Methods in Molecular Biology 1455

Springer Protocols

Attila Németh *Editor*

The Nucleolus

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

The Nucleolus

Methods and Protocols

Edited by

Attila Németh

Department of Biochemistry III, University of Regensburg, Regensburg, Germany

Humana Press

Editor
Attila Németh
Department of Biochemistry III
University of Regensburg
Regensburg, Germany

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-3790-5 ISBN 978-1-4939-3792-9 (eBook)
DOI 10.1007/978-1-4939-3792-9

Library of Congress Control Number: 2016943074

The cover image was taken in Angus Lamond's laboratory by Saskia Hutten

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

The nucleolus is a prominent structural and functional compartment of the cell nucleus. The main function of the nucleolus is the transcription and processing of ribosomal RNA and the first steps of ribosome assembly; however, it is also involved in numerous other cellular activities, comprising stress response, RNA metabolism, and cell cycle regulation. This volume in the Methods in Molecular Biology series aims at providing an up-to-date compilation of current methodological approaches utilized for the exploration of nucleolar structure and function.

The book is divided into three parts: (I) Imaging of the Nucleolus, (II) Analysis of Ribosomal RNA Transcription and Processing, and (III) Genomics and Proteomics of the Nucleolus.

Part I brings together various techniques to visualize nucleolar structure and function in different model organisms. The methodologies range from immunofluorescence and fluorescence in situ hybridization approaches through correlative light and electron microscopy to live-cell imaging and use *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Spodoptera frugiperda*, *Danio rerio*, *Rattus norvegicus*, *Mus musculus*, and *Homo sapiens* cells as model systems.

Part II includes a collection of protocols, which were developed to analyze the basic task of the nucleolus, the synthesis of ribosomal RNA in *S. cerevisiae* and in mammals. Since the production of ribosomal proteins and several steps of the ribosome assembly take place outside of the nucleolus, this part is concerned exclusively with methods to analyze ribosomal RNA transcription and processing. The protocols describe in vitro transcription assays, purification of RNA polymerase I, furthermore quantitative analyses of ribosomal RNA synthesis and processing by in vivo labeling and northern blot techniques.

Part III focuses primarily on high-throughput wet laboratory and computational experimental approaches, which deliver essential data for the system-level understanding and integrative analysis of different nucleolar functions. The chapters cover the following topics: ribosomal DNA sequence assembly in eukaryotic genomes, investigation of ribosomal DNA methylation and fluorescence-activated nucleolus sorting in *Arabidopsis thaliana*, purification of RNA polymerase I-associated chromatin from *S. cerevisiae* for "omics" studies, genomics of ribosomal DNA chromatin, and RNomics and proteomics of the human nucleolus.

The diversity of the protocols mirrors the multifaceted biology of the nucleolus, and the book should provide biochemists, molecular biologists, cell biologists, and bioinformaticians with a reliable practical handbook to facilitate the investigation of this nuclear compartment at the advanced level.

I had the distinct honor to work on the compilation of this book with 43 authors, the Series Editor, Professor John M. Walker, and David Casey at Humana Press, Springer, and would like to express my gratitude for their great contribution.

Regensburg, Germany — *Attila Németh*

Contents

Contributors

Saumya Agrawal • *Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand; School of Biological Sciences, University of Auckland, Auckland, New Zealand*

Baoyan Bai • *Department of Immunology, Institute for Cancer Research, Oslo University Hospital, Oslo, and Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway*

Dalila Bensaddek • *Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee, UK*

Maxime Berthaud • *CRCNA-UMR 892 INSERM, Nantes, France; CNRS, Nantes, France; University of Nantes, Nantes, France*

Myriam Boyer-Clavel • *Institut de Génétique Moléculaire de Montpellier, BioCampus Montpellier, Montpellier, France*

Astrid Bruckmann • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie I, Universität Regensburg, Regensburg, Germany*

Kaspar Burger • *Sir William Dunn School of Pathology, University of Oxford, Oxford, UK*

M. Cristina Cardoso • *Department of Biology, Technische Universität Darmstadt, Darmstadt, Germany*

Stefan Dillinger • *Biochemistry Center Regensburg, University of Regensburg, Regensburg, Germany*

Dirk Eick • *Department of Molecular Epigenetics, Helmholtz Center Munich and Center for Integrated Protein Science Munich (CIPSM), Munich, Germany*

Christoph Engel • *Department of Molecular Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany*

Olivier Gadal • *Laboratoire de Biologie Moléculaire Eucaryote, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, Toulouse, France*

Austen R.D. Ganley • *Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand; School of Biological Sciences, University of Auckland, Auckland, New Zealand*

Joachim Griesenbeck • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie III, Universität Regensburg, Regensburg, Germany*

Henry D. Herce • *Department of Biology, Technische Universität Darmstadt, Darmstadt, Germany*

Marikki Laiho • *Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA; Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

Angus I. Lamond • *Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee, UK*

Christian Lanctôt • *Institute of Cellular Biology and Pathology, First Faculty of Medicine, Charles University in Prague, Prague, Czech Republic*

Gernot Längst • *Biochemistry III, University of Regensburg, Regensburg, Germany*

Isabelle Léger-Silvestre • *Laboratoire de Biologie Moléculaire du CNRS, University of Toulouse, Toulouse, France*

Jan Linnemann • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie I, Universität Regensburg, Regensburg, Germany*
Anne K. Ludwig • *Department of Biology, Technische Universität Darmstadt, Darmstadt, Germany*
Robert M. Martin • *Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal*
Brian McStay • *Centre for Chromosome Biology, National University of Ireland, Galway, Ireland*
Philipp E. Merkl • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie III, Universität Regensburg, Regensburg, Germany; Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA*
Philipp Milkereit • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie III, Universität Regensburg, Regensburg, Germany*
Gireesha Mohannath • *Department of Biology, Indiana University, Bloomington, IN, USA; Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN, USA*
Tom Moss • *Laboratory of Growth and Development, St-Patrick Research Group in Basic Oncology, Cancer Division of the Quebec University Hospital Research Centre, Québec, Canada; Department of Molecular Biology, Medical Biochemistry and Pathology, Faculty of Medicine, Laval University, Québec, Canada*
Attila Németh • *Department of Biochemistry III, University of Regensburg, Regensburg, Germany*
Armel Nicolas • *Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee, UK*
Christophe Normand • *Laboratoire de Biologie Moléculaire Eucaryote, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, Toulouse, France*
Jorge Perez-Fernandez • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie I, Universität Regensburg, Regensburg, Germany*
Dimitri G. Pestov • *Department of Cell Biology, Rowan University School of Osteopathic Medicine, Stratford, NJ, USA*
Craig S. Pikaard • *Department of Biology, Indiana University, Bloomington, IN, USA; Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN, USA; Howard Hughes Medical Institute, Indiana University, Bloomington, IN, USA*
Michael Pilsl • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie III, Universität Regensburg, Regensburg, Germany*
Frédéric Pontvianne • *CNRS, Laboratoire Génome et Développement des Plantes, Perpignan, France; University Perpignan Via Domitia, Laboratoire Génome et Développement des Plantes, Perpignan, France*
Julio Sáez-Vásquez • *CNRS, Laboratoire Génome et Développement des Plantes, Perpignan, France; University Perpignan Via Domitia, Laboratoire Génome et Développement des Plantes, Perpignan, France*
Uwe Schwartz • *Biochemistry III, University of Regensburg, Regensburg, Germany*
Marjolein van Sluis • *Centre for Chromosome Biology, National University of Ireland, Galway, Ireland*

Viktor Y. Stefanovsky • *Laboratory of Growth and Development, St-Patrick Research Group in Basic Oncology, Cancer Division of the Quebec University Hospital Research Centre, Québec, Canada; Department of Molecular Biology, Medical Biochemistry and Pathology, Faculty of Medicine, Laval University, Québec, Canada*

Herbert Tschochner • *Biochemie-Zentrum Regensburg (BZR), Lehrstuhl Biochemie III, Universität Regensburg, Regensburg, Germany*

Chelly van Vuuren • *Centre for Chromosome Biology, National University of Ireland, Galway, Ireland*

Minshi Wang • *Department of Cell Biology, Rowan University School of Osteopathic Medicine, Stratford, NJ, USA; Blood Cell Development and Function Program, Fox Chase Cancer Center, Philadelphia, PA, USA*

Renjie Wang • *Laboratoire de Biologie Moléculaire Eucaryote, Centre de Biologie Intégrative (CBI), Université de Toulouse, Toulouse, France*

Part I

Imaging of the Nucleolus

Chapter 1

The Relationship Between Human Nucleolar Organizer Regions and Nucleoli, Probed by 3D-ImmunoFISH

Marjolein van Sluis, Chelly van Vuuren, and Brian McStay

Abstract

3D-immunoFISH is a valuable technique to compare the localization of DNA sequences and proteins in cells where three-dimensional structure has been preserved. As nucleoli contain a multitude of protein factors dedicated to ribosome biogenesis and form around specific chromosomal loci, 3D-immunoFISH is a particularly relevant technique for their study. In human cells, nucleoli form around transcriptionally active ribosomal gene (rDNA) arrays termed nucleolar organizer regions (NORs) positioned on the p-arms of each of the acrocentric chromosomes. Here, we provide a protocol for fixing and permeabilizing human cells grown on microscope slides such that nucleolar proteins can be visualized using antibodies and NORs visualized by DNA FISH. Antibodies against UBF recognize transcriptionally active rDNA/NORs and NOP52 antibodies provide a convenient way of visualizing the nucleolar volume. We describe a probe designed to visualize rDNA and introduce a probe comprised of NOR distal sequences, which can be used to identify or count individual NORs.

Key words 3D-immunoFISH, Nucleolus, Nucleolar organizer region (NOR), Ribosomal genes, Human acrocentric chromosomes

1 Introduction

Nucleoli form around arrays of ribosomal gene (rDNA) repeats that are transcribed by RNA polymerase I, to produce the major RNA components of ribosomes, 18S, 28S, and 5.8S ribosomal RNAs. In humans, these arrays, termed nucleolar organizer regions (NORs), range in size from 50 kb to >6 Mb and are positioned on the short or p-arms of each of the five acrocentric chromosomes HSA13-15, HSA21, and HSA22 [1]. During cell division, the nucleolus disappears and NORs can be observed as achromatic gaps on DAPI-stained metaphase chromosomes. These chromosomal features are sometimes referred to as 2° constrictions. Unlike the highly condensed heterochromatic 1° constrictions, centromeres, the rDNA in 2° constrictions is under-condensed compared to the surrounding heterochromatin. Upstream binding factor (UBF), a nucleolar-specific HMG box protein that binds extensively

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_1,

across the rDNA array is directly responsible for this morphology [2, 3]. In many or possibly most human cell types, a proportion of NORs fail to show this morphology. In these silent NORs, the rDNA is condensed and was not transcriptionally active in the previous interphase and will not be active in the subsequent interphase [4]. These silent NORs lack UBF and will not contribute to nucleolar formation. Upon reactivation of transcription in anaphase, nucleoli begin to form around individual active NORs [5]. Then in most cell types, these small nucleoli will fuse producing the characteristic large nucleoli observed in most human cells.

Nucleoli are subdivided into three distinct components reflecting the stages of ribosome biogenesis [6]. Fibrillar centers (FCs) contain unengaged pools of transcription factors and nontranscribed rDNA sequences. rDNA transcription occurs at the interface between FCs and the associated dense fibrillar component (DFC) where early processing of resulting pre-rRNA takes place. Late processing and ribosome assembly occurs in the surrounding granular component (GC).

Combining fluorescent in situ hybridization (FISH) with antibody staining on cells in which three-dimensional structure is preserved, 3D-immunoFISH, is a powerful tool for examining the intimate relationship between NORs and nucleoli [3, 7, 8]. In principle, it allows one to distinguish active from silent NORs. It is possible to count the individual NORs associated with each nucleolus. Finally, using 3D-immunoFISH, alterations in NOR activity and organization can be visualized in cells as they respond to changes in growth status, drug treatments, or genetic manipulation. In this chapter, we describe DNA probes, antibodies, and protocols for performing 3D-immunoFISH to study the relationship between NORs and nucleoli in human cells.

The 3D-immunoFISH procedure we use is similar to a previously described protocol (*see* ref. [9]). It consists of four essential steps: (1) preparation of slides, (2) preparation and validation of probes, (3) hybridization of slides, and (4) antibody staining.

To visualize nucleoli, we commonly use antibodies against UBF or NOP52. UBF is an FC/DFC protein that associates with active rDNA repeats and NORs [10, 11]. In contrast, NOP52, also known as RRP1, is a GC protein involved in pre-rRNA maturation [12]. It provides a convenient marker for delineating the extent of the nucleolar volume (Fig. 1a).

The human rDNA repeat is approximately 43 kb in length, 13 kb of which are transcribed to produce pre-rRNA and a 30 kb intergenic spacer (IGS) [13]. To visualize rDNA by FISH, we use a 12 kb EcoR I restriction fragment that is positioned in the IGS immediately upstream of the gene promoter (Fig. 1b). As these sequences are nontranscribed, we can be certain that we are visualizing rDNA as opposed to rRNA. This rDNA FISH probe can readily identify individual NORs on metaphase spreads and reveal the variation in their rDNA content. However, it cannot distinguish individual NORs within

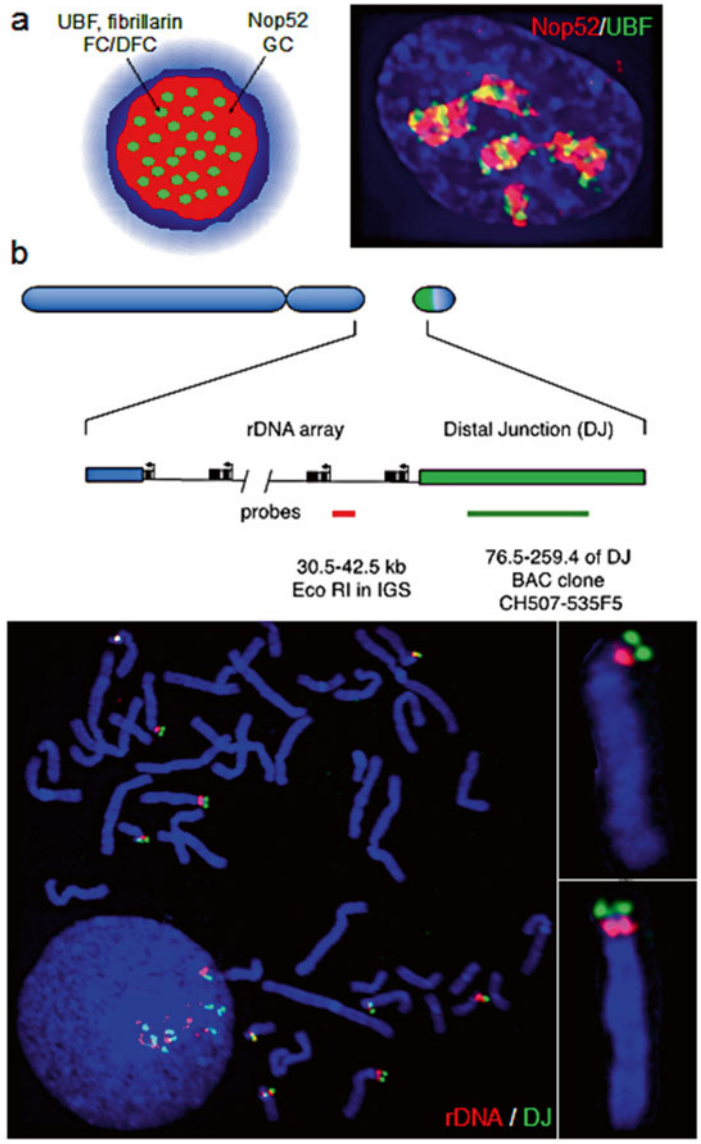

Fig. 1 Antibodies and DNA probes used in probing the relationship between NORs and nucleoli by 3D-immunoFISH. (**a**) Nucleolar fibrillar centers (FC) and dense fibrillar components (DFC) are typically not resolved in the light microscope and can be stained with antibodies against UBF or fibrillarin, resulting in a punctate staining pattern within the nucleolar interior. Antibodies against NOP52, located in the granular component (GC), provide a convenient way of visualizing the nucleolar volume. Staining of an individual human nucleolus is shown in *cartoon* form on the *left*, note the perinuclear heterochromatin (*dark blue*). An individual DAPI-stained human nucleus containing five nucleoli visualized with UBF and NOP52 antibodies is shown on the *right*. (**b**) rDNA is visualized using a 12 kb EcoRI DNA fragment that lies immediately upstream of the gene promoter in the nontranscribed intergenic spacer. This fragment represents sequences between 30.5 and 42.5 kb of the rDNA repeat described in GenBank Accession U13369. Sequences immediately distal to rDNA arrays are referred to as the distal junction (DJ). These sequences are virtually identical on all five human acrocentrics. BAC clone CH507-535F5 contains the sequences that lie between 76.5 and 259.4 kb distal to the rDNA array. FISH using rDNA and DJ probes performed on a normal human male metaphase chromosomes spread is shown below a cartoon representation of the probes. Enlarged individual acrocentric chromosomes are shown on the *right*

interphase nucleoli. The identification of sequences immediately distal to rDNA arrays and the demonstration that these are virtually identical on each acrocentric p-arm, offers a unique opportunity for visualizing and counting NORs [7]. Thus far nearly 400 kb of rDNA distal sequences, distal junction (DJ), has been identified. BAC clone CH507-535F5 (GenBank: CT476834) covers the interval between 76.5 and 259.4 kb distal to rDNA (Fig. 1b). Probing of metaphase spreads with this DJ BAC results in signal that is consistent between NORs (Fig. 1b). Application of this DJ probe to interphase cells can therefore easily enumerate NORs associated with individual nucleoli. DJ sequences associated with active NORs are usually associated with perinucleolar heterochromatin, visualized by DAPI staining (Fig. 2a). Silent NORs are typically observed as isolated DJ signals dissociated from nucleoli. By employing fluorescently labeled secondary antibodies, nucleoli, rDNA, and DJ can be visualized within the same cell (Fig. 2b, c). BACs derived from the long or q-arms can be used to identify each individual acrocentric chromosomes (*see* ref. 7).

