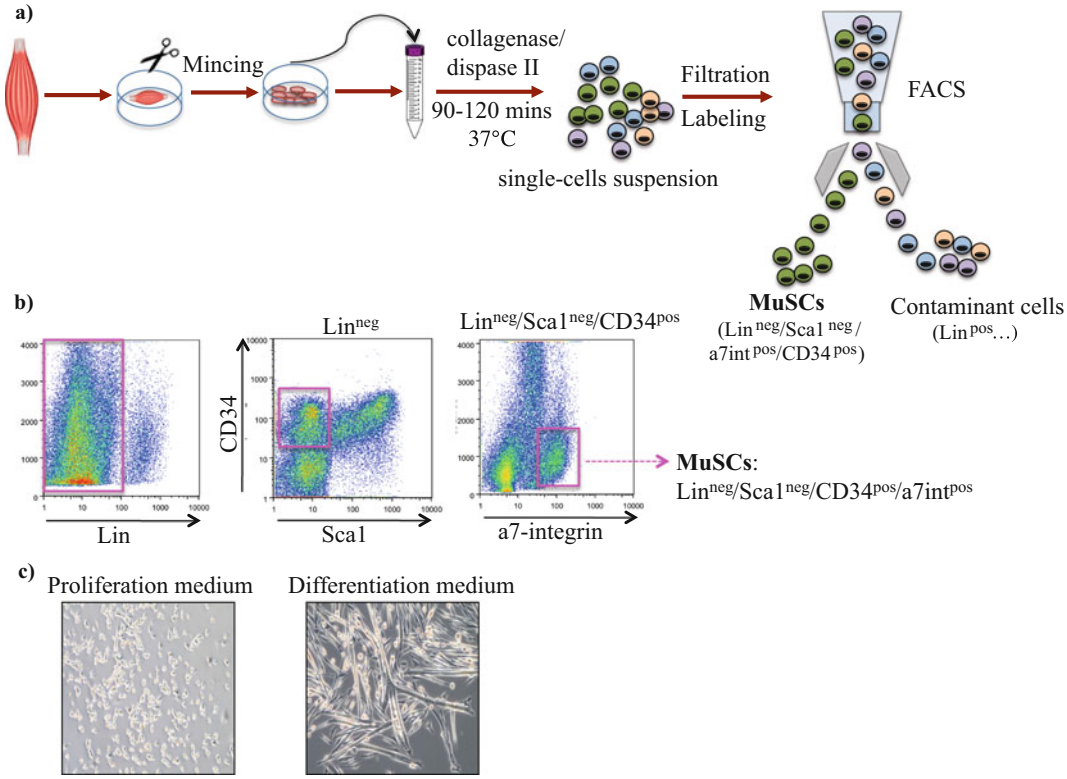


Fig. 1 MuSC isolation by FACS. (**a**) Schematic representation of the experimental steps required to isolate MuSCs by fluorescence-activated cell sorting (FACS). (**b**) Representative FACS plot to sort MuSCs as lineage negative (Lin–: $CD31^-/CD45^-/Ter119^-$) and $Sca1^-/a7integrin^+/CD34^+$. (**c**) Representative phase- contrast images of proliferating (*left*) and differentiating MuSCs (*right*)

2 Materials

1. Water bath at 37 °C.
2. Dissection microscope.
3. Stereo dissection microscope with transmission illumination.
4. Tissue culture hood or lamina flow cabinet.
5. Tissue culture incubator (humidified, 37 °C, 5% CO).
6. Dissection pins.
7. Fine forceps.
8. Fine scissors.
9. Micro-dissecting scissors.
10. Glass Pasteur pipettes (22 cm), sterile.
11. Rubber pipette bulbs.
12. Petri dishes (50 and 35 mm diameter).
13. Plastic pipettes (5, 10, 25 ml volumes), sterile.