2 Materials

The list of materials that we present here is cumulative. Within each section, we list only those materials that are required for the first time.

2.1 Preparation of Slides

1. Superfrost® Plus Microscope slides (Menzel-Glaser) (*see* **Note 1**).
2. QuadriPerm (Sarstedt) or 150 mm cell culture dishes (*see* **Note 2**).
3. The 3D-immunoFISH experiments described here were performed on hTERT-RPE1 cells [14] (ATCC) maintained in DMEM/Nutrient Mixture F-12 Ham containing 2 mM L-Glutamine, 10% v/v FBS, and 0.25% v/v sodium bicarbonate.
4. 10× and 1× Phosphate buffered saline (PBS).
5. Fixative: 4% w/v paraformaldehyde (PFA) in PBS; 4 g in 100 mL PBS, heat at 60 °C until dissolved, then cool to room temperature. Prepare freshly on the day.
6. Rocking platform.
7. Permeabilization buffer: 0.5% w/v saponin, 0.5% v/v Triton X-100 in 1× PBS; 0.5 g saponin and 0.5 mL Triton X-100 in 100 mL PBS. Filter sterilize with a 0.45 μm syringe filter.
8. 20% v/v Glycerol in 1× PBS.
9. Liquid nitrogen.
10. Plastic slide boxes.

2.2 Preparation and Validation of Probes

1. BAC clone CH507-535F5 (GenBank: CT476834) obtained from the BACPAC Resource Center, Children's Hospital Oakland Research Institute in Oakland, California, USA (*see* **Note 3**).

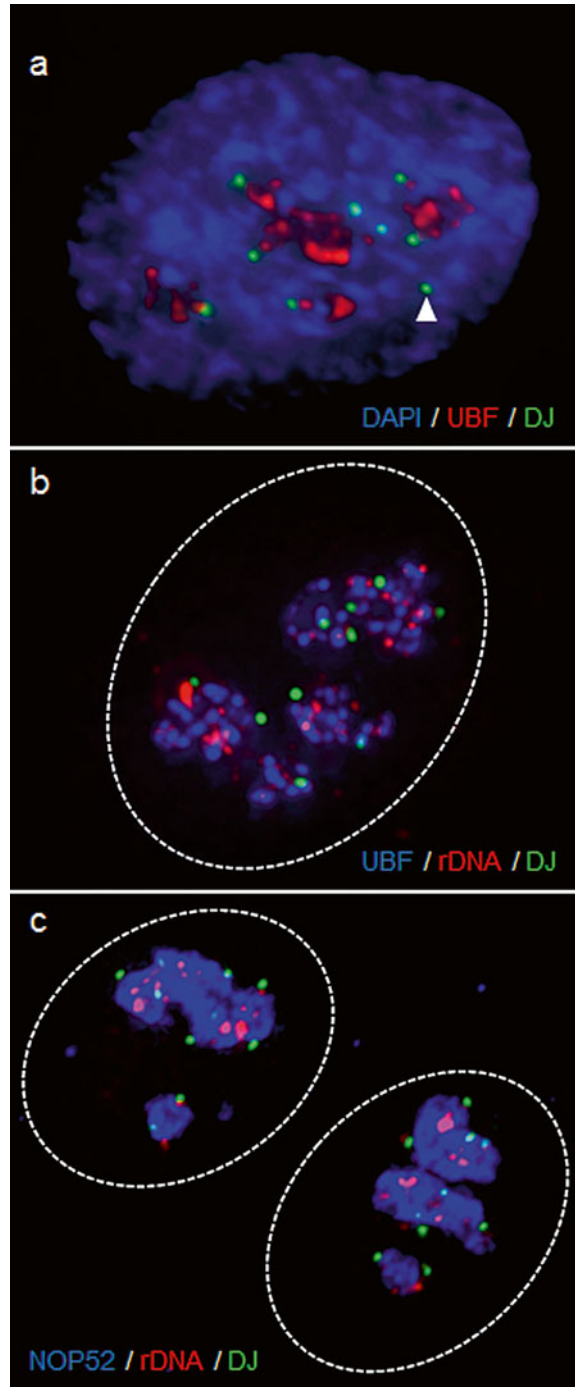

Fig. 2 3D-immunoFISH performed on hTERT-RPE1 cells. (**a**) An individual DAPI-stained cell probed with UBF antibodies and DJ FISH probe. Note that DJ signals are associated with DAPI dense perinucleolar heterochromatin. A silent NOR, lacking UBF signal, is indicated by an *arrowhead*. (**b**) An individual cell probed with UBF antibodies and both rDNA and DJ FISH probes. The nucleus is outlined. (**c**) Two hTERT-RPE1 cells probed with NOP52 antibodies and both rDNA and DJ FISH probes. Note how NOP52 antibodies more clearly delineate the nucleolar volume than UBF antibodies

2. Human rDNA plasmid pUC-hrDNA-12.0: A plasmid containing a 12.0 kb EcoRI fragment that corresponds to sequences between 30.5 and 42.5 kb of the human rDNA repeat (GenBank: U13369) (*see* **Note 4**).
3. Nick translation kit, Green dUTP, and Red dUTP (Abbott Molecular 07J00-001, 02N32-050, and 02N34-050, respectively).
4. Aluminum foil.
5. Dryblocks at 15 and 37 °C.
6. Waterbaths (ranging from 42 to 75 °C).
7. 1 mg/mL Human Cot-1 DNA (Applied Genetics, Hybloc™ Competitor DNA, HHB) (*see* **Note 5**).
8. 10 mg/mL Herring or Salmon sperm DNA.
9. 3 M sodium acetate, pH 5.5.
10. 100, 85, and 70% Ethanol.
11. Hybridization buffer, Hybrisol® VII (MP Biomedicals, 11RIST139010).
12. Human Male Metaphase Slides (Applied Genetics, HMM).
13. Coplin Jars (*see* **Note 6**).
14. Formamide, ACS grade deionized with Amberlite® mixed bed ion exchange resin (Merck) (*see* **Note 7**).
15. 20× SSC: 88.2 g sodium citrate tribase dihydrate, 175.3 g sodium chloride, adjust the pH to 7 with HCl and adjust the volume to 1 L.
16. M-FISH Denaturation buffer: 70% v/v deionized formamide, 2× SSC.
17. Coverslips, 24 mm × 40 mm.
18. Rubber cement (Marabu-Fixogum).
19. Humidity chamber at 37 °C (*see* **Note 8**).
20. Single edge razor.
21. M-FISH Wash buffer I: 0.4× SSC, 0.3% NP-40.
22. M-FISH Wash buffer II: 2× SSC, 0.1% NP-40.
23. VectorShield mounting medium +/− DAPI (4,6-diamidino-2-phenylindole, Vector Laboratories).
24. Nail varnish.
25. Fluorescent microscope, camera, and image capture/analysis software package (*see* **Note 9**).

2.3 Hybridization to Fixed Cells on Slides

1. Formamide, molecular biology grade.
2. 0.1 N HCl.
3. 2× SSC.

4. 3D-FISH Equilibration buffer: 50% v/v molecular biology grade formamide, 2× SSC.
5. Dry slide incubator capable of heating slides to 73 °C for denaturation (*see* **Note 10**).
6. 3D-FISH Wash I: 50% v/v deionized formamide, 2× SSC.
7. 3D-FISH Wash II: 0.1× SCC.

2.4 Antibody Staining

1. Primary antibodies recognizing UBF and/or Nop52 (*see* **Note 11**).
2. 1% w/v BSA in 1× PBS: 0.1 g BSA (IgG free) in 10 mL PBS.
3. Fluorophore (Fluorescein FITC, Rhodamine, or Coumarin AMCA) conjugated whole IgG secondary antibodies (Jackson Immuno Research).

3 Methods

3.1 Preparation of Slides

1. Open a box of Superfrost Plus slides in a tissue culture hood and place slides, frosted side up, in either quadriPERM (1 slide per well) or 150 mm (up to five slides) cell culture dishes, add media, and seed with logarithmically growing cells (*see* **Note 12**).
2. When cells reach 50–70% confluence, replace culture media with the same volume of fixative and incubate for 10 min at room temperature on a rocking platform.
3. Pour off fixative, add the same volume of PBS and incubate for 10 min at room temperature on a rocking platform. This wash step should be repeated a further two times.
4. Pour off the final PBS wash solution, replace with the same volume of permeabilization buffer and incubate for 10 min at room temperature on a rocking platform.
5. Wash the slides three times in PBS as in **step 3**.
6. Pour off the final PBS wash solution and incubate the slides in 20% glycerol in 1× PBS for 2 h at room temperature on a rocking platform. Alternatively, the slides can be incubated overnight at 4 °C.
7. Snapfreeze the slides in liquid nitrogen. Slides can be stored in plastic slide boxes at −80 °C indefinitely.

3.2 Preparation and Validation of Probes

1. Label 1 μg of BAC or plasmid DNA with Green dUTP or Red dUTP by nick translation in a 50 μL reaction according to the manufacturers (Abbott Molecular) protocol. Labeling reactions are protected from light by aluminum foil and incubated at 15 °C for 8 h and terminated by heat inactivation at 70 °C for 10 min. Labeled DNA can be stored at −20 °C in the dark indefinitely.

2. For hybridization to normal human male metaphase chromosome slides, combine 100 ng BAC DNA (5 μL of labeling reaction) and/or 50 ng rDNA (2.5 μL of labeling reaction) with 2.5 μL Human Cot-1 DNA and 5 μL Herring Sperm DNA. Add one-tenth volume of 3 M sodium acetate, pH 5.5 and precipitate the DNA by addition of 2.5 volumes of 100% ethanol. Following washing with 70% ethanol and brief drying, resuspend the DNA pellet in 25 μL Hybrisol® VII by incubation at 37 °C with intermittent mixing.
3. Denature normal human male metaphase chromosome slides by incubating for 5 min in 40–50 mL of M-FISH Denaturation buffer in a Coplin jar sitting in a 75 °C water bath.
4. After denaturation, remove slides and dehydrate by incubating for 1 min in 70% ethanol, then 1 min in 85% ethanol and finally 1 min in 100% ethanol. Air-dry slides.
5. Denature the DNA probe mix in Hybrisol® VII by incubating at 75 °C for 5 min. Apply the denatured probe immediately to a coverslip placed on a 37 °C heating block. Invert the denatured chromosome slide and lower onto the coverslip/probe, holding it in position until the meniscus of the hybridization mix reaches the edge of the coverslip. Turn the slide the right way up and seal the edges of the coverslip with rubber cement.
6. Incubate slides at 37 °C in a humidity chamber for between 16 and 48 h (*see* **Note 13**).
7. After hybridization, remove coverslips and rubber cement using a single edge razor blade. Immediately immerse slides in M-FISH Wash buffer I in a Coplin jar equilibrated to 73 °C in a water bath. Agitate each slide after immersion for 2–3 s and let stand for 2 min.
8. Remove the slide to M-FISH Wash buffer II at room temperature. Agitate each slide after immersion for 2–3 s and let stand for 1 min. Remove slides and air-dry in darkness.
9. Place a drop of Vectashield plus DAPI onto a coverslip. Lower the inverted slide onto it, as before. Remove excess mounting media and seal coverslips using nail varnish (preferably clear).
10. Image slides using an appropriate fluorescent microscope, camera, and image capture/analysis software package (Fig. 1b) (*see* **Note 9**).

3.3 Hybridization to Fixed Cells on Slides

1. Prepare hybridization mix containing labeled probe, human Cot-1 DNA, and Herring sperm DNA exactly as described above (*see* Subheading 3.2, **step 2**) (*see* **Note 14**).
2. Remove slides of fixed and permeabilized cells (*see* Subheading 3.1, **step 7**) from the −80 °C freezer and rinse three times in PBS.

3. Depurinate the slide by placing in a Coplin jar containing 0.1 N HCl for 5 min. Then wash in 2× SSC.
4. To pre-equilibrate the slide, prior to applying the probe and denaturing, remove the slide from the 2× SSC, shake off excess liquid and dry the back and sides of the slide with absorbent tissue. Pipet 250 μL of 3D-FISH Equilibration buffer onto the center of the slide, loosely apply a coverslip and incubate for 15 min in humidity chamber at 37 °C. Care must be taken here that the fixed cells on the slide surface do not dry out.
5. Just prior to the end of the pre-equilibration period, apply the hybridization mix to a coverslip that has been placed on a 37 °C heating block.
6. Remove the coverslip from the pre-equilibrating slide, shake off excess liquid and lower the inverted slide onto the hybridization mix. Again care must be taken here that the cells do not dry out.
7. Seal the coverslip using rubber cement, allowing it to harden by placing the slide (right way up) on the 37 °C heating block for a few minutes.
8. Denature the slide, now containing the hybridization mix, by placing it in a dry slide incubator set at 73 °C for 12 min.
9. After denaturation remove the slide to a humidity box and incubate at 37 °C for between 16 and 48 h (*see* **Note 13**).
10. After hybridization, remove the coverslip and rubber cement and wash the slides in a Coplin jar, three times for 5 min each, with 3D-FISH Wash I at 42 °C.
11. Place the slides contained within a Coplin jar on a rocking platform at room temperature, and wash three times for 5 min each with 3D-FISH Wash II preheated to 60 °C.
12. After the final hybridization wash, prepare the slides for antibody staining by washing three times (5 min each) in PBS.

3.4 Antibody Staining

1. Dilute the primary antibody in 200 μL 1% BSA/PBS at an appropriate dilution and pipet onto the center of the slide. Loosely apply a coverslip. Incubate the slide for 1 h in humidity chamber at 37 °C. Care must be taken here that the fixed cells on the slide surface do not dry out.
2. Remove the coverslip and wash the slide in PBS at room temperature three times for 10 min each.
3. Dilute the fluorophore-conjugated secondary antibodies in 200 μL 1% BSA/PBS (1/200 for FITC and Rhodamine-labeled secondary antibodies and 1/100 for AMCA) and pipet onto the center of the slide. Loosely apply a coverslip. Incubate the slide for 40 min in humidity chamber at 37 °C.
4. Remove the coverslip and wash the slide in PBS at room temperature three times for 10 min each.

5. Place a drop of Vectashield with or without DAPI onto a coverslip. Lower the inverted slide onto it, as before. Remove excess mounting media and seal coverslips using nail varnish (preferably clear).
6. Image slides using an appropriate fluorescent microscope, camera, and image capture/analysis software package (Fig. 2) (*see* **Notes 9** and **15**).

4 Notes

1. SuperFrost® Plus glass slides are made by a process that places a permanent positive charge on standard microscope slides. This forms a bridge so that covalent bonds develop between formaldehyde fixed cells and the glass.
2. In QuadriPerm (Sarstedt) dishes, parallel cell culture on up to four slides can be performed. This is useful for comparing various treatments on cells. 150 mm cell culture dishes can be used to grow cells on up to five slides at a time in the same conditions.
3. BAC DNA is prepared from 800 mL cultures using a Nucleobond Xtra Maxi kit (Macherey Nagel) and following the manufacturer's instructions for low-copy number plasmids.
4. Although direct labeling of the rDNA containing plasmid works in 3D-immunoFISH, we prefer to label gel purified insert.
5. Both the human rDNA IGS and the DJ BAC probes contain repetitive elements, such as *alu* repeats, found elsewhere in the genome. It is therefore essential that hybridization signals from these be suppressed by inclusion of Cot-1 DNA [7, 13].
6. We prefer polypropylene Coplin jars with lids (Fisher 10082844) to glass.
7. Add 5 g of mixed bed resin to 1 L of ACS grade formamide in a glass beaker and mix on a magnetic stirrer for 30 min. Then filter through a 0.2 μM bottle top filter into a brown glass bottle.
8. Slides are placed within sealed plastic food storage boxes with clip-on lids and containing water soaked absorbent tissue to maintain humidity. Sealed boxes are placed in a standard 37 °C incubator.
9. In our opinion, the major determining factors for producing high quality 3D-immunoFISH images are the slide processing, quality of the probes and antibodies used. Images can be captured using a basic fluorescent microscope and imaging software. Currently, we use a Photometric® Coolsnap HQ camera and Volocity 6 imaging software (PerkinElmer) with a 63× Plan Apochromat Zeiss objective mounted on a Zeiss Axioplan2 imaging microscope. Typically we capture Z-stacks of fluorescent images, 30–40 slices for interphase cells and ~5 for meta-

phase spreads. Images are deconvolved by iterative restoration in Volocity 6 with a confidence limit of 95% and a maximum of 50 iterations. In most cases, extended focus projections of deconvolved Z-stacks are presented (*see* Fig. 2).

10. We use a TruTemp DNA MicroHeating System with heating block capable of holding up to 16 slides, originally supplied by Robbins Scientific. While this is discontinued, other systems that could substitute include CytoBrite and CytoBrite Duo Slide incubation systems (SciGene) and ThermoBrite Programmable Temperature Controlled Slide Processing System (Abbott Molecular).
11. While we typically use our own antibodies raised in sheep against recombinant UBF and Nop52, there are many commercial UBF antibodies available and we have good experience with Nop52 antibodies supplied by Novus (NBP1-85338). Examples of other antinucleolar antibodies that work in 3D-immunoFISH can be found in other publications from this laboratory [2, 3, 8, 15].
12. The protocol we describe here works in every established adherent human cell line that we have tested thus far, ranging from primary, to hTERT immortalized and cancer cell lines. Care should be taken using cancer cell lines as they often have complex poorly defined karyotypes. While cells in metaphase are rounded up and make minimal contact with the surface of the slide, they can still be observed if care is taken in slide handling.
13. For repetitive probes, such as rDNA, 16 h hybridization is sufficient. For low-copy number probes, such as the DJ BAC, signals are improved by longer hybridization times, up to 48 h.
14. In the protocol described here, probe and cellular DNA are denatured together in the sealed slide. Accordingly, separate denaturation of the probe mix is not required.
15. Antibody staining is usually performed after the hybridization. However in some cases, determined empirically, it is preferable to perform antibody staining prior to depurination and hybridization. In this case cells are stained with both primary and secondary antibodies. The antibodies are then fixed with 2%PFA/PBS prior to depurination and hybridization as outlined above (*see* Subheading 3.3).

Acknowledgments

B.M. acknowledges the Science Foundation Ireland-Health Research Board-Wellcome Trust Biomedical Research Partnership (Investigator Award 106199/Z/14/Z) for funding work in his laboratory. C.v.V. is funded by a NUIG Hardiman scholarship.

References

1. Stults DM, Killen MW, Pierce HH, Pierce AJ (2008) Genomic architecture and inheritance of human ribosomal RNA gene clusters. Genome Res 18(1):13–18. doi:10.1101/gr.157941.113
2. Mais C, Wright JE, Prieto JL, Raggett SL, McStay B (2005) UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery. Genes Dev 19(1):50–64
3. Grob A, Colleran C, McStay B (2014) Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. Genes Dev 28(3):220–230. doi:10.1101/gad.234591.113
4. McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. Annu Rev Cell Dev Biol 24:131–157. doi:10.1146/annurev.cellbio.24.110707.175259
5. Sirri V, Urcuqui-Inchima S, Roussel P, Hernandez-Verdun D (2008) Nucleolus: the fascinating nuclear body. Histochem Cell Biol 129(1):13–31
6. Raska I, Shaw PJ, Cmarko D (2006) Structure and function of the nucleolus in the spotlight. Curr Opin Cell Biol 18(3):325–334
7. Floutsakou I, Agrawal S, Nguyen TT, Seoighe C, Ganley AR, McStay B (2013) The shared genomic architecture of human nucleolar organizer regions. Genome Res 23(12):2003–2012. doi:10.1101/gr.157941.113
8. van Sluis M, McStay B (2015) A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. Genes Dev 29(11):1151–1163. doi:10.1101/gad.260703.115
9. Bridger JM, Lichter P (1999) Analysis of mammalian interphase chromosomes by FISH and immunofluorescence. In: Bickmore WA (ed) Chromosome structural analysis. A practical approach. Oxford University Press, Oxford, pp 103–121
10. Roussel P, Andre C, Comai L, Hernandez-Verdun D (1996) The rDNA transcription machinery is assembled during mitosis in active NORs and absent in inactive NORs. J Cell Biol 133(2):235–246
11. O'Sullivan AC, Sullivan GJ, McStay B (2002) UBF binding in vivo is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat. Mol Cell Biol 22(2):657–668
12. Yoshikawa H, Ishikawa H, Izumikawa K, Miura Y, Hayano T, Isobe T, Simpson RJ, Takahashi N (2015) Human nucleolar protein Nop52 (RRP1/NNP-1) is involved in site 2 cleavage in internal transcribed spacer 1 of pre-rRNAs at early stages of ribosome biogenesis. Nucleic Acids Res 43(11):5524–5536. doi:10.1093/nar/gkv470
13. Gonzalez IL, Sylvester JE (1995) Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer. Genomics 27(2):320–328
14. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, Harley CB, Shay JW, Lichtsteiner S, Wright WE (1998) Extension of life-span by introduction of telomerase into normal human cells. Science 279(5349):349–352
15. Prieto JL, McStay B (2007) Recruitment of factors linking transcription and processing of pre-rRNA to NOR chromatin is UBF-dependent and occurs independent of transcription in human cells. Genes Dev 21(16):2041–2054

Chapter 2

Analysis of the *C. elegans* Nucleolus by Immuno-DNA FISH

Christian Lanctôt

Abstract

Caenorhabditis elegans is a well-established model organism which allows, among others, to investigate the link between nucleolar structure/function on the one hand and cell fate choices and cellular differentiation on the other. In addition, *C. elegans* can be used to study the role of the nucleolus in processes that can be difficult to faithfully reproduce in vitro, such as gametogenesis, disease development, and aging. Here I present two complementary techniques, immunofluorescent staining and DNA fluorescence in situ hybridization, that have been adapted to label nucleolar components at various stages of the life cycle of the worm.

Key words *Caenorhabditis elegans*, Nucleolus, Immunofluorescence, DNA FISH, Embryogenesis

1 Introduction

Caenorhabditis elegans is a small nematode that is well established as a premier model organism for biomedical research. However, it is fair to say that so far *C. elegans* has not been widely used to study the biology of the nucleolus. Yet, the worm offers a well-characterized system to investigate the various aspects of nucleolar functions, especially the link between nucleolar activity and cell fate choices/differentiation [1]. In addition, as a whole organism, *C. elegans* allows one to study nucleolar structure and function in processes that can be difficult to faithfully reproduce in culture, such as gametogenesis, disease development, and aging [2, 3]. Furthermore, the model is easily amenable to genetics and cell biology approaches. For instance, it was recently found that mutations in the *C. elegans* ortholog of the Nopp140 nucleolar protein lead to abnormal gonadal development and oogenesis [4]. Similarly, depletion of RBD-1, an homolog of the yeast RNA-binding protein Mrd1p, was found to be associated with abnormal processing of pre-rRNA, growth retardation, and larval arrest in the worm [5]. *C. elegans* was also used in an elegant study to show that the embryonic nucleoli form through condensation of a fixed number of nucleolar components when these reach a certain threshold concentration, an assembly which the authors modeled based on the

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_2,

physics of phase transition [6]. In the present paper, I describe how to use immunostaining combined with DNA fluorescence in situ hybridization (FISH) to investigate the *C. elegans* nucleolus.

2 Materials

2.1 Preparation of Samples

1. Microscope slides, 76 mm × 26 mm × 1 mm, precleaned, twin frosted end.
2. Coverslips, 18 mm × 18 mm, standard no. 1, and high quality no. 1.5.
3. Filter paper: e.g., no. 1 circles (cat. 1001 150), cut in quadrants.
4. Diamond pen.
5. Razor blade.
6. Metal block in dry ice: a block from a dry bath (e.g., 10 cm × 8 cm) is turned upside down, half-buried in pellets of dry ice and stored at −80 °C.
7. Gelatin, from porcine skin.
8. Chrome alum.
9. 1 M sodium azide (NaN_3), *caution*, this is a toxic substance.
10. Poly-L-lysine (hydrobromide, M_w > 300,000 g/mol, Sigma-Aldrich).
11. PLL solution (*see* **Note 1**): heat 40 mL of ddH_2O to 60 °C. Dissolve 0.08 g of gelatin. Cool the solution to 37 °C. Add 0.008 g of Chrome Alum. Mix well. Add 40 µL of 1 M NaN_3 (final concentration is 1 mM). Dissolve 40 mg of poly-L-lysine (final concentration is 1 mg/mL). Transfer to a 50-mL tube. Store at 4 °C.
12. 10 N NaOH.
13. Sodium hypochlorite solution ("bleach," available chlorine contents 4.00–4.99 %, Sigma-Aldrich).
14. M9: dissolve the following in 200 mL of ddH_2O: 0.6 g of KH_2PO_4 (anhydrous, final concentration 22 mM), 1.495 g of $Na_2HPO_4 \cdot 2H_2O$ (42 mM), 1 g of NaCl (86 mM). Add 0.2 mL of 1 M $MgSO_4$ (1 mM). Sterilize on 0.22 µm filter (optional). Store at room temperature.
15. Egg Buffer: dissolve the following in 450 mL of ddH_2O: 3.448 g NaCl (final concentration 118 mM), 1.789 g KCl (48 mM), 0:147 g $CaCl_2 \cdot 2H_2O$ (2 mM), 0.203 g $MgCl_2 \cdot 6H_2O$ (2 mM), 2.979 g HEPES (M_w 238.3 g/mol, 25 mM). Adjust pH to 7.3 with 1 N NaOH (no more than 10 mL). Complete to 500 mL with ddH_2O. Sterilize on 0.22 µm filter (optional). Store at room temperature.
16. 100 mM levamisole-HCl (in water, Sigma-Aldrich), *caution*, this is a toxic substance.

2.2 Immuno-fluorescent Staining

1. 1× PBS.
2. 10× PBS.
3. Tween-20™, polyoxyethylene-sorbitan monolaurate.
4. Coplin jars.
5. Silicon isolator rings, Press-to-seal Gasket, 13 mm diameter, 2.5 mm depth (Electron Microscopy Sciences Inc.).
6. Methanol, in a Coplin jar, at –20 °C.
7. Acetone, in a Coplin jar, at –20 °C.
8. 70, 50 and 30% acetone, (v/v) in ddH_2O, precooled at 4 °C in Coplin jars.
9. 4% formaldehyde: mix 26 mL of ddH_2O, 4 mL of 10× PBS and one vial (10 mL) of 16% paraformaldehyde solution, EM Grade (Electron Microscopy Sciences Inc.).
10. PBST (PBS-0.02% Tween): 10 µL of Tween-20™ in 50 mL of 1× PBS.
11. PBST-BM: dissolve 0.05 g bovine serum albumin fraction V and 0.05 g dry milk (low fat) in 10 mL of PBST.
12. Primary antibodies.
13. Secondary antibodies coupled to fluorophores, e.g., goat anti-mouse IgG, Alexa Fluor 488 conjugate (Life Technologies) or goat anti-rabbit IgG, Cy3 conjugate (Jackson Immunoresearch).
14. 0.5 mg/mL DAPI, in 1× PBS, Sigma-Aldrich.
15. Vectashield mounting medium, Vector Laboratories.
16. Nail polish.
17. Microbeads stained with multiple fluorophores (e.g., TetraSpeck™ from Life Technologies).
18. Wide field or confocal fluorescence microscope.

2.3 DNA FISH

1. 5 U/µL Taq DNA polymerase, or equivalent.
2. 10× Taq enzyme buffer.
3. 1.5 mM dTTP, in water.
4. 2 mM dAGC: dATP, dGTP, dCTP, 2 mM each (in water).
5. 1 mM Cy3-dUTP, GE Healthcare Amersham.
6. Forward primer, 25 µM, 5′-TTGTGCAAGCGGCCGAGGTC-3′.
7. Reverse primer, 25 µM, 5′-AGACTCAAGCGCCTCGACGC-3′.
8. PCR cycler.
9. Ion-exchange columns for the purification of DNA, e.g., Zymoclean™, QiaQuick™, or equivalent.
10. Nano-spectrophotometer.
11. 0.1 N HCl: add 0.44 mL of concentrated (11.3 N) HCl to 49.5 mL of ddH_2O.

12. 0.01 N HCl: add 5 mL of 0.1 N HCl to 45 mL of ddH_2O.
13. 20× SSC: add 17.5 g of NaCl (3 M) and 8.8 g of sodium citrate tribasic dehydrate (0.3 M) to 80 mL of ddH_2O. Adjust the pH to 7 with 1 N HCl. Complete to 100 mL with ddH_2O.
14. 2× SSC.
15. 0.2× SSC.
16. 10 mg/mL RNAse A, in water, store at −20 °C.
17. Formamide, molecular biology grade (deionized).
18. 2× SSC/50% formamide: add 5 mL of 20× SSC to 20 mL of ddH_2O. Add 25 mL of molecular biology-grade (deionized) formamide.
19. Hybridization chambers: 13 mm diameter, 20 μL volume (Electron Microscopy Sciences Inc.).
20. 2× Hyb Buffer: weigh 1 g of dextran sulfate (Sigma-Aldrich) in a 15 mL tube. Add 4× SSC to 5 mL. Vortex and rotate to dissolve completely. Make 0.5 mL aliquots and store at −20 °C.

3 Methods

3.1 Preparation of Samples

Worms are easily and cheaply maintained according to standard procedures [7]. At 20 °C, *C. elegans* goes from embryo to adult in 3–4 days, passing through four larval stages [8]. The animals are transparent and their typical lifespan is on the order of 15–30 days, depending on the culture conditions. The vast majority of adults are self-fertilizing hermaphrodites. A number of well-maintained websites cover various aspects of the worm's biology: anatomy (http://www.wormatlas.org/), strains (http://cbs.umn.edu/cgc/home), genomics (www.wormbase.org), general (www.wormbook.org), to name just a few.

3.1.1 Treatment of Microscope Slides

Although the procedures described here can be performed in solution, I prefer to stain samples that are first affixed to a microscope slide. This ensures a better penetration of reagents and, to some extent, a better preservation of the sample. However, it is crucial to treat the slides in order to promote adhesion of the worms. The importance of this step cannot be overstated, especially if working with adult worms. Indeed, partial or even complete loss of the sample is a frequent problem when using the protocol described here. To minimize this problem, slides are usually coated with high molecular weight poly-L-lysine, which provides a positive surface to which the negatively charged cuticle of the worm can attach (*see* **Note 2**).

1. Bring the poly-L-lysine solution (PLL) to room temperature.
2. Place two precleaned microscope slides back to back, dip in the PLL solution, and incubate 2 min.

3. Remove the slides from the PLL solution and place them upright against a support. Leave to dry at least 20 min, protected from dust.
4. Use on the same day (*see* **Note 3**).

3.1.2 Isolation of Early Embryos

C. elegans embryos develop in the uterus until shortly after gastrulation (~25–30-cell stage). The easiest way to obtain a large number of embryos is therefore to dissolve the hermaphrodite body in conditions that minimally affect the embryos. This is done by a brief incubation in a NaOH/bleach solution, followed by several washes in a neutral buffer.