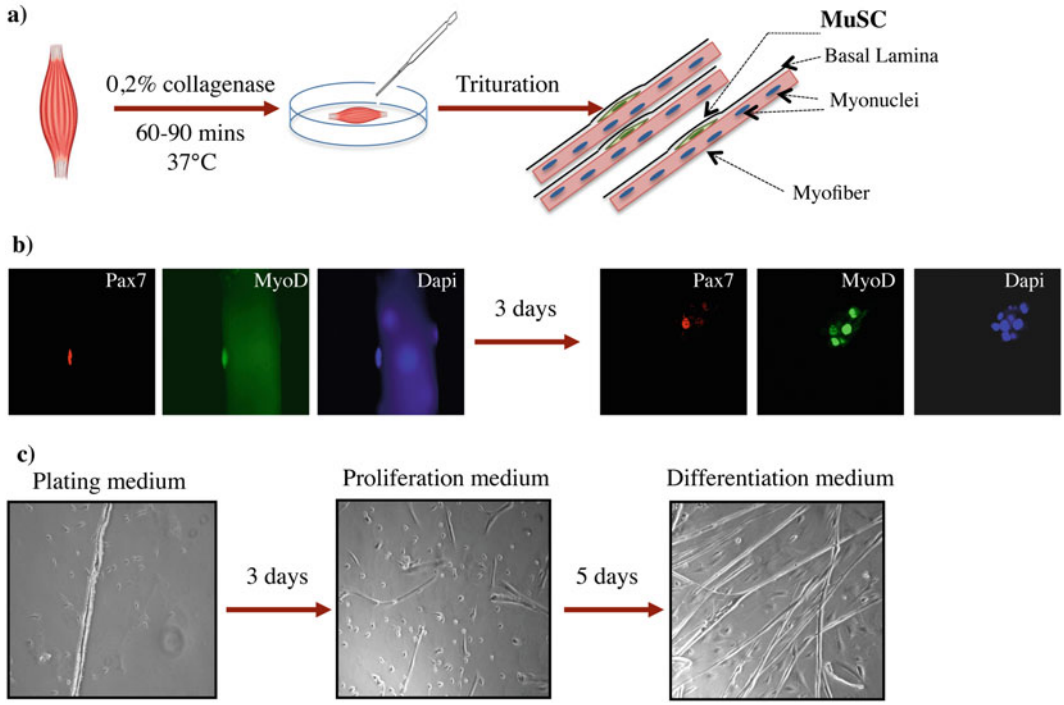


Fig. 2 Single myofiber-derived MuSCs. (**a**) Schematic representation of the procedure to isolate single myofibers. (**b**) Representative images of immunofluorescence for Pax7 (a marker of quiescent and activated MuSCs) and MyoD (a marker of activated and differentiated MuSCs) in myofiber-derived MuSCs either right after isolation (*left panels*) or upon 3 days in non-adherent culture (*right panels*). (**c**) Representative images of myofiber-derived MuSCs grown in adherent culture. Single myofibers are plated on Matrigel-coated dishes in plating medium to allow delamination of associated MuSCs (*left*). Proliferation is then promoted by shifting in proliferation medium for further 5 days (*middle*), after which MuSCs can be induced to differentiate into large myotubes by shifting in differentiation medium (*right*)

14. Syringes 5 and 20 ml, sterile.
15. 0.2 μM Syringe filters.
16. Falcon tubes, 15 and 50 ml.
17. FACS tubes.
18. Cell strainers, 100, 70, and 40 μm pore diameter.
19. Bunsen burner.
20. $CaCl_2$ 50 mM stock concentration, sterile.
21. $MgCl_2$ 1 M stock concentration, sterile.
22. Bovine serum albumin (BSA), sterile.
23. Collagenase A.
24. Dispase II.
25. DNase I.
26. Collagenase from *Clostridium histolyticum*.
27. Penicillin/streptomycin (P/S).
28. Phosphate-buffered solution, PBS (no Ca or Mg).

29. Hanks' balanced salt solution, HBSS (with Ca and Mg).
30. Normal goat serum.
31. Dulbecco's modified Eagle's medium (DMEM), high glucose, GlutaMAX™, Pyruvate.
32. Horse serum (HS).
33. Chick embryo extract (CEE).
34. Basic fibroblast growth factor (bFGF).
35. Fetal bovine serum (FBS).
36. Matrigel.

3 Methods

All steps should be performed under sterile conditions in a tissue culture hood or lamina flow cabinet. Sacrifice the mouse by anesthesia followed by cervical dislocation.

3.1 MuSC Isolation by FACS

3.1.1 Preparation of Solutions and Dishes

1. Prepare a sterile solution of HBSS+ (HBSS with 1% P/S and 0.2% BSA).
2. Prepare the digestion solution as follows:
   - Dispase II (final concentration: 2.4 U/ml) dissolved in PBS 1× (w/o Ca and Mg); this has to be freshly prepared just before muscle harvesting.
   - Collagenase A (final concentration: 2 μg/μl).
   - $CaCl_2$ (final concentration: 400 μM).
   - $MgCl_2$ (final concentration: 5 mM).
   - DNase I (final concentration: 10 μg/ml).
3. In the tissue culture hood, filter-sterilize the digestion solution using a sterile syringe with a 0.2 μM filter.
4. If you plan to grow isolated MuSCs in culture perform Matrigel coating of tissue culture dishes. Prepare 1 mg/ml of Matrigel solution (Matrigel stock aliquots should be defrosted overnight at 4 °C), diluted in DMEM (w/o serum). Cover the dish to be coated with Matrigel solution and remove the excess, leaving only a film of it (*see* **Note 1**). Incubate the plate at 37 °C for at least 1 h. Perform a wash with PBS and then plate the cells.

3.1.2 Skeletal Muscle Harvesting

1. Wash and wipe the mouse with 70% ethanol solution.
2. With scissors cut the skin all around the belly area, and then pull the skin until the ankles and wrists to expose hindlimb and forelimb muscles, respectively.