1. Collect healthy young adult worms in ddH_2O and transfer to a 15 mL tube. Let the worms settle by gravity. Wash once with 10 mL of ddH_2O.
2. Resuspend the worm pellet in a total of 3.75 mL of ddH_2O.
3. Mix 0.25 mL 10 N NaOH and 1 mL of sodium hypochlorite solution. (This mix must be made fresh). Add to the worm solution. Vortex briefly.
4. Incubate at room temperature until the worms are no longer visible with the naked eye (usually 5–7 min). Vortex briefly every other minute.
5. Centrifuge 30 s at 1200 × *g*. Quickly remove the supernatant and add 5 mL of Egg Buffer to the pellet. Vortex briefly.
6. Repeat **step 5** three times.
7. After the last wash, leave ~50 µL/60 mm petri of worms above the egg pellet. If needed, let embryos grow at room temperature until they reach the desired stage (e.g., ~60 min to get to the ~200-cell stage).

3.1.3 Freeze-Crack

The *C. elegans* embryonic egg shell and larval/adult cuticle are largely impermeable. To be able to stain internal structures, it is necessary to break open and tear away part of the egg shell or cuticle. This is done by freezing the sample under a glass coverslip and then popping off the coverslip in order to rip up, so to speak, the egg shell/cuticle [9]. Please note that this step takes some practicing before the optimal "tearing" conditions can be reproducibly generated.

1. If starting with isolated embryos, please go to **step 4**.
2. Collect worms in M9 and transfer to a 15 mL tube. Spin 30 s at 200 × *g* at room temperature. Wash two times with 5 mL of M9 and once in ddH_2O (*see* **Note 4**).
3. After the last wash, leave ~50 µL of supernatant/60 mm petri of worm. Add 0.5 µL of 100 mM levamisole. Mix gently.
4. Label a PLL-treated slide and, with a diamond pen, gently draw a circle of approximately 1 cm in diameter *on the reverse side*, in the center. This slight etching helps to locate the sample in subsequent steps.

5. Using a tip rinsed in M9 + 0.002 % (v/v) Tween-20™ (*see* **Note 5**), pipette 25 µL of the embryo/worm solution inside the circle onto a PLL-treated slide.
6. Using forceps, cover gently with an 18 mm × 18 mm coverslip. Working at low magnification under the stereomicroscope, adsorb the liquid using a filter paper until the worms/embryos are slightly compressed between the slide and the coverslip (*see* **Note 6**).
7. Place the slide on a metal block in dry ice. Incubate at −80 °C for at least 30–60 min.
8. Holding the slide almost vertically, wedge a razor blade under the corner of the coverslip and pop it off. Take care not to thaw the slide before popping.
9. Proceed immediately to fixation.

3.2 Immunofluorescence Staining

Although a large collection of transgenic strains that express a variety of GFP fusion proteins is now available to the *C. elegans* research community [10], immunostaining remains a powerful technique due to its ability to reveal multiple antigens simultaneously and its high sensitivity when compared to GFP-based labeling. It is estimated that up to 80 % of human genes have orthologs in *C. elegans* [11]. Unfortunately, due to the relatively limited homology of proteins between these two species, typically around 40 %, many of the antibodies that were raised against mammalian antigens fail to detect their counterpart in the worm. It is therefore advisable to use antibodies raised against *C. elegans* proteins, many of which can be searched for in Wormbase (www.wormbase.org) and obtained directly from *C. elegans* researchers or from the Developmental Studies Hybridoma Bank (http://dshb.biology.uiowa.edu/). If a *C. elegans*-specific reagent cannot be located, then it is worth testing several antibodies against the target protein to find one that reacts against the worm protein. Figure 1 shows examples of an embryo and a young larva immunostained for both a nucleolar protein (DAO-5) and a germ line marker (PGL-1). These results indicate that the formation of the nucleolus is delayed in the embryonic germ line, thus highlighting how *C. elegans* can be used to reveal unique features of nucleolar biology in vivo.

3.2.1 Fixation

The fixation conditions have to be optimized depending on the target protein(s) and the extent of structural preservation that is wished for. I have found that fixation with acetone followed by progressive rehydration usually provides the most intense immunofluorescent signal and satisfactorily preserves intracellular structures, especially in L4 larvae and adult worms. However, it should be kept in mind that formaldehyde fixation is preferred if one is to perform DNA FISH after immunostaining, since acetone-fixed material does not withstand well the FISH procedure.

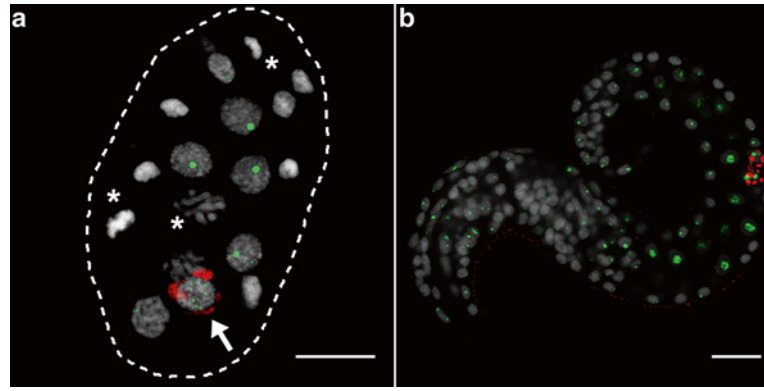

Fig. 1 Nucleologenesis is delayed in the *C. elegans* embryonic germ line. Samples were immunostained with an antibody against DAO-5 (*green*), a Nopp140 homolog, and PGL-1, a marker of germ line cells (*red*). DNA was counterstained with DAPI (*gray*). Scale bar, 10 μm. (**a**) In the 24-cell embryo nucleoli have formed in somatic cells, but not in the single germ cell precursor (*arrow*), in which the DAO-5 signal is seen as ~10–15 *small dots* found throughout the nucleoplasm. Note that this embryo contains numerous mitotic cells (some labeled with *asterisks*), none of which display nucleoli. The contour of the embryo is indicated by a *dotted line*. Shown here is only a cross-section of the embryo (4 μm in thickness). (**b**) In the L1 larva, the two germ cell precursors (*red*) show prominent DAO-5-labeled nucleoli

1. After freeze-crack and removal of the coverslip, immediately fix embryos in cold 100% methanol, 2 min at −20 °C.
2. Transfer the slide(s) to acetone at −20 °C and fix a further 5 (for embryos) or 10 min (for larvae and adults).
3. Incubate the slides successively in precooled 70% acetone, 50% acetone, and 30% acetone, 1 min each at 4 °C (*see* **Note 7**) (*see* **Note 8** if DNA FISH is to be performed after immunostaining).
4. Rinse slides in PBS. Wash 3 min in fresh PBS at room temperature.

3.2.2 Staining

From this point on, all incubations and washes are performed at room temperature unless otherwise stated. These steps are performed in small volumes (~200–500 μL) pipetted in a well that is created around the sample by gluing a silicon ring adaptor on the slide. Solutions should not be squirted directly onto the samples but rather dispensed on the side of the well.

1. Remove the slides one at a time from PBS. Carefully wipe out the liquid around the sample, i.e., outside the etched circle. Glue a 13 mm silicon ring adaptor around the sample. Press firmly using a second microscope slide. Immediately pipette 500 μL of PBS in the well thus created. Repeat for every slide.
2. Permeabilize the sample by incubating for 5 min in PBST (*see* **Note 9**).

3. Wash 3 min in PBS.
4. Dilute the primary antibody(ies) in 200 µL PBST-BM/sample (*see* **Note 10**).
5. Pipette the primary antibody solution onto the samples. Place the slides in a humidified chamber and incubate 2 h at room temperature or overnight at 4 °C.
6. Wash two times 3 min with 500 µL of PBST.
7. Dilute the secondary antibody(ies) in 200 µL PBST-BM, typically 1/400 (*see* **Note 11**).
8. Pipette the secondary antibody solution onto the samples. Place the slides in a humidified chamber and incubate 2 h at room temperature or overnight at 4 °C.
9. Wash two times 3 min with 500 µL PBST.
10. Wash once 3 min with 500 µL PBS.
11. If DNA FISH is to be performed, fix the immunocomplexes using 2 % formaldehyde in PBS (10 min), wash twice with PBS, and proceed to Subheading 3.3.2.
12. Counterstain with a 1/500 dilution of DAPI (final concentration of 1 µg/mL) for 2 min.
13. Rinse with 500 µL PBS. Remove the silicon ring adaptor. Mount in Vectashield. Use high quality 18 mm × 18 mm no. 1.5 coverslips. Seal with nail polish. Store slides at 4 °C until and between observations. Proceed to Subheading 3.4.

3.3 DNA FISH

While in some cases immunostaining is sufficient to reveal nucleolar structure, in others it is necessary to combine this technique with DNA FISH in order to highlight the position of the rDNA repeats in the nucleus. For instance, using immuno-DNA FISH, we have found that nucleolar proteins such as DAO-5 and FIB-1 label extranucleolar structures in the early *C. elegans* embryo (Fig. 2), the identity of which remains to be determined [12]. There is a single array of about 50 rDNA repeats (the *rrn* genes) in the *C. elegans* genome [13]. It is located at the end of chromosome I and encodes the 18S, 5.8S, and 26S rRNA molecules [14].

3.3.1 Probe Synthesis

Since the rDNA sequence is repeated many times in the genome, a short fragment that is synthesized and labeled by PCR can serve as an efficient FISH probe against rDNA. In the following protocol, a 390-bp fragment encompassing nt 172–561 upstream of the 18S rRNA sequence is amplified from genomic DNA and labeled with Cy3-dUTP.

1. Prepare the following reaction.

C. elegans genomic DNA	200 ng
Forward primer	1 µL (final 500 nM)
Reverse primer	1 µL (final 500 nM)
2 mM dAGC	5 µL (final 200 µM)
1.5 mM dTTP	4.3 µL (final 130 µM)
1 mM Cy3-dUTP	3.25 µL (final 65 µM) (*see* **Note 12**)
10× Taq Buffer	5 µL
H_2O	complete to 49.5 µL

2. Program the following PCR.

   94 °C, 3 min (initial denaturation).

   25 cycles [94 °C 30 s, 58 °C 30 s, 72 °C 30 s].

   72 °C 10 min (final elongation).

3. Hot start the PCR by adding 2.5 U (0.5 µL) of Taq DNA polymerase.
4. Analyze a 2.5 µL aliquot of the PCR reaction (1/20) on an agarose gel.
5. If a single band is visible (*see* **Note 13**), purify the rest of the probe on an ion-exchange column (Zymoclean™, QiaQuick™ or equivalent). Elute in 50 µL of elution buffer.

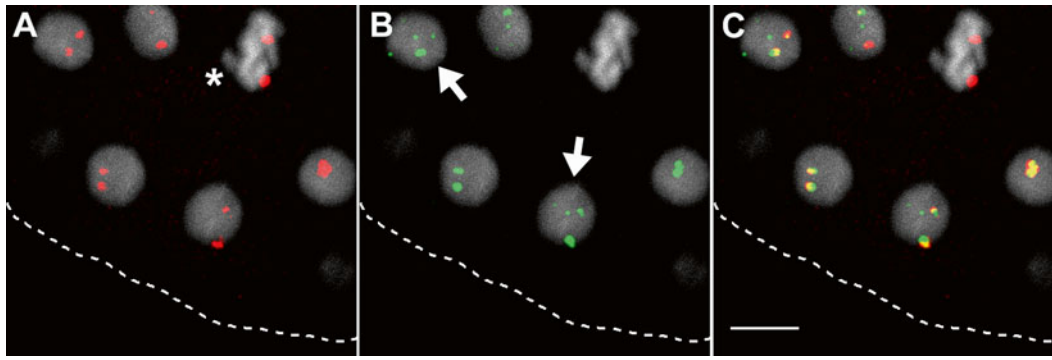

Fig. 2 Immuno-DNA FISH can be used to identify nucleoli. Samples were first immunostained with an antibody against DAO-5 and then hybridized with a labeled DNA probe against the first intervening sequence (ITS1) of *C. elegans* rDNA. Shown here is a cross-section (4 µm in thickness) of part of a ~40-cell embryo. The *dotted line* indicates the egg shell. Scale bar, 10 µm. (**a**) The DNA FISH signal (*red*) appears as two distinct dots in each nucleus, including in mitotic cells (*asterisk*). (**b**) The DAO-5 signal (*green*) localizes to rDNA spots, but is also found in ~4–5 *smaller dot*-like structures throughout the nucleoplasm. The *arrows* point to nuclei in which these structures are clearly seen. (**c**) Merged image showing only partial colocalization of the immunofluorescence (DAO-5, *green*) and DNA FISH (rDNA, *red*) signals

6. Assay the probe concentration by nano-spectrophotometry. Store the probe at –20 °C.

3.3.2 DNA FISH

1. Wash 2 min with 500 µL PBS.
2. Rinse once in 0.01 N HCl.
3. Incubate *precisely 2 min* in 0.1 N HCl at room temperature.
4. Wash once with PBS, then with 2× SSC, each time for 2 min.
5. Treat with 50 µg/mL RNAse A in 2× SSC, 45 min at 37 °C.
6. Remove the silicon ring adaptor.
7. Transfer the slides to a Coplin jar filled with 2× SSC.
8. Incubate in 2× SSC/50% formamide for at least 2 h.
9. Dilute the probe to a concentration of 2 ng/µL in molecular biology-grade (deionized) formamide (12 µL per slide). Add an equal volume of 2× Hyb. The final probe concentration is 1 ng/µL.
10. Mix well by pipetting and vortexing. Spin 20 s in a tabletop centrifuge to remove air bubbles. Keep the probe at room temperature.
11. Pipette 20 µL of the probe solution into a hybridization chamber.
12. Take a slide out of the 2× SSC/50% formamide solution. Gently wipe off excess liquid around the sample using a soft paper.
13. Invert the sample onto the hybridization chamber prepared in **step 11**. Press firmly to seal the chamber.
14. Repeat **steps 11–13** for each slide.
15. Prehybridize the slides at 37 °C overnight (*see* **Note 14**).
16. Denature the probe and the genomic DNA simultaneously by leaving the slides for 3 min upright on a heating block set at 76 °C.
17. Incubate at 37 °C for 2 days.
18. Fill four Coplin jars with 2× SSC. Preheat three of those in a water bath at 37 °C. Fill two other Coplin jars with 0.2× SSC and preheat at 55 °C.
19. Using the tip of a scalpel or a razor blade, gently lift and remove the hybridization chamber.
20. Place the slide in a Coplin jar filled with 2× SSC at room temperature.
21. Wash three times 5 min in 2× SSC at 37 °C. During these and subsequent washes, the slides should occasionally be gently pulled out of the solution and plunged back in.
22. Wash two times 5 min in 0.2× SSC at 55 °C.
23. Counterstain with a 1/500 dilution of DAPI (final concentration of 1 µg/mL) for 2 min. Pipette 500 µL of the DAPI dilution onto the sample.

24. Rinse with 500 μL of 2× SSC. Mount in Vectashield. Use high quality 18 mm × 18 mm no. 1.5 coverslips. Seal with nail polish. Store slides at 4 °C until and between observations.

3.4 Imaging

The labeled samples can be imaged on a wide field or confocal fluorescence microscope. We routinely image *C. elegans* samples on a Leica TCS SP5 scanning confocal microscope equipped with a 63× oil immersion objective (numerical aperture 1.4). The distance between individual optical sections (*z* interval) is set at 400 nm and the *xy* pixel size, at ~80 nm. In the case of multicolor imaging experiments designed to assess the extent of colocalization of two different nucleolar or cellular components, care should be taken to correct for the chromatic aberration that is introduced by the optical system (if any). This can be done by imaging microbeads stained with multiple fluorophores (e.g., TetraSpeck™ from Life Technologies), measuring the distance between the focal points for the different fluorophores and using this measure to align image stacks.

4 Notes

1. The poly-L-lysine solution should be stored at 4 °C. It is stable for at least 2–3 months and can be reused a number of times. However, it should be made fresh once a decrease in the adhesion of samples to the slides is noticed.
2. Microscope slides can also be treated with a silane derivative to increase adhesiveness (e.g., aminoethylamine-propyltrimethoxysilane APTS). However, in my experience, this treatment is less efficient than the one with the poly-L-lysine solution.
3. To provide even greater adhesiveness, a small drop of the PLL solution (~20 μL) can be pipetted in the center of the slide, which is then placed on a hot plate (~75 °C) and left there until the PLL solution has evaporated. The spot of evaporated PLL should be very sticky.
4. It is important to thoroughly wash out the bacteria and to resuspend the worms in water, otherwise they will not stick as well to the PLL-treated slide.
5. Worms and embryos tend to stick to plastic tips. To prevent loss of material, the tips are first rinsed in a solution that contains traces of detergent (0.002% v/v) before being used to pipette worms or embryos.
6. This is a critical step. The coverslip should touch the worms and embryos, but not compress them to the point that these are grossly deformed.

7. The rehydration solutions should be cooled to 4 °C in ice or in the refrigerator before use. I have found this precooling to be very important in order to avoid detachment of the samples from the slides.
8. Alternatively, the slides can be rinsed in cold (4 °C) PBS after the incubation in methanol, transferred to cold (4 °C) 4% formaldehyde, and fixed for 10 min *at room temperature*. As earlier, the use of precooled solution is essential to prevent loss of material.
9. Permeabilization with low concentrations of Tween-20™ is sufficient in most cases. If needed, a more thorough permeabilization can be achieved using PBS supplemented with 0.2% (v/v) Triton X-100.
10. The optimal dilution must be determined empirically. The typical range is 1/20 to 1/400.
11. The immunofluorescent signal afforded by strong fluorophores such as Alexa Fluor 488 or Cy3 generally withstands well a subsequent FISH procedure. However, if a higher sensitivity is required, the secondary antibody can be coupled to a biotin moiety instead of a fluorophore. In this case, the immunocomplexes should be labeled after DNA FISH using streptavidin or avidin fluorescent conjugates.
12. The ratio of Cy3-dUTP to dTTP is 1:2. A higher ratio is not recommended.
13. Cy3-labeled double stranded DNA does not bind SyBr Green efficiently, which means that only the quality, but not the quantity of the PCR product can be assessed by agarose gel electrophoresis.
14. Hybridization can be carried out on a hybridization plate. Alternatively, slides can be placed in a small, flat-bottom metal container, which is left to float for the duration of the hybridization in a water bath set at 37 °C.

Acknowledgements

I gratefully acknowledge the financial support of the Czech Science Foundation (grant P305/12/1246). This work was also supported by Charles University in Prague (PRVOUK P27/LF1/1). The Imaging Center at the institute is supported by the European Regional Development Fund (OPPK CZ.2.16/3.1.00/24010).

References

1. Lee LW, Lee CC, Huang CR et al (2012) The nucleolus of *Caenorhabditis elegans*. J Biomed Biotechnol 2012:601274
2. Baumeister R, Ge L (2002) The worm in us - *Caenorhabditis elegans* as a model of human disease. Trends Biotechnol 20:147–148

3. Gonzalez-Aguilera C, Palladino F, Askjaer P (2014) *C. elegans* epigenetic regulation in development and aging. Brief Funct Genomics 13:223–234
4. Lee CC, Tsai YT, Kao CW et al (2014) Mutation of a Nopp140 gene *dao-5* alters rDNA transcription and increases germ cell apoptosis in *C. elegans*. Cell Death Dis 5:e1158
5. Saijou E, Fujiwara T, Suzaki T et al (2004) RBD-1, a nucleolar RNA-binding protein, is essential for *Caenorhabditis elegans* early development through 18S ribosomal RNA processing. Nucleic Acids Res 32:1028–1036
6. Weber SC, Brangwynne CP (2015) Inverse size scaling of the nucleolus by a concentration-dependent phase transition. Curr Biol 25:641–646
7. Stiernagle T (2006) Maintenance of *C. elegans*. In: Community TCeR (ed) WormBook, doi: 10.1895/wormbook.1.101.1, pp 1–11
8. Corsi AK, Wightman B, Chalfie M (2015) A transparent window into biology: a primer on *Caenorhabditis elegans*. Genetics 200:387–407
9. Strome S, Wood WB (1982) Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 79:1558–1562
10. Sarov M, Murray JI, Schanze K et al (2012) A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*. Cell 150:855–866
11. Lai CH, Chou CY, Ch'ang LY et al (2000) Identification of novel human genes evolutionarily conserved in *Caenorhabditis elegans* by comparative proteomics. Genome Res 10:703–713
12. Korčeková D, Gombitová A, Raška I et al (2012) Nucleologenesis in the *Caenorhabditis elegans* embryo. PLoS One 7:e40290
13. Albertson DG (1984) Localization of the ribosomal genes in *Caenorhabditis elegans* chromosomes by in situ hybridization using biotin-labeled probes. EMBO J 3:1227–1234
14. Ellis RE, Sulston JE, Coulson AR (1986) The rDNA of *C. elegans*: sequence and structure. Nucleic Acids Res 14:2345–2364

Chapter 3

Correlative Light and Electron Microscopy of Nucleolar Transcription in *Saccharomyces cerevisiae*

Christophe Normand, Maxime Berthaud, Olivier Gadal, and Isabelle Léger-Silvestre

Abstract

Nucleoli form around RNA polymerase I transcribed ribosomal RNA (rRNA) genes. The direct electron microscopy observation of rRNA genes after nucleolar chromatin spreading (Miller's spreads) constitutes to date the only system to quantitatively assess transcription at a single molecule level. However, the spreading procedure is likely generating artifact and despite being informative, these spread rRNA genes are far from their in vivo situation. The integration of the structural characterization of spread rRNA genes in the three-dimensional (3D) organization of the nucleolus would represent an important scientific achievement. Here, we describe a correlative light and electron microscopy (CLEM) protocol allowing detection of tagged-Pol I by fluorescent microscopy and high-resolution imaging of the nucleolar ultrastructural context. This protocol can be implemented in laboratories equipped with conventional fluorescence and electron microscopes and does not require sophisticated "pipeline" for imaging.

Key words Correlative light and electron microscopy (CLEM), Nucleolus, RNA polymerase I (Pol I), Ribosomal RNA genes (rDNA), Pol I transcription, Transmission electron microscopy (TEM), Fluorescence microscopy, Yeast, *Saccharomyces cerevisiae*

1 Introduction

Correlative light and electron microscopy (CLEM) combines the benefits of light microscopy (LM) and transmission electron microscopy (TEM). On the one hand, LM can provide information about distribution of fluorescent tagged proteins and their dynamic properties when performing live-cell imaging; rare events can also be preselected by LM before TEM analysis. On the other hand, TEM produces, at high resolution, images of the ultrastructural context of the proteins of interest. The main determinant of success in CLEM is the sample preparation. It should ensure the preservation of the cellular ultrastructure and the fluorescence of the tags simultaneously.

Here, we present a detailed protocol including sample preparation and an imaging workflow adapted for CLEM analysis of the

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_3,

transcription of ribosomal RNA (rRNA) genes by the RNA polymerase I (Pol I) in the nucleus of the budding yeast *Saccharomyces cerevisiae*. The advantages of this model stem from the repeated nature of the ribosomal DNA (rDNA) (~150 copies, half of them being transcribed) and the local very high transcriptional activity of the Pol I; it results in the formation of a subnuclear domain, the nucleolus, easy to observe and considered to be the "transcription factory" for rRNA production (Fig. 1a). Contrariwise, RNA polymerase II (Pol II) transcription, producing messenger RNA (mRNA), is a phenomenon intrinsically difficult to characterize at the single molecule level due to the thousands of different transcripts Pol II has to generate, scattered throughout the nucleoplasm. The ultrastructural characterization of Pol I transcribed genes is comparatively simple and was introduced in 1969 by O. Miller [1] by TEM analysis of extracted and spread nuclear chromatin (Fig. 1b, c). This direct electron microscopy observation constitutes to date the only system to quantitatively assess transcription at a single molecule level: each Pol I transcription unit adopts a characteristic "Christmas tree" conformation with each Pol I molecule loaded on the rDNA (tree's trunk) associated with nascent rRNA appearing as growing "branches" decorated by "terminal knobs" (Fig. 1c). However, the spreading procedure is likely generating artifact and despite being informative, these spread rRNA genes are far from their in vivo situation! The integration of the structural characterization of spread rRNA genes in the three-dimensional (3D) organization of the nucleolus by CLEM would represent an important scientific

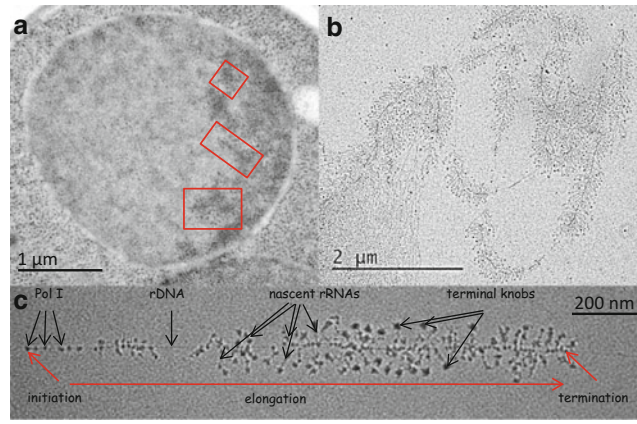

Fig. 1 Electron micrograph of the ribosomal RNA genes. (**a**) On a yeast section, the nucleus of *S. cerevisiae* displays a crescent-shaped dense nuclear subdomain, the nucleolus organized around the repeated rRNA genes transcribed by the Pol I machinery. Note that rod-like structures strongly counterstained, reminiscent of spread transcription units are visible on sections (*red rectangles*). (**b**) By Miller's treatment, the head-to-tail repetitions of rRNA genes are spread and can be individually observed. (**c**) High magnification of one rDNA transcription unit adopting characteristic "Christmas tree" configuration

achievement. *S. cerevisiae* provides an ideal model system for such studies, since rRNA genes represent about 10% of the genome, rendering their detection relatively simple. In addition, thanks to a diversity of genetic tools, the Pol I transcription apparatus can be easily modified in baker's yeast. Finally, while Pol I was for long time considered as a very specialized transcriptional machinery for the rDNA, recent structural studies provided evidences that molecular details of transcription by Pol I and Pol II are very related, consolidating the relevance of rDNA transcription as a model for highly transcribed genes [2].

The CLEM method here described allows detection of Pol I and histones molecules tagged with fluorophores on yeast sections which are then imaged by TEM. This long procedure includes (1) the fixation step and resin embedding that preserve both ultrastructure of the cell nucleus and the fluorescence of tagged proteins, (2) the sectioning and harvesting of the sections in a compatible setup for fluorescent and TEM microscopy, and (3) the development of an "imaging workflow" to sequentially image sections in fluorescence and in TEM and align the pictures.

2 Materials

Prepare all solutions using deionized ultrapure water with a resistivity of 18 MΩ at 25 °C (milli-Q water) and analytical grade reagents. Carefully follow all waste disposal instructions when disposing waste materials; pay especially attention with reagents for TEM which are often hazardous.

2.1 Fixation and Resin Embedding

1. Cryoprotective solution to resuspend the yeasts: dissolve 200 mg agarose Low Melting Point and 140 mg sucrose in 0.1 M sodium cacodylate buffer (EMS). 0.2 M cacodylate buffer (pH 7.2) in milli-Q water is stored at 4 °C for few months. (Take care, cacodylate contains sodium dimethyl arsenate!) Just before using, melt the cryoprotective solution few seconds in a microwave oven and keep it fluid on a water bath at 40–42 °C.
2. Cryogenic fluid: liquid nitrogen (LN_2) is extremely cold (−196 °C) and can cause severe burns to the skin and eyes. The experimenter has always to work with protective gloves and glasses. When LN_2 evaporates, the air oxygen content decreases; therefore, LN_2 is stored in Dewars stored in well ventilated area and the working place has to be kept in a well aerated room.
3. The fixative: 0.1% uranyl acetate, $2H_2O$ (LauryLab) is dissolved in glass distilled acetone. The use of uranyl acetate is under strict regulation. To avoid any risk of contamination, manipulate the uranyl acetate exclusively under the fume hood. Dispose a small quantity of uranyl acetate crystals in a preweighed Eppendorf tube. Close the tube and weigh it to calculate the quantity of the

uranyl acetate. Therefore, adapt the volume of acetone required to reach the correct dilution. A quantity of crystals equivalent to a matchstick tip weighs approximately 50 mg and has therefore to be dissolved in 50 ml of acetone.

4. The resin: the lowicryl HM20 resin (EMS) is a highly cross-linked acrylate-based embedding resin, providing low viscosity at low temperature. It is stored at 4 °C for few months and ready, meaning that no addition of any catalyzer or accelerator is required. Before use, it is precooled in the freeze substitution system or in a –80 °C freezer. The resin is photo-polymerized by wavelength (360 nm) with a LED UV lamp adapted to the freeze substitution system. Pay attention with resin, it is armful if inhaled, and causes skin and eyes irritation.
5. Freeze Substitution system: Leica EM AFS2 is a user-programmed freeze substitution and low temperature embedding system. It consists of a freeze substitution chamber with a stainless steel working platform equipped with an integrated LED UV lamp. The chamber is cooling down thanks to LN_2 stored in a Dewar vessel mounted on a mobile trolley. The programming is made with a control unit and program steps are displayed on a color screen. The AFS2 is equipped with an automatic reagent handling system EM Freeze Substitution Processor (EM FSP, Leica) which dispenses reagents, dilute media and resins from 100% stock solutions in containers.

2.2 Sectioning and Sections Harvesting

1. Plastic support on TEM grids: collodion (Parlodion) 2% in amyl acetate (EMS).
2. Carbon rods 3 mm diameter (EMS).
3. Ultra-microtome: Ultracut (Reichert).
4. Diamond knife 35° (Diatome Inc).
5. Finder grids with an alphanumerical code: Gilder Grids G200F2, in copper or nickel (EMS).
6. Carbon Coater: the Q150T Turbo-Pumped Carbon Coater (Quorom Technologies) is a compact turbomolecular-pumped coating system that can evaporate carbon from carbon rods (EMS) with a precise thickness control (quartz crystal oscillator). The grids are installed on a stage that can rotate for a homogeneous covering.

2.3 Fluorescence and TEM Acquisitions

1. Glass coverslips for fluorescence acquisition: long (24 × 60 mm) glass coverslips (Menzel-Gläser) are used. This size is required because these coverslips are used as slides and have to fit with the fluorescent microscope stage.
2. Fluorescence imaging was performed with a Nikon TI-E/B inverted microscope featuring an EMCCD camera (Ixon Ultra

DU897-ANDOR) and a HG intensilight illumination. Images were acquired using Nikon CFI Apo TIRF 100× (NA = 1.49) objective and Semrock filters sets for GFP (Ex: 482BP35; DM: 506; Em: 536BP40) and mCherry (Ex. 562BP40, DM593, Em.641BP75).

3. Heavy metals for contrast:

   Uranyl acetate: uranyl acetate (LauryLab) 5 % in water is stored at 4 °C, for months, protected from light.

   Lead Citrate: lead citrate, trihydrate (EMS) 0.3 % is prepared in 0.1 M NaOH and stored at room temperature.

4. Sodium hydroxide: ultra, pellets (Fluka).
5. Transmission electron microscope: Jeol-JEM 1400.

3 Methods

In order to study the nuclear organization of the Pol I transcription, we generated a collection of yeast strains bearing genetically engineered fluorescent tags on proteins of the Pol I transcription machinery (Pol I subunits, transcription factors, etc.). In order to establish preparative conditions for CLEM, i.e., that preserve both ultrastructure of cells and the fluorescence of tagged proteins, we performed experiments with a strain containing a reduced number of rRNA genes copies (25 instead of 150) and expressing the Pol I subunit Rpa190 in fusion with the GFP and the H2B histone in fusion with m-Cherry (*see* **Note 1**). The advantages of using this strain stem from the abundance of H2B and Pol I correlated with very strong unambiguous fluorescent signals in living cells, therefore increasing the chance to keep fluorescence after the long procedure to prepare sections for TEM analysis. Moreover, in this strain all the rRNA genes are transcriptionally active; therefore, the Pol I is highly concentrated on the rDNA in the nucleolus, while the histones are concentrated in the nucleoplasm (actively transcribed rRNA genes are largely devoid of histones molecules, [3]); therefore the green and red fluorescent signals are complementary in the nucleus allowing the clear identification of each domain and reducing the probability of confusion with cellular auto-fluorescence emanating, for example, from vacuoles (*see* **Note 2**).