3. Remove muscles of interest. Unless interested in isolating MuSCs from a particular muscle, you can harvest and pool hindlimb and/or forelimb muscles.
4. As a general guideline, for each muscle, locate and grip one muscle's tendon with fine forceps and gently ease it away from the underlying musculature and bone. Using forceps or scissors to gently disrupt connections cut the muscle at the level of the opposite tendon.
5. Gently tear/cut away any connective or adipose tissue overlying the muscle with forceps or scissors.

3.1.3 Skeletal Muscle Digestion

1. Immediately transfer the harvested muscles into a Petri dish (35 mm diameter) containing 1 ml of HBSS+ and keep it on ice until the mincing step (**step 2**).
2. Mince the collected muscles extensively with sharp microdissecting scissors (*see* **Note 2**).
3. Transfer the minced muscles into a 15 ml Falcon tube.
4. Use 5 ml of HBSS+ to wash the Petri dish and to recover any leftover of minced muscles and transfer into the 15 ml Falcon tube (**step 3**). If necessary repeat this step twice to be sure to recover as much as possible from the Petri dish.
5. Quickly spin down the minced muscles and remove the excess of HBSS+ to have only minced muscles at the bottom of the tube.
6. Add 4 ml of digestion solution (*see* **Note 3**).
7. Incubate in a water bath at 37 °C for 90/120 min, with agitation every 15 min unless you use a shaking water bath (*see* **Note 4**).
8. To stop digestion move the tubes under the culture hood, add 10 ml of HBSS+, and filter the cell suspension through a 100 μm cell strainer into a new 50 ml Falcon tube. Add HBSS+ to a final volume of 30 ml (*see* **Note 5**).
9. Pellet cells by centrifugation at 300 × *g* for 5 min, at 4 °C.
10. Resuspend the pellet in 10 ml HBSS+ and filter the cell suspension through a 70 μm cell strainer into a new 50 ml Falcon tube. Add HBSS+ to a final volume of 30 ml.
11. Pellet cells by centrifugation at 300 × *g* for 5 min, at 4 °C.
12. Resuspend the pellet in 10 ml HBSS+ and filter the cell suspension through a 40 μm cell strainer into a new 50 ml Falcon tube. Add HBSS+ to a final volume of 30 ml.
13. Pellet cells by centrifugation at 300 × *g* for 5 min, at 4 °C.
14. Resuspend the pellet in HBSS++ (HBSS+ with 1% DNase I) and proceed with the staining for FACS isolation (*see* **Note 6**).

3.1.4 MuSC Staining for FACS Analysis and Isolation

1. Block the cell suspension by incubating for 5 min with normal goat serum (diluted 1:10) on ice.
2. Prepare a mix of the antibodies (each used to a final concentration of 10 μg/ml) necessary to isolate MuSCs by FACS (*see* **Note 7**) and dispense it into each sample to be stained and analyzed by FACS (*see* **Note 8**).
3. Incubate for 30 min on ice.
4. Add HBSS+ to fill the entire tube and centrifuge at 300 × *g* for 5 min, at 4 °C.
5. Remove supernatant and add further HBSS+ to wash cells. Centrifuge at 300 × *g* for 5 min, at 4 °C.
6. Resuspend pellet in 100/300 μl of HBSS+ (*see* **Note 9**) and proceed to FACS analysis.
7. Isolated MuSCs can be either analyzed right after sorting or placed in culture on Matrigel-coated dishes.
8. To maintain MuSCs in proliferation, grow the cells in DMEM + 20% FBS + 10% HS + 2.5 ng/ml bFGF. To induce differentiation, change the medium to DMEM+2% HS.

3.2 MuSCs from Single Myofibers

3.2.1 Preparation of Dishes

1. Prepare DMEM + 10% HS + 1% P/S (DMEM/10% HS).
2. Prepare plating medium: composed of DMEM + 10% HS + 0.5% CEE + 1% P/S.
3. To prevent fibers from sticking to the dishes or pipettes, rinse Petri dishes and pipettes with DMEM/10% HS.
4. Prepare 3–4 50 mm by 18 mm Petri dishes per muscle and add 8 ml of DMEM/10% HS. Place dishes in the incubator at 37 °C for at least 30 min to allow the DMEM to warm.
5. Prepare glass Pasteur pipettes, needed for muscle trituration. Score around the Pasteur pipette using a diamond pen and snap away the end to create pipettes with diameters at the mouth of approximately 1, 4, and 6 mm.
6. Heat polish the cut ends of the Pasteur pipettes using a Bunsen burner to melt the glass at the mouth to remove the sharp and jagged edges (*see* **Note 10**).
7. If you plan to plate myofibers for delamination of associated MuSCs perform Matrigel coating of tissue culture dishes as described in Subheading 3.1.1, **step 4**.
8. Immediately before dissection, freshly prepare a solution of 0.2% collagenase (from *Clostridium histolyticum*), diluted in DMEM + 1% P/S (w/o serum), and sterilize it by filtering it using a sterile syringe with a 0.2 μM filter under the tissue culture hood (*see* **Note 11**).

3.2.2 Skeletal Muscle Harvesting

1. Isolate skeletal muscles of interest (*see* **Note 12**) as described in Subheading 3.1.2. For single myofibers isolation is extremely important to handle the muscles from the distal tendons (*see* **Note 13**). Immediately transfer the liberated muscles to a pre-warmed tube of 0.2% collagenase/DMEM (*see* **Note 14**).

3.2.3 Skeletal Muscle Digestion

1. Incubate the muscles in 0.2% collagenase/DMEM in a shaking water bath for 60–90 min at 37 °C. As a general guide, muscles from a 6- to 12-week-old mouse require approximately 90 min of digestion (*see* **Note 15**).

3.2.4 Skeletal Muscle Trituration and Myofiber Isolation

1. When digestion is complete, remove the tubes containing the muscle from the water bath, wipe dry, and then wipe with 70% ethanol before placing in the culture hood. Also place one deep Petri dish containing DMEM/10% HS per muscle in the culture hood.
2. Use the largest diameter heat-polished glass Pasteur pipette with a rubber pipette bulb and coat the inside with DMEM/10% HS just before use to prevent myofibers from sticking. Carefully remove the muscle from the tube, place it into a deep Petri dish containing DMEM/10% HS, and then put back the Petri dish into the incubator for 20–30 min.
3. Place the stereo-dissecting microscope directly in the tissue culture hood, together with the deep Petri dish containing the digested muscle and a second containing pre-warmed DMEM/10% HS.
4. Using the largest diameter pre-coated heat-polished Pasteur pipette, take up the muscle and triturate repeatedly. This procedure will liberate highly refractive, hairlike fibers.
5. Continue the trituration until the muscle becomes clearly fragmented. Take up a small-diameter Pasteur pipette and pre-coat with DMEM/10% HS. Carefully collect intact liberated fibers and place them in a fresh, pre-warmed dish of DMEM/10% HS, being careful to avoid collecting hypercontracted myofibers and debris.
6. Depending on the amount of needed myofibers, at this point it may be useful to move the remaining muscle to another deep Petri dish and continue to triturate to isolate myofibers using a fire-polished, coated, Pasteur pipette with a slightly smaller aperture. Using the fine-aperture Pasteur pipette continue to collect liberated myofibers (*see* **Note 16**).
7. Serially transfer isolated myofibers through 1–2 clean and pre-warmed deep Petri dishes of DMEM/10% HS to ensure that any contaminating cells are removed (*see* **Note 17**).
8. At this point fibers can be either used for analysis of quiescent MuSCs by fixation and immunostaining (this procedure is detailed in ref. 11).

3.2.5 Myofiber Non-adherent Culture

Myofibers can be cultured in non-adherent culture in vitro to allow associated MuSCs to activate, proliferate, differentiate, and self-renew.

1. Under the microscope, transfer clean myofibers into a DMEM/10% HS-coated deep Petri dish containing plating medium composed of DMEM + 10% HS + 0.5% CEE. Myofibers can be maintained in non-adherent cultures up to 96–120 h, depending on the experimental needs.

3.2.6 Myofiber-Derived MuSC Adherent Culture

Myofibers can also be plated to allow their associated MuSCs to delaminate and attach onto the tissue culture substrate. With this procedure large quantities of MuSC-derived myoblasts can be obtained and studied up to the stage of fusion into large multinucleated myotubes.

1. Under the microscope, transfer the needed amount of clean myofibers into a Matrigel-coated dish. Allow the fibers to settle and attach to Matrigel for 3–5 min, without any media, and then add the lowest possible amount of plating medium (DMEM + 10% HS + 0.5% CEE +1% P/S) (*see* **Note 18**).
2. The day after, add the entire volume of plating medium extremely carefully and slowly, to avoid fibers to detach from the plate.
3. 3–4 days following plating there will be 50–300 MuSCs surrounding fiber; to promote proliferation the medium should be switched (*see* **Note 19**) to DMEM + 20% FBS + 10% HS + 1% CEE + 1% P/S, upon removal of attached fibers (*see* **Note 20**).
4. Grow the cells to the desired confluency. At 75% confluency MuSCs should be splitted for sub-culture. MuSCs can be passaged 3–4 times and will remain in a proliferative state for up to 1 week, but will then begin to differentiate.
5. Differentiation can be induced by switching proliferation medium to differentiation medium (DMEM + 2% HS + 0.5% CEE + 1% P/S).