3.1 Fixation and Resin Embedding

The unique nucleolus of *S. cerevisiae* is a heterogeneous electron dense crescent-shaped nuclear subdomain attached to the nuclear envelope (Fig. 1a); it is organized into distinct morphological regions (related to functional regions) defined by their fibrillar or granular appearance [4]. The nucleolus of yeast is sensitive to fixation conditions and its structure is only well preserved using methods without chemical fixation of cells. To minimize artifactual structural reorganizations within

the cell and to produce biological material embedded in resin blocks that can be sectioned and observed at room temperature, we developed cryo-fixation and cryo-substitution methods [5]. Numerous resins with various properties are available and should be tested since embedding is highly sample specific. For CLEM, resins should have besides basic features (low shrinkage and absence of specimen damage during cure, minimal compression of sections, relative absence of electron beam-induced section damage, etc.) the lowest auto-fluorescent properties without quenching fluorescence from genetically encoded fluorescence epitope (GFP and m-Cherry).

1. Harvest 1 ml of the cell culture in exponential phase growth (OD = 0.5 or 3.5×10^6 cells) (*see* **Notes 3** and **4**). Centrifuge 3 min at 800 rcf to pellet the cells and trash the supernatant. Resuspend the yeast pellet in 5–20 µl (according to the size of the pellet) prewarmed (42 °C) 1% Agarose Low Melting Point and 7% sucrose dissolved in 0.1 M cacodylate buffer (pH 7.2).
2. Immediately freeze the cells embedded in Agarose by high-pressure cryo-fixation (EMPACT, Leica) (*see* **Note 5**).
3. Transfer in liquid nitrogen the frozen pellets into glass distilled acetone containing 0.1% uranyl acetate at –90 °C in a programmed Freeze Substitution System equipped with an automatic reagent handling system. This machine controls temperature and incubation times and dispenses reagents for freeze substitution, resin infiltration, and resin embedding steps according to the following program:

48 h	0.1% uranyl acetate in acetone	–90 °C
10 h	acetone	increase gradually the temperature from –90 °C to –50 °C
24 h	HM20 30% in acetone	–50 °C
24 h	HM20 70% in acetone	–50 °C
12 h	HM20 100%	–50 °C
8 h	HM20 100%	–50 °C (*see* **Note 6**)
48 h	HM20 100%	–50 °C with UV for polymerization (*see* **Note 7**)
20 h	HM20 100%	increase the temperature from –50 °C to 20 °C (3.5 °C/h) with UV
24 h	HM20 100%	20 °C with UV

Store the blocks, shielded from light at RT.

3.2 Sectioning and Sections Harvesting

1. With an ultra-microtome, cut the blocks to produce sections of 100–200 nm thickness using an ultra 35° diamond knife.
2. Deposit the sections on collodion-carbon coated Finder grids (EMS) (*see* **Note 8**).

3.3 Fluorescence and TEM Acquisitions

1. For observation of the fluorescent signal (*see* **Note 9**), sink one grid in a water drop spread on a glass coverslip, paying attention that the sections are on the side in contact to the coverslip to allow correct focusing. To avoid any fluorescence background, the coverslips are treated (*see* **Note 10**).
   Do the observations using an inverted fluorescent microscope (*see* **Note 11**).
2. First, acquire the red fluorescence which corresponds to the strongest signal due to the abundance of the histones (*see* **Note 12**).
3. Then acquire the green fluorescence.
4. Finally, acquire the light transmitted image (conventional or interferential contrast for better contrast) at the same magnification and also at a lower magnification (or of a microscopic field a bit shifted) in order to locate the cells of interest using the alphanumerical code of the Finder grids (*see* **Note 13**).
5. Carefully, remove the grids from the drop and let it dry protected from dust.
6. Counterstain the sections with 12 min incubation in uranyl acetate 5% in water, shielded from light. Rinse extensively in milli-Q water. Incubate 30 s in lead citrate in NaOH saturated atmosphere (*see* **Note 14**). Rinse extensively in milli-Q water.
7. Evaporate 5 nm of carbon over the sections in order to spread the electron charges during TEM process (*see* **Note 15**).
8. Sections are analyzed with a Jeol-JEM 1400 electron microscope. With the help of the alpha numerical code, locate the region of interest where fluorescent signals were detected (*see* **Note 16**). Do acquisition at low magnification in order to be able to recognize patterns possibly formed by the cells distribution in the section. Also do acquisitions at high magnification of cells that were positive in fluorescence (*see* **Note 17**).

3.4 Superimposition of the LM and TEM Images

1. Align the pictures using particular areas visible both in LM and TEM (e.g., grids bars or collodion cracks, etc.) (*see* **Note 18**). In Fig. 2, two examples of microscopic fields imaged in TEM, fluorescence and Interferential contrast (DIC) are presented.

3.5 Conclusion

The CLEM protocol described here can be implemented in laboratories equipped with conventional fluorescence and electron microscopies facilities and does not require sophisticated "pipeline" for imaging. It is adapted to strong fluorescent signals. It requires manual and tedious exploration of the grids; semiautomation of samples screening and improvement of images alignment have still to be developed. Moreover, even if some patterns are clearly visible on fluorescent pictures of yeast sections, the conventional fluorescence microscopy cannot distinguish between objects

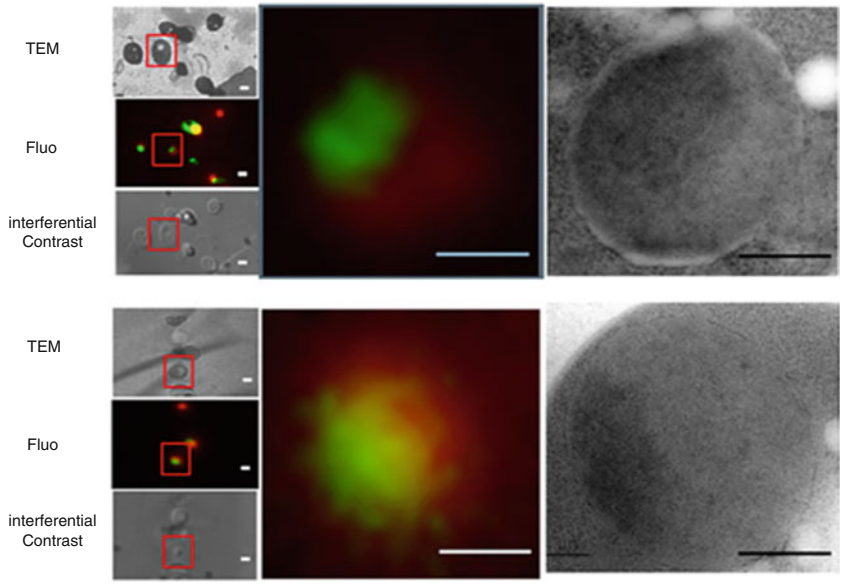

Fig. 2 CLEM—correlative light-electron microscopy. Sections of yeast strains expressing H2B histone fused to mCherry (*red*) and Pol I fused to GFP (*green*) are imaged by interferential- and fluorescence microscopy and by transmission electron microscopy (TEM). The *green* fluorescent signal corresponds to the GFP-Pol I which are concentrated in the nucleolus. The *red* fluorescent signal corresponds to mCherry-histones which are concentrated in the nucleoplasm. The same cells imaged by TEM providing the ultrastructural context of the fluorescent proteins. Scale bars: 1 μm

separated by less than 200 nm (resolution inherently limited by the wavelength of the light); enhanced resolution is required to precisely relate fluorescent signal to nucleolar subcompartments. Therefore, the challenge is now to develop CLEM method making use of advanced high resolution fluorescent imaging. The group of A. S. Frangakis in Frankfurt (Germany) published CLEM of a tagged nucleolar protein using Direct Stochastic Optical Reconstruction Microscopy (STORM) [6]. We are currently developing CLEM for analysis of nucleolar transcription with PALM. PALM (PhotoActivated Localization Microscopy) is based on the serial photoconversion and localization of hundreds to thousands of individual molecules to build an image at near-molecular resolution [7]. For this "high resolution CLEM procedure," our group is establishing a collection of yeast strains producing proteins of the Pol I machinery in fusion with photoactivable tags (mEos2) adapted for detection of single molecules. By correlating the PALM images of Pol I machinery distribution with electron microscope images of the nucleolar structure, it will be possible to decipher if the observed rods on TEM micrographs (Fig. 1a) could correspond to individual rRNA genes. CLEM with PALM would provide new possibilities to understand how individual transcribed genes are organized in the volume of the nucleolus.

4 Notes

1. The genotype of the strain we are using is as follows: ade2-1 ura3-1 his3-1,15, trp1-1 leu 2-3,112 can1-100 fob1D::KAN-MX4,rDNAcopies25,RPA190::GFP-URA3,HTB1::mcherry-HIS3. HTB1 is the gene coding the H2B histone. GFP (Green Fluorescent Protein) has a major excitation peak at a wavelength of 488 nm (blue) and its emission peak is at 509 nm (green). m-Cherry has a major excitation peak at a wavelength of 587 nm (green) and its emission peak is at 610 nm (red).
2. The vacuole is in yeast sample a potential problem for fluorescence detection. Indeed, the vacuoles contain pigments that can be excited at certain wavelengths and therefore generate a fluorescent signal. The number, the size, and the composition of the yeast vacuole vary from one strain to another and depend on the physiological state of the cells (growth phase, cell cycle phase, nutrient starvation, carbon source, etc.). Note also, that we generally add an excess of adenine in the growth medium to inhibit formation of metabolic intermediates that are pigmented and can also be a source of auto-fluorescence.
3. The nucleolus is the morphological counterpart of the ribosome biogenesis of the cells and its ultrastructure is therefore closely correlated to this activity. The harvesting of cells if they are not in exponential growth phase can give completely different results.
4. Before any CLEM experiments, one should of course check the intensity of the fluorescence in living cells!
5. To ensure a good morphological preservation, the decrease of temperature has to be around 10,000 °C/s and the pressure is 2000 bars. In such conditions, the water from the sample is transformed in amorphous (vitreous) ice because ice crystals that cause damage to the specimen's ultrastructure are unable to form [8]. After cryofixation, cryomicroscopy of vitreous sections is possible but requires high skills and specific expensive microscope. We chose to substitute at low temperature the ice with chemical fixative in order to be able to work with samples at room temperature. Sectioning as well as imaging with "conventional" electron microscopy is much easier. Moreover, cryofixation/cryosubstitution provided very good quality of images of the yeast nucleolus [4, 5]. Finally, the correlation of the TEM picture with the fluorescent image is also much easier at room temperature.
6. Long infiltration times guarantee a good quality of embedding.
7. Other resins can be used for embedding samples dedicated to CLEM. Indeed, some fluorescent tags are quenched by long UV exposition. When using mEOS2 photo convertible tags (that are

adapted to do high resolution fluorescent microscopy like PALM), the GMA (EMS) resin that can polymerize at –20 °C without UV by adding *N*,*N*-dimethyl-*p*-toluidine is a good alternative [9].

8. Without plastic support, the sections are not stable under the electron beam. Therefore, the grids are covered with a plastic film of collodion according to the following procedure. Fill a crystallizer of milli-Q water up to overflowing. Let fall a drop of collodion on the meniscus of water. When contacting the water, the collodion polymerizes. With a Pasteur pipette, remove the collodion film and trash it: this step will clear the water of any dust contaminant. Let fall another drop of collodion; when the collodion has polymerized, deposit carefully as many grids as possible on the plastic film. Thanks to parafilm or filter paper, recover all the collodion covered grids at the same time. Dry and store them shielded from dust. Handle collodion carefully because it forms toxic gases and can cause skin, eyes, respiratory, and central nervous system irritation. However, polymerization of the collodion under a fume hood is not advised because turbulences can cause wavy collodion films. When dried, the collodion can be covered with 5 nm of evaporated carbon/graphite (graphite is the softest stable form of carbon). Because of the low molecular weight and unique structure where atoms all lie in a plane, graphite films have excellent mechanical stability; moreover, carbon is able to disperse electron charges under the microscope beam. Therefore, carbon coating is increasing the resistance of the support.
9. Observations in fluorescence have always to be done before observations in TEM because the electron beam bleaches fluorophores.
10. To avoid fluorescence backgrounds, the coverslips are treated twice with sequential washes in methanol, acetone, and milli-Q water. The coverslips are then plunged in 1 M KOH and treated with ultra sounds for 1 h. Coverslips are finally rinsed and stored in milli-Q water at 4 °C.
11. We never succeeded to perform CLEM with grids mounted between slide and coverslip (while that would have been adapted for observation with a straight microscope, more common in laboratories). The problem is to take apart the coverslip without damaging and cracking the collodion–carbon support and the sections. Contrariwise to published work, we were also not able to do CLEM with grids mounted in glycerol (even at low concentration); extensive washings did not remove fully the glycerol that impairs TEM observation.
12. The acquisition time required to detect fluorescent signal from sections is drastically increased (×30) when compared to living cells. This decrease in signal intensity is not only

due to the reduction of the quantity of the tagged protein but also to the degradation of the fluorescent signal during the procedure for preparation of the sample for TEM. Indeed, considering a 200 nm thick section of a nucleus with a radius of 0.87 μm (average radius of yeast nucleus measured in living cells) (R. Wang, LBME, unpublished results), the maximal volume of the sectioned nucleus is only three times smaller than the volume of the full nucleus. The protocol here described is probably not adapted for detection of low fluorescent signal.

13. The acquisitions at the fluorescent microscope have to be performed as fast as possible (or one has to take care to regularly add water on the grid), because the drop of water is drying by heating under the fluorescent lamp.
14. When staining with lead citrate, do so in a closed container with sodium hydroxide pellets in order to absorb atmospheric CO_2. Indeed, the lead citrate forms water-insoluble precipitate (lead carbonate) if not used under CO_2-free conditions.
15. Because carbon is uniformly amorphous and highly transparent to electron beam, it can be used to protect the sample during the TEM observation. Coating the sections with 5 nm of carbon will stabilize them under the electron beam because, as explained in **Note 8**, carbon can spread the electron charges.
16. Unfortunately, it often happens that due to the repeated handling of the grids, the collodion film breaks; therefore some regions of interest can have "disappeared" between the acquisition of the fluorescent signal and the TEM acquisition. We multiply the number of acquired regions and the number of grids to bypass this problem.
17. We often print the photo of a region of interest in light microscopy (DIC) and indicate on this print which cells are positive for fluorescence. This print is used when acquisitions are performed on the TEM to try to recognize the cells. The alphanumerical code is essential to spot the region of interest, but often, the way cells are distributed in the section finally enables to recognize them. The manual screening of the grids is long and tedious. We are presently developing a semiautomated protocol using ImageJ macro to improve this step of exploring regions of interest.
18. In parallel with the implementation of high resolution fluorescence microscopy instead of conventional microscopy, the alignment step of the LM and TEM photos has to be improved. The deposit of fiducial markers that can be visualized with LM and TEM will increase the superimposition accuracy.

Acknowledgement

This work is supported by ATS-Nudgene and Emergence-CLEMgene of the Toulouse-IDEX. O.G. and C.N. are supported by Agence Nationale de la Recherche (ANDY). TEM images were taken with the facilities on the METI platform, part of TRI-Genotoul.

References

1. Miller O, Beatty B (1969) Visualization of nucleolar genes. Science 164:955–957
2. Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P (2013) RNA polymerase I structure and transcription regulation. Nature 502:650–655
3. Merz K, Hondele M, Goetze H, Gmelch K, Stoeckl U, Griesenbeck J (2008) Actively transcribed rRNA genes in *S. cerevisiae* are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. Genes Dev 22:1190–1204
4. Léger-Silvestre I, Trumtel S, Noaillac-Depeyre J, Gas N (1999) Functional compartmentalization of the nucleus in the budding yeast Saccharomyces cerevisiae. Chromosoma 108:103–113
5. Léger-Silvestre I, Noaillac-Depeyre J, Faubladier M, Gas N (1997) Structural and functional analysis of the nucleolus of the fission yeast *Schizosaccharomyces pombe*. Eur J Cell Biol 72:13–23
6. Perkovic M, Kunz M, Endesfelder U, Bunse S, Wigge C, Yu Z, Hodirnau VV, Scheffer MP, Seybert A, Malkusch S, Schuman EM, Heilemann M, Frangakis AS (2014) Correlative Light- and Electron Microscopy with chemical tags. J Struct Biol 186:205–213
7. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF (2006) Imaging intracellular fluorescent proteins at nanometer resolution. Science 313:1642–1645
8. Kanno H, Speedy RJ, Angell CA (1975) Supercooling of water to -92°C under pressure. Science 189:880–881
9. Paez-Segala M, Sun MG, Shtengel G, Viswanathan S, Baird MA, Macklin JJ, Patel R, Allen JR, Howe ES, Piszczek G, Hess HF, Davidson MW, Wang Y, Looger LL (2015) Fixation-resistant photoactivatable fluorescent proteins for CLEM. Nat Methods 12:215–218

Chapter 4

High-Throughput Live-Cell Microscopy Analysis of Association Between Chromosome Domains and the Nucleolus in *S. cerevisiae*

Renjie Wang, Christophe Normand, and Olivier Gadal

Abstract

Spatial organization of the genome has important impacts on all aspects of chromosome biology, including transcription, replication, and DNA repair. Frequent interactions of some chromosome domains with specific nuclear compartments, such as the nucleolus, are now well documented using genome-scale methods. However, direct measurement of distance and interaction frequency between loci requires microscopic observation of specific genomic domains and the nucleolus, followed by image analysis to allow quantification. The fluorescent repressor operator system (FROS) is an invaluable method to fluorescently tag DNA sequences and investigate chromosome position and dynamics in living cells. This chapter describes a combination of methods to define motion and region of confinement of a locus relative to the nucleolus in cell's nucleus, from fluorescence acquisition to automated image analysis using two dedicated pipelines.