4 Notes

1. This procedure should be performed on ice and with cooled pipettes and media, to prevent Matrigel from gelification.
2. The yield of MuSCs recovery highly depends on the efficiency of mincing. Triturated muscles should appear as a mush with only very small pieces of muscles left.
3. Use no more than 2.5 g of muscle for 4 ml of digestion solution; otherwise it will become too viscous.
4. Check the digestion carefully after 60 min; if fully dissociated, stop digestion. The time of digestion should be determined

empirically as it does depend on the amount of muscles. A fully dissociated cell suspension does not contain any visible piece of muscle; it appears as a liquid, though a bit viscous, solution.

5. Reach the final volume of 30 ml of HBSS+ by adding 10 ml per time into the digestion tube, to recover any leftover of the cell suspension, and pass it through the filter.
6. We generally perform staining for FACS in a final volume of 500 μl. Thus, considering the volume of goat serum (*see* Subheading 3.1.4, **step 1**) and of antibodies (*see* Subheading 3.1.4, **step 2**) we generally resuspend cells at this step in a volume = 500 μl—volume of goat serum (*see* Subheading 3.1.4, **step 1**)—total volume of antibodies (*see* Subheading 3.1.4, **step 2**). We then transfer the cell suspension directly into FACS tubes, whose characteristics depend on the instrument that will be used and should be chosen accordingly.
7. We generally isolate MuSCs as $CD45^-/CD31^-/Ter119^-/CD11b^-/Sca1^-$ (lineage negative, Lin^-) and $a7\text{-integrin}^+/CD34^+$ (as in [15]). However, other MuSC-specific markers can be used, as highlighted in the introduction. To ease FACS procedure, we advise to use all the antibodies for the negative selection (Lin−), conjugated with the same fluorophore, as such only one channel will be used to gate Lin-negative cells. The fluorophores to which antibodies are conjugated should be chosen to be compatible to and visible by the set of lasers of the FACS instrument that will be used.
8. It is extremely important to save some cells to perform the proper controls for FACS analysis, that is, cells unstained, cells stained with each single antibody and fluorophore, and cells with all the fluorophores except the one that is being measured (fluorescence minus one, FMO).
9. Cells should be ideally resuspended to a final concentration of 1×10^7 cells/ml.
10. To test whether the edges are smooth, circle on a piece of aluminum foil; if the foil tears, it is not smooth enough and will damage fibers. Reheat carefully to sterilize and store in the tissue culture hood [11].
11. 3–5 ml per muscle is generally sufficient. It is important that the muscle is completely immersed in the digestion solution.
12. Single myofibers are generally and more efficiently isolated from tibialis anterior (TA), extensor digitorum longus (EDL), and soleus (for a detailed review on how to properly harvest muscles for myofiber isolation refer to ref. 11).
13. Never touch the muscles directly, and only manipulate it via the distal tendons, as this is critical to prevent myofiber contraction and damage.

14. Muscles can be stored in DMEM (without collagenase) for transportation, but no longer than ~90 min.
15. The precise time of digestion depends upon both the age and size of the mouse and the activity of the batch of collagenase used, and should be determined empirically. Digestion is complete when the muscle looks less defined and slightly swollen, and under the microscope, hairlike single fibers are seen from the edge of the muscle.
16. It is extremely important not to allow the digested muscles and isolated myofibers to cool too much. Dishes should be returned to the incubator every 20 min for a minimum of 10 min. Switch between muscles every 15–30 min.
17. Ensure that fibers are free of vascular and connective tissue debris, as this will decrease myogenic purity. If necessary increase the number of passages through clean Petri dishes.
18. For example, add only 0.5 ml in 35 mm dish or 1 ml in a 50 mm dish. This step will increase the chance that fibers attach to the Matrigel-coated dish and will increase the yield of delaminating MuSCs.
19. If plating medium is maintained for longer periods MuSCs will start to differentiate.
20. At this time the fiber can be easily removed using a Pasteur pipette without disturbing the satellite cells. Removal at a later point is more difficult due to stronger binding with Matrigel.

References

1. Caretti G, Di Padova M, Micales B, Lyons GE, Sartorelli V (2004) The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes Dev 18:2627–2638
2. Palacios D, Mozzetta C, Consalvi S, Caretti G, Saccone V, Proserpio V et al (2010) TNF/p38α/polycomb signaling to Pax7 locus in satellite cells links inflammation to the epigenetic control of muscle regeneration. Cell Stem Cell 7:455–469
3. Asp P, Blum R, Vethantham V, Parisi F, Micsinai M, Cheng J et al (2011) Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. Proc Natl Acad Sci U S A 108:E149–E158
4. Stojic L, Jasencakova Z, Prezioso C, Stutzer A, Bodega B, Pasini D et al (2011) Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells. Epigenetics Chromatin 4:16
5. Mozzetta C, Consalvi S, Saccone V, Forcales SV, Puri PL, Palacios D (2011) Selective control of Pax7 expression by TNF-activated p38α/polycomb repressive complex 2 (PRC2) signaling during muscle satellite cell differentiation. Cell Cycle 10:191–198
6. Juan AH, Derfoul A, Feng X, Ryall JG, Dell'Orso S, Pasut A et al (2011) Polycomb EZH2 controls self-renewal and safeguards the transcriptional identity of skeletal muscle stem cells. Genes Dev 25:789–794
7. Yaffe D, Saxel O (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature 270:725–727
8. Blau HM, Pavlath GK, Hardeman EC, Chiu CP, Silberstein L, Webster SG et al (1985) Plasticity of the differentiated state. Science 230:758–766
9. Yin H, Price F, Rudnicki MA (2013) Satellite cells and the muscle stem cell niche. Physiol Rev 93:23–67
10. Rosenblatt JD, Lunt AI, Parry DJ, Partridge TA (1995) Culturing satellite cells from living single muscle fiber explants. In Vitro Cell Dev Biol Anim 31:773–779

11. Moyle LA, Zammit PS (2014) Isolation, culture and immunostaining of skeletal muscle fibres to study myogenic progression in satellite cells. Methods Mol Biol 1210:63–78
12. Beauchamp JR, Heslop L, Yu DS, Tajbakhsh S, Kelly RG, Wernig A et al (2000) Expression of CD34 and Myf5 defines the majority of quiescent adult skeletal muscle satellite cells. J Cell Biol 151:1221–1234
13. Zammit PS, Golding JP, Nagata Y, Hudon V, Partridge TA, Beauchamp JR (2004) Muscle satellite cells adopt divergent fates: a mechanism for self-renewal? J Cell Biol 166: 347–357
14. Rosenblatt JD, Parry DJ, Partridge TA (1996) Phenotype of adult mouse muscle myoblasts reflects their fiber type of origin. Differentiation 60:39–45
15. Sacco A, Doyonnas R, Kraft P, Vitorovic S, Blau HM (2008) Self-renewal and expansion of single transplanted muscle stem cells. Nature 456:502–506

ERRATUM TO

Chromatin Preparation and Chromatin Immuno-precipitation from Drosophila Embryos

Eva Löser, Daniel Latreille, and Nicola Iovino

Chiara Lanzuolo and Beatrice Bodega (eds.), Polycomb Group Proteins: Methods and Protocols, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5_3, © Springer Science+Business Media New York 2016

DOI 10.1007/978-1-4939-6380-5_28

There was a typo in the value of Glycine in Chapter 3, under the sub-heading 2.2, point #4 in the initial version published online and in print. It was printed as "1 M of Glycine" but the actual value should be "2.5 M of Glycine".

The updated original online version for this chapter can be found at 10.1007/978-1-4939-6380-5_3

Index

A

Acceptor photobleaching....148
Active contour model....181–196
Adapter clipping....42
Algorithm....48, 49, 74, 82, 83, 126, 132, 141, 182–187, 190, 192, 247, 251, 252, 258, 260, 287
Alignment....41, 47, 48, 129, 255
Antibodies....10–12, 19, 20, 26, 38, 70, 71, 78–80, 83, 87, 103, 113, 156, 162, 163, 170, 171, 175, 188, 289, 312, 317, 320
Automated imaging analysis....141

B

Beads buffer....90, 95
Bioinformatics....38, 70, 128, 225, 235
Biotin....99, 100
  labeling....100, 106, 112
B lymphoma Mo-MLV insertion region 1 (Bmi1)....4, 37, 39, 171, 176, 188, 189, 191
Bromodeoxyuridine triphosphate (BrdU)....5, 58, 61, 63

C

Capture hybridization analysis of RNA targets (Chart)....71, 87, 88, 90–96, 115
Capture oligonucleotides (CO)....88, 92–94
Cellular differentiation....139, 311
CFP-tagged protein....148
Chromatin....3–5, 7, 9–12, 15–16, 18–20, 23–35, 37, 38, 69, 71, 79, 84, 87, 92, 115, 122, 126, 127, 141, 168, 178, 201–205, 223–239, 291
  fractionation....140, 167–178
Chromatin immunoprecipitation (ChIP)....4, 5, 7–20, 38, 115, 127
Chromatin immunoprecipitation sequencing (ChIP-seq)....12, 27, 30, 37–50, 127
Chromatin isolation by RNA purification (ChIRP)....71, 115–122
Chromosome conformation capture (3C)....207–212, 223, 243, 244
Chromosome Conformation Capture on Chip (4C)....243–260
Circular Chromosome Conformation Capture (4C-seq)....224, 245
Classification technique....181–196
Clustered regularly interspaced short palindromic repeats (CRISPRs)....266
Competition analysis....149
Confocal microscopy....287
CpG islands....38
CpG-rich promoter elements....4
Cross-linking/cross-linked....9, 10, 12, 13, 15–16, 18, 19, 23, 24, 26, 29, 35, 38, 40, 74–76, 78–79, 88, 90–93, 140, 216, 219, 224–226, 228, 230–232, 236, 239
  agent....70, 75, 76, 219
  reversal....96, 122
Cufflinks....83, 126, 128, 130–133