Key words Nucleolus, FROS, Nuclear organization, Live-cell imaging, Fluorescence microscopy, Chromosome domain dynamics, Yeast, *Saccharomyces cerevisiae*

1 Introduction

Nuclear organization is investigated using a large panel of advanced molecular biology and imaging techniques. Chromosome conformation capture (3C) and derived methods are more and more popular to explore 3D organization of genomes [1]. However, contact frequencies obtained using 3C-derived methods are indicative of the frequency at which sequences are ligated together by cross-linking [2]. High frequency of product captured correlated strongly with spatial proximity but should be combined with direct imaging observation to exclude local discrepancy [3]. Thanks to the fluorescent labeling of chromosome loci in living cells, advanced imaging techniques can also provide considerable amount of data describing the spatial organization of chromosomes. Fluorescent labeling in living cells is mostly performed using Fluorescent

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_4,

operator-repressor system (FROS) fluorescent repressor-operator system, which combines the expression of a bacterial repressor fused to a fluorescent protein and the integration of operator sequence as tandem arrays at a specific locus (*see* **Note 1**). A unique FROS tagged locus appears as a bright fluorescent spot in the nucleus. However, to be informative, imaging data must be quantitative, and "high throughput" to generate statistically robust data comparable with 3C-derived dataset. Here, we describe two types of analysis to explore motion and position of a locus relative to the nucleolus. Motion addresses local biophysical properties of chromatin, while position gives information about confinement (gene territory within chromosome territory) and long-range nuclear organization [4, 5].

Locus motion can be followed using time-lapse imaging of fluorescently tagged loci. Time-lapse imaging can be used over broad time scales (from 0.2 to 400 s time intervals) and then the images analyzed by a dedicated image analysis system [6] which can be used to extract trajectories of loci in yeast cell nucleus with high precision (below 30 nm). Our previous work had explored the dynamics of chromosomes relative to the nucleolus [7, 8]. The nucleolar association of loci is evaluated using a double detection of loci and nucleolar region in the nucleus. When chromatin is out of the nucleolus, we propose that chromatin motion is described by the Rouse polymer model, which assumes that chromatin fiber behaves as a homogeneous series of beads connected by elastic springs [7]. The motion of nucleolar associated locus in yeast is not compatible with such model and remains largely uncharted [7, 8].

Trajectories are classically analyzed using mean-square displacement (MSD). For every time interval (s), the square of the mean displacement observed (μm^2) can be calculated. Confinement of a locus in areas of the nucleus appears as a plateau at longer time scale, defining a radius of constrains (Rc, expressed in μm) (typically measured at 200 s) [4]. However, for such long intervals, the number of distance measurements decreases, which increases the standard deviation of the measured Rc. Therefore, Rc only approximates the volume of the nucleus in which loci are confined. Typically, the explored volume for different loci (0.57–1.5 μm^3) can be determined using spherical approximations [9]. How to explore accurately association of the locus to the nucleolus? To analyze the spatial location of a given locus in yeast nucleus, 'Nucloc' algorithm determines, in each cell nucleus of a population, the three-dimensional position of the locus relative to the nuclear envelope, and the nuclear center and the nucleolus as landmarks [10]. Automated detection "pipeline" allows high-throughput analysis of a large number of cells (>1000). The 'Nucloc' algorithm [10] can create genemap: a high-resolution cumulative percentile map of subnuclear domains occupied by individual loci in a population of cells, thereby defining the domain

in the nucleus in which locus is confined: the gene territory. Importantly, the dimensions of a gene territory, as defined by the volume in which 50% of the gene positions are detected in statistical maps, are compatible with explored volumes estimated using MSD (from 0.57 and 1.5 μm^3 using MSD to 0.8 and 1.2 μm^3, respectively, using gene territory) [8]. Therefore, motion analysis and high-resolution cumulative percentile maps of subnuclear domains occupied by individual locus are nice complementary methods.

Here, we present a detailed protocol to determine the motion of a FROS tagged locus relative to the nucleolus and explore the shape and the position of its confinement area. This protocol describes sample preparation, image analysis, and visualization of data generated. The protocol includes a specific focus on validation procedure of imaging data prior to automated processing and on controls required to assess quality of data collected.

2 Materials

2.1 Cell Culture

1. YP medium (1% w/v yeast extract, 2% w/v peptone) complemented with required sugar source (glucose, galactose, raffinose) 2% w/v each.
2. Synthetic medium (SC): Due to high autofluorescence of yeast extract in most optical setup, cells are washed and mounted on slide using synthetic medium (SC) 0.67% nitrogen base w/o amino acids, 2% w/v dextrose supplemented with amino acids mixture supplemented with adenine (20 mg/L).
3. Molecular biology grade agarose (2% w/v) is used to immobilize cells for microscopic inspection. We prepare an agar pad using SC media.
4. "VaLaP" is a mixture of 1/3 vaseline, 1/3 lanoline, and 1/3 paraffin which is used to seal coverslip.
5. Slide (Knittel Glass, Germany, 76×26 mm).
6. Coverslip (Knittel Glass, Germany, 22×22 mm).

2.2 Live-Cell Fluorescence Microscopy for Time-Lapse Imaging

1. An inverted microscope (Ti-E/B; Nikon).
2. A system for long-range time-lapse experiments (Perfect Focus System).
3. Filters tubes (e.g., Semrock; GFP (Ex: 482BP35; DM: 506; Em: 536BP40) and mCherry (Ex. 562BP40, DM593, Em.641BP75)).
4. EM CCD camera (e.g., DU-897; Andor Technology).
5. 100× oil immersion objective lens with variable NA (e.g., CFI Plan-fluor 100×, NA 0.7–1.30, Dt0.2) (*see* **Note 2**).

2.3 Live-Cell Fluorescence Microscopy for Determination of 3D Gene Position in Cell Population

1. A spinning disk confocal system (e.g., Andor Revolution Nipkow-disk confocal system).
2. A CSU22 confocal spinning disk unit (Yokogawa).
3. EM CCD camera (e.g., DU888, Andor).
4. 100× oil immersion objective lens (e.g., Plan APO, 1.4 NA, Olympus).
5. Diode pumped solid state lasers (e.g., DPSSL; at 488 nm (50 mW, Coherent) and mCherry fluorescence at 561 nm (50 mW, CoboltJive)).
6. Filters (e.g., Semrock bi-bandpass emission filter, Em01-R488/568-15).

2.4 Image Analysis Workstation

1. Tracking gene motion: Executable file derived from multiple-target tracing (MTT) [11] is freely available upon request [6]. This executable file used is compiled from Matlab application. Therefore, the compiled executable file can be used on image analysis workstation without Matlab installed. A freely accessible application named component Runtime (MCR; a stand-alone set of shared libraries to run compiled MATLAB applications) should be installed.
2. Determination of gene position: 'Nucloc' is a suite of Matlab programs, requiring Matlab (Mathworks) to be used (tested with releases R2007b and R2008a) including curve fitting, image processing optimization, signal processing statistic, and wavelet toolbox. The suite is freely available upon request (http://www.nucloc.org/) [10].

3 Methods

3.1 Preparation of the Samples

To determine locus motion and position relative to the nucleolus, we start our experiments with a strain bearing three fluorescent proteins (TetR-GFP, GFP-Nup49, and mCherry-Nop1) (No-FROS strain; *see* **Note 3**). Insertion of a transgene bearing high number of operator (250 repeats) results in individual yeast clones with various number of tetO repeats insertion in the genome (from about 50 to 250 repeats) (*see* **Note 4**). For each targeted locus, we prepare for imaging from two to four individual FROS labeled clones (the quality of the acquired images are evaluated postacquisition, *see* Subheading 3.3). For image acquisition, we need a device allowing long-term growth but maintaining cells immobile close to a microscopic coverslip. We immobilize the yeast cells and distribute them in a single cell monolayer (*see* **Note 5**).

1. Preculture the strain at 30 °C in YP media containing 2 % w/v glucose and grow overnight.
2. Dilute cells at 10^6 cells/mL in 50 mL YP glucose media. For some mutant strains, culture condition should be modified to avoid autofluorescence (*see* **Note 6**).
3. Incubate at 30 °C till concentration reaches 4×10^6 cells/mL (at least two doublings after dilution).
4. Collect 1 mL cultured cells and centrifuge $1000 \times g$, 3.5 min at room temperature (low speed centrifugation is used to preserve nuclear organization). Additional steps are required when preparing fixed cell samples (*see* **Note 7**).
5. Rinse cell pellet twice with 1 mL of SC media complemented with suitable sugar.
6. Resuspend cells and concentrate in 5–10 µL SC media.

   Steps 7–11 can be done during cell culture (**step 3**).

7. Prepare 20 mL 2% w/v agarose in SC media containing 2% w/v of sugar. Melt agarose slowly in microwave oven (defrosts program for several minutes). An Erlenmeyer in water filled beaker is used. Melt agarose is kept in 50 °C water bath before pouring pads.
8. Align side by side three microscopy slides on the bench and stick with laboratory tape on the two slides on each side. The one in the middle, that will be used to pour pad, can be freely moved. The tape thickness will define the thickness of the agarose pad. For long time imaging, we increased thickness with a second tape.
9. Drop 80 µL (thickness 2 tapes) or 40 µL (thickness 1 tape) of melted agarose on the central glass slide.
10. A regular slide is immediately placed across the top to create a flat surface on the pad. Press against the top slide for 1 min (*see* **Note 8**).
11. Recovering the slide is the tricky part. Remove the central slide with the cross slide. Slowly move the two slides against each other without damaging the pad. A slide with a pad can be stored in wet atmosphere for 1–2 h.
12. Place 3 µL of the concentrated cells on the pad and cover by a regular coverslip. Gently press to distribute the cells into a monolayer.
13. Seal the coverslip with "VaLaP". "VaLaP" is melted at 150 °C and applied with a cotton swab.

    Live microscopy with such agar pad should be limited to 20 min after mounting, at temperature from 20 to 37 °C.

3.2 Image Acquisition

1. For observation of a gene motion, image acquisition was performed on wide-field microscope equipped with a single photon sensitivity camera. Time interval is usually set to 200 ms but can vary from 20 ms to 10 s depending on the time range to be analyzed.
2. For observation of the gene territory, microscopy was performed with a spinning disk microscope with a sensitive camera with large field of view. For 3D analysis, Z-stacks of 41 images with a 250 nm Z-step were used. Exposure time was 200 ms.

3.3 Validation of the Acquired Fluorescent Images

Gene motion and gene territories are both automated imaging pipeline, allowing high-throughput image analysis. Automation has a clear cost: without proper calibration, it will generate meaningless dataset. Therefore, such optimized algorithms require a well-calibrated and homogeneous fluorescent signal in the cell population (*see* **Note 9**). In case of low signal-to-noise ratio (SNR) for FROS labeling, detection of locus can be inaccurate (*see* **Note 10**). Furthermore, automatic detection of FROS tagged locus can generate meaningless dataset (*see* **Note 11**). Postacquisition control must be performed before running automated image analysis for time-lapse imaging (tracking gene motion) or confocal stack (gene position):

1. Evaluate qualitatively the fluorescent signal obtained in each FROS tagged clone (*see* **Note 12**).
2. From a representative set of acquisition (10–20 cell's nucleus), measure locus intensity (I_{max}), background mean (I_{noise}), and standard deviation (σ_{noise}) intensity (Fig. 1).
3. Determine SNR (dB) of your acquisition (*see* **Note 13**).

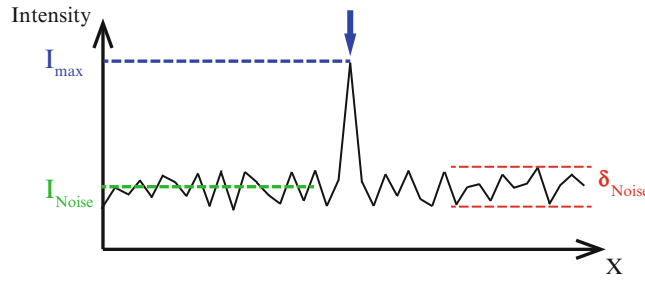

Fig. 1 Signal-to-noise ratio determination. One-dimensional representation of an intensity plot is shown. A point-like fluorescent object (*blue arrow*) is detected. The signal-to-noise ratio depends on the intensity of the signal (I_{max}), the background intensity (I_{noise}), and the variance of the background. The signal-to-noise ratio can be estimated using two different equations, which give different results: $SNR = (I_{max} - I_{noise})/\delta_{noise}$ or $SNR = (I_{max} - I_{noise})/\sqrt{I_{max}}$

3.4 Image Analysis and Visualization of Data

3.4.1 Tracking Gene Motion

To determine trajectory with high resolution, we use a freely accessible executable script [6]. This script is based on the algorithm Multiple-Target Tracing (MTT), an advanced version of SPT [11]. MTT's derived executable file can generate dynamic maps at high densities of tracked loci, thereby providing global representation of chromosome dynamics in nucleus. In MTT, the detection, estimation of positions, and reconnection are done sequentially.

1. Detect the position $r(t)$ of a fluorescently labeled locus in each frame of a time-lapse movie (Fig. 2a) (*see* **Note 14**).
2. Evaluate extracted trajectories (*see* **Note 15**). Raw measurement is the sum of true motion and experimental errors. Visual inspection of the tracking allows rejection of aberrant tracking. User should also evaluate microscopic slide drift, easily detectable by correlated motion of all genes in the frame of view.
3. Determine the contribution of nuclear motion to gene trajectories (*see* **Note 16**).
4. Estimate accuracy of extracted trajectories (*see* **Note 17**).
5. Compute ensemble-averaged MSD from high quality trajectories (Fig. 2b, *see* **Note 18**). To identify possible heterogeneity in the data, rather than mean displacement, step distribution for a given time interval can also be informative [12]. The loci in nucleolus have a different behavior compared with the loci on the same chromosome but in nucleoplasm (Fig. 2c vs d).

3.4.2 Determination and Validation of Genemap

1. Detect cells by importing the confocal stack of images (Fig. 3a). Individual round nuclei (*see* **Note 19**) are detected and cropped as region of interest (ROI). Automation of the script allows processing of about 1000 nuclei without user input. Cells also can be selected manually.
2. 3D localization of FROS position is automatically performed by detection of chosen landmarks (one or two tagged loci; NPC labeling and nucleolus) (Fig. 3b).
3. Quality control of extracted distances is performed (*see* **Note 20**).
4. Evaluate position of loci in cell nucleus. At this stage, distances distribution can be generated (locus-NE, locus-Nuclear center, Locus-Nucleolar centroid) (Fig. 3c).
5. Alignment is computed to generate probability map (Fig. 4). After alignment (Fig. 4a), loci are projected in two-dimension (2D) by cylindrical projection, preserving distance to nuclear and nucleolar center (*see* **Note 21**).
6. Generate the 2D locus probability map in color-coded histogram (Fig. 4b). Rather than histogram, we now use kernel density representation (Fig. 4c), which allows a smooth visualization of color-coded locus probability density map and does not rely on arbitrary histogram bin [13] (*see* **Note 22**).

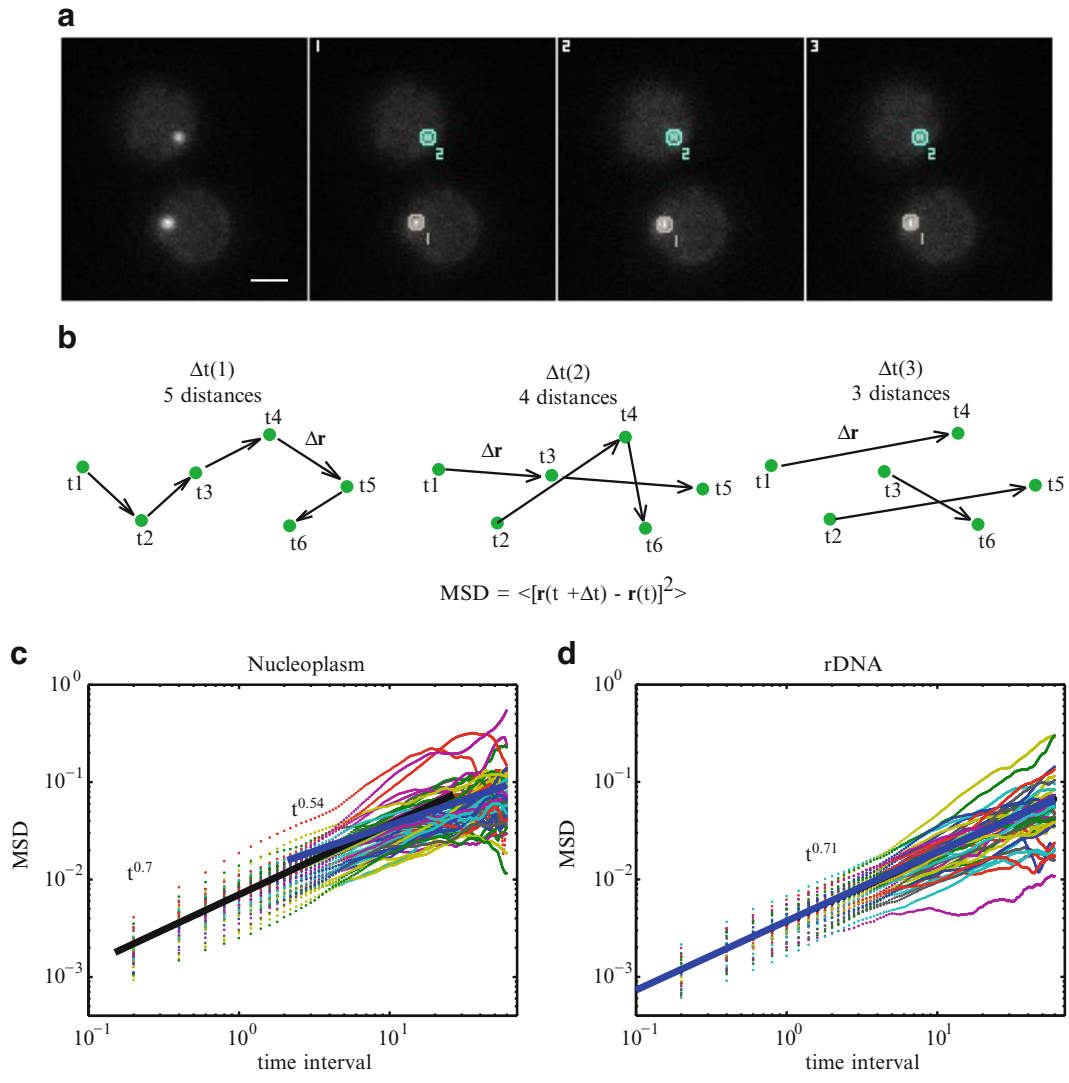

Fig. 2 Mean square displacement (MSD) calculation and interpretation. (**a**) Locus position determined using an MTT-derived algorithm. *Left* is the initial fluorescent image, which is followed by images taken at 200 ms intervals in time-lapse imaging. Scale bar 1 μm. (**b**) Robust MSD calculation requires high numbers of data points. From six positions, detected from t1 to t6 by MTT, the MSD value at $\Delta t(1)$ can be calculated from the mean of five distances. For $\Delta t(2)$ and $\Delta t(3)$, MSD values are the mean of four and three distances, respectively. Robustness of MSD decreases when time intervals (Δt) are increasing. (**c**) MSD extracted from 45 trajectories for one selected FROS labeled locus (position 380 kb on chromosome XII). The *black* and *blue solid lines* represent the average MSD of these trajectories. Motion of loci in the nucleoplasm scale with an exponent of 0.54. (**d**) MSD extracted from 41 trajectories for FROS inserted in rDNA. Blue line represents average MSD. The motion of loci in the nucleolus scales with an exponent of 0.7, which is clearly different from the motion of loci in the nucleoplasm

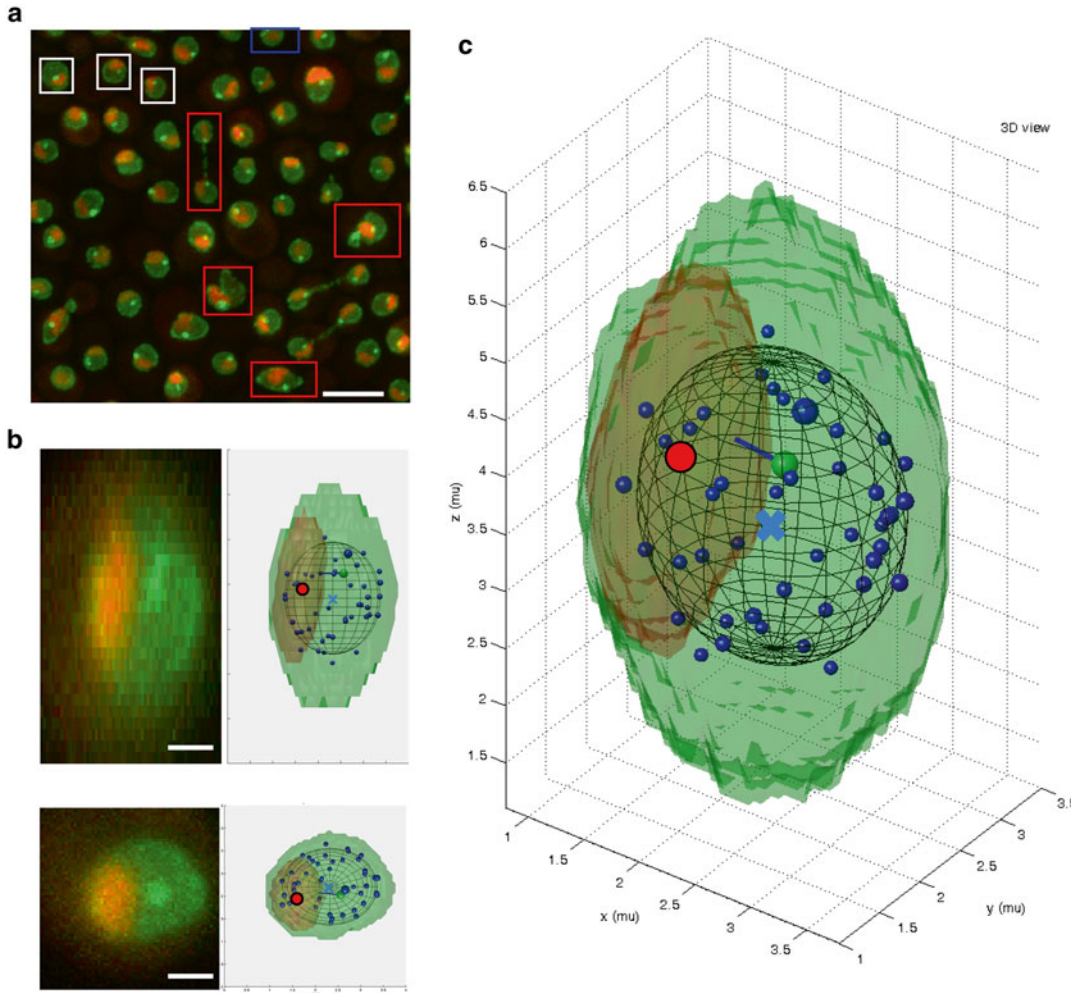

Fig. 3 Determination of three-dimensional position of gene and nuclear landmarks: nuclear center, nuclear periphery, and nucleolus. (**a**) Selection of interphase cells can be performed manually or automatically. Selected nuclei (*white box*) are round and correspond to G1/S phase. The nucleus at the edge of the field of view (*blue box*), mitotic cells, and cells with abnormal shape are excluded (*red boxes*). (**b**) X-projected (*upper panel*) or Z-projected (*lower panel*) fluorescent stack (*left*) and extracted position (*right*). The 'Nucloc' package extracts point-like structures (gene-*green* and NPC-*blue*) and performs thresholding of GFP and mCherry signal used, respectively, to determine nuclear center (*blue cross*) and nucleolar centroid (*red circle*). (**c**) 3D representation of the detection output of 'Nucloc'

4 Conclusion

Both methods give access to different chromatin properties and can be combined using the FROS tagged strains. The robustness and accuracy of the datasets generated by our automated gene motion tracking pipelines are of high quality. Confinement can be best explored using Genemap visualization (Fig. 4), which allows determination of gene territories (50% of distribution). Possible interpretation of such data is still the topic of extensive discussion.

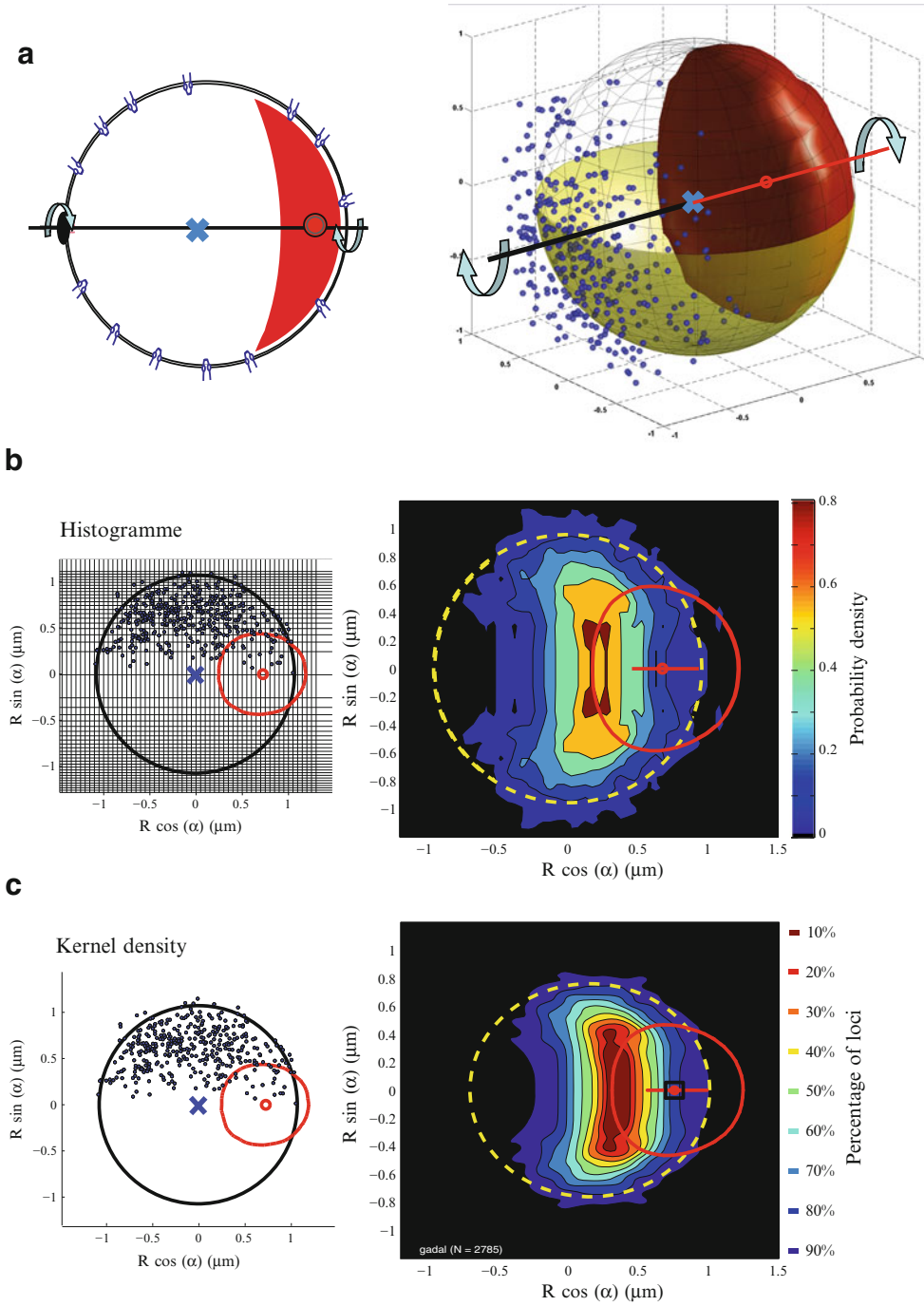

Fig. 4 Aggregation of individual detection in a single probability density map. (**a**) The yeast nucleus is organized around a central axis connecting the nuclear and nucleolar center. (**b**) Alignment of all individual detections. All nuclear centers are translated to the origin (*blue cross*). All nucleolar centers are aligned on *X*-semiaxis (*red* segment). All detected loci (*blue sphere*) are represented in an aggregated 3D map. The median nucleolus (*red shape*) and median nucleus (*yellow circle* with a *dashed line*) are also shown. Rotation around the central axis (*blue arrows*) does not modify distances to nuclear center or nucleolar center. (**c**) Cylindrical projection of genes around the central axis and histogram representation. Gene positions in a 2D map are calculated corresponding to an iso-volume (projected volume per surface unit scale quadratic along the *Y*-axis). A color-coded histogram is determined (*right panel*), in which the colors depict probability density. The kernel distribution does not rely on a histogram, but on kernel density, which allows a smooth representation of the data. The colors indicate the percentage of loci in the enclosing contour

We have suggested that the motion of the locus in the nucleoplasm can be analyzed using the Rouse model (Fig. 2c) [6, 8]. The analysis of MSD is further described in a recent review [7]. Briefly, in a short time interval, the motion of the locus is not restricted by elastic interactions with its neighboring polymer segments, leading to a regime of free diffusion. The second regime corresponds to the Rouse model of polymer dynamics, in which the MSD increases with time with a power-law scaling of 0.54. However, alternative models such as fractional Brownian gene motion in the nucleus have also been proposed [14, 15].

5 Notes

1. The FROS system includes TetR/*tetO* [16], LacI/*lacO* [17], λcl/*λOp* [18]. Genome biology is affected by FROS insertion, and biological consequences should be tested before and after insertion [19]. FROS insertion can affect position of the tagged gene [20], expression of the neighboring gene, and generate artificial tethering between two tagged loci [19]. Due to close proximity of 5′ or 3′ regulatory elements in yeast genome, we insert FROS in both 3′ and 5′ of targeted locus (in 100–800 bp) to explore possible side effect. Nonrepetitive array using ParB-INT appears as an alternative system for gene labeling [21].
2. For fast time-lapse imaging, a single acquisition at each time point (no Z-step) is chosen to shorten time intervals. However, FROS locus can move out of the focal plan. Depth of field is small when using high numerical aperture (NA), typically 1.4. Using variable NA objective (from 1.3 to 0.7), depth of field can be increased. At the cost of a reduced collection of light, single Z-step using NA 1.25 strongly increases trajectory length (with such a setup FROS loci stay in focal plan).
3. Nuclear envelope position is determined using Nup49, an abundant nuclear pore complex (NPC) protein, fused to GFP. Nucleolar volume and position are extracted thanks to the nucleolar protein Nop1, the yeast ortholog of fibrillarin, fused to mCherry. Before tetO insertion, TetR-GFP is distributed throughout the nucleus and appears as a diffuse nuclear staining, which allows (together with GFP-Nup49) a robust determination of the nuclear center. This strain has no loci labeled and must be used as negative control for fluorescent imaging (No-FROS).
4. Methods for tetO insertion including plasmid preparation, yeast transformation, and verification of sequence integration sites have been described previously [18]. We perform an additional pulsed-field gel electrophoresis (PFGE) step to check

possible chromosome abnormalities observed at low frequency (5–10 % of clones) after FROS insertion.

5. Multiple designs exist for live yeast cell imaging. The simplest (agar pad—presented here), concavity slide (Electron Microscopy Science; 1.4–1.6 mm thick), or adhesive gene frame (ABgene; 1.7 × 2.8 cm) allows cells to grow in monolayer for up to 30 min. Longer inspection requires microfluidic device, commercially available (CellASIC™ ONIX Microfluidic Live Cell) or homemade PDMS device, to flow fresh medium in the microfluidic chamber during inspection.
6. Autofluorescence (GFP or mCherry channel) is low in exponentially wild-type yeast cells. Some stress condition (starvation, oxidative stress, aging) can generate autofluorescence and should be carefully investigated. Well-known mutations (*ade