D

Data processing....37–50, 243–260
Deep-sequencing analysis....33–34
Denaturant buffer....90, 95
Development....4, 34, 38, 71, 113, 139, 153, 168, 213, 253, 266, 267, 284, 301
Differential expression....131
Differentiation....140, 153, 159, 160, 168, 169, 172, 173, 182, 265, 283, 311, 312, 314, 317
DNA–DNA interactions....208
DNA-protein interactions....19, 60
DNA sonication....57, 59–60
Donor larva....301, 304, 305
Double-strand breaks (DSBs)....269
Drosophila abdomen....266, 301, 302
Drosophila embryonic S2 cells....169, 171
Drosophila embryos....5, 23–35, 246, 267, 283–288
Drosophila melanogaster....167

E

Early S-phase....56, 57, 60–63
Egg-laying....28, 34
Embryonic stem cells (ESCs)....140, 153–164
Enhancer of zeste (E(z))....13, 38
Epigenetic factors....153, 311
Epimorphic regeneration....301
Extra sex combs (ESCs)....38, 202

Chiara Lanzuolo and Beatrice Bodega (eds.), *Polycomb Group Proteins: Methods and Protocols*, Methods in Molecular Biology, vol. 1480, DOI 10.1007/978-1-4939-6380-5, © Springer Science+Business Media New York 2016

F

Fluorescence activated cell sorting (FACS) 58, 59, 61, 265, 297, 298, 312–317, 320
Fluorescence in situ hybridization (FISH) 118, 141, 203, 205, 207, 208, 298
Fluorescence resonance energy transfer (FRET) 140, 141, 143, 144, 147–151
  acceptor 148, 149
  donor 149
Formaldehyde 4, 7, 9, 13–15, 18, 20, 23, 25, 29, 30, 32, 34, 35, 38, 75–77, 79, 84, 88, 89, 91–93, 96, 117, 118, 121, 122, 208, 209, 213, 215, 217–219, 224–226, 232, 239

G

Glycerol buffer 90, 91
Green fluorescence protein (GFP) 208, 283, 287, 290, 299

H

H2AK119ub 39, 153
H3K4me3 ..... 69, 127
H3K27me3 ..... 5, 26, 33, 37–41, 49, 223, 224, 244
H3K36me3 ..... 69, 127
Heatmap 45–47, 50, 130
Heterokaryon 266, 290, 291, 293–299
Higher order chromatin structure 207
High resolution capture (Hi-C) 201–203, 205, 254
High-throughput microscopy images 181–196
Histone modification 10, 23, 37–50, 167
Homologous recombination (HR) 266, 280
Human B lymphocytes 293
Human Genome Nomenclature Committee (HGNC) 134
Hybridization buffer 90, 95, 117, 119

I

Imaginal discs 13, 208, 209, 211, 266, 270, 301–310
Immunofluorescence (IF) 20, 141, 291, 298, 314
Immunoprecipitation (Ip) 7, 10–12, 15–17, 19, 24, 26, 27, 33, 38, 58–60, 74–78, 80, 83, 84, 140, 155
Induction of pluripotent stem cells (iPS) 266
Insulator proteins 223
Intraspecies hybrids 290, 296
In vitro transcription 99, 100, 102, 106, 111
ISODATA 185, 186

J

Jun N-terminal kinase (JNK) 270

L

Labeling 58, 106, 107
Lac operator (laco) 208
Lac repressor (LacI) 208
Lamin 167–169, 176
Late S-phase 61, 63
Ligation reaction 213, 219, 280
Live microscopy 283, 284
Long noncoding RNAs (lncRNAs) 69–71, 73–75, 115, 116, 122, 125–127, 131–134

M

Magnetic bead 20, 78–80, 83, 89, 100, 117
Mapping 3–5, 10, 12, 27, 42–43, 47, 48, 83, 88, 91, 118, 126, 129, 131, 132, 235, 250
Massively parallel sequencing 38
Matrigel-coated dishes 314, 317
Mean square change (M.S.C.) 283, 285, 288
Mean square displacement (M.S.D.) 283, 285, 288
Medium containing retinoic acid (ATRA) 158, 160
Metamorphosis 301, 308
Methyl transferase activity 153
Microarray processing 82
MicroRNAs (miRNAs) 125
Mid S-phase 56
Mitochondrial sequences 61, 63
Mouse skeletal myoblasts C2C12 169
Muscle differentiation 141, 168, 311, 312
Muscle stem cells (MuSCs) 187, 188, 265, 311–321
Myotubes 169, 173, 176, 177, 311, 314, 319

N

Native RIP 76–78
Next-generation sequencing (NGS) 4, 5, 38, 75, 76, 82, 127, 128
Non-canonical PRC1 (ncPRC1) 153
Noncoding RNAs (ncRNAs) 69–71, 73, 74, 82, 87, 99, 125–134
Non-homologous end-joining (NHEJ) 270, 271, 279–280
Nuclear architecture 167
Nuclear compartmentalization 168
Nuclear matrix 140, 168, 174, 178
Nuclei rinse buffer 90

O

Oases 126

P

Paraformaldehyde (PFA) 29, 147, 148
PcG RING fingers (PcGFs) 154
Photobleaching 148, 149, 151
Piwi-interacting RNAs (piRNAs) 125
Pluripotent stem cells 266, 289
Pol II occupancy 127
Polyacrylamide gel electrophoresis 103

Polycomb (Pc) 3–5, 9, 33, 37, 100, 139–141, 167–178, 223, 302
Polycomb Group proteins (PcG)
bodies..... 168, 182, 183, 185–189, 191–196
dynamics..... 139
protein 3–5, 12, 13, 18, 37–50, 139–141, 143, 168, 169, 176, 178, 182, 189, 223, 245, 266, 267, 302, 311, 312
silencing..... 270, 302
Polycomb-repressive complex 1 (PRC1) 37, 38, 41, 140, 153, 154, 162, 163, 171
Polycomb repressive complex 2 (PRC2)..... 37, 38, 41, 74, 75, 153, 171, 244, 311
Polycomb response element (PRE) 3, 33, 38, 208, 214, 245, 253
Polyethylene glycol (PEG) 107, 112, 266, 290, 291, 293, 294, 296–298
Polyhomeotic (Ph)..... 37
Polymerase chain reaction (PCR) analysis..... 17, 83
Polymer physics 201–205
Posterior sex combs (Psc)..... 37
Preprocessing..... 128, 189–190, 192, 193
Probe design 71, 115, 118
Promoter..... 3, 13, 33, 69, 100, 101, 104, 122, 126, 270
Propidium iodide (PI) 57, 59, 62, 299
Proteinase K..... 14, 15, 17, 20, 26, 30, 32, 57–60, 77, 80, 89, 95, 96, 117, 170, 175, 210, 215, 217, 227, 230, 231
Protein-RNA interaction..... 74, 99
Pull-down..... 71, 99–113

Q

Quenching buffer 90, 93

R

Read counts 129, 130
Recruitment..... 3, 4, 38, 41, 100, 154
Reference genome 42, 47, 48, 82, 129, 235
Reference indexing 129
Repeat variable diresidue (RVD)..... 269, 271
Replication timing..... 5
Repressor 3, 10, 265
Reprogramming 265, 266, 289–299
RNA antisense purification (RAP)..... 71, 115
RNA-binding proteins (RBPs)..... 74, 76
RNA, genomic DNA 23, 88, 96, 104, 111, 115, 177, 216, 247, 279, 281
RNA immunoprecipitation (RIP) 70, 73–84
RNA purification 80–81, 89, 94
RNase A 26, 30, 59, 117, 119, 120, 122, 170, 175, 215
RNase H..... 88–96, 117, 120

S

Satellite cells 321
Scripture 126, 132
SDS-PAGE..... 84, 156, 163, 164, 175
Segmentation..... 141, 181–196, 245, 251–253, 258–260, 285, 287, 288
Sex combs extra (Sce/RING) 37
Short interfering RNAs (siRNAs)..... 125
Silver staining 100, 103, 109, 110, 156, 164
Single myofiber 265, 312, 314, 317–320
Skeletal muscles..... 311, 315–316, 318
Small nucleolar RNAs (snoRNAs)..... 125
SOAPdenovo..... 126
Somatic cell nuclear transfer (SCNT) 266, 289
Sonication buffer 90, 92, 93
Sonicator 14, 15, 18, 19, 58, 62, 89, 95, 96, 104, 112
Strings & binders switch (SBS) model..... 203
Sucrose buffer 90, 91
Suppressor of zeste 12 (Su(z)12) 38

T

Thresholding techniques 182
Tiling array probes 249, 250, 257, 259
Tissue regeneration 302
Topological associated domains (TADs) 201, 204, 205, 223
Transcription 3, 4, 38, 45, 46, 69, 70, 75, 81, 94, 102, 120, 131, 207, 266, 269
Transcription activator-like effector nucleases (TALENs) 266, 269–281
Transcription factor (TF) 4, 10, 38, 167, 270, 289
Transplantation..... 266, 301–310
Trimming 129
Trinity..... 126, 132

U

Unpaired (upd)..... 270

V

Variational models 183, 195, 253

W

Wash buffer..... 32, 93, 108, 117, 119, 162, 227, 233, 234
Wingless (wg) genes..... 270

X

XIST 100, 110, 111

Y

YFP-tagged protein..... 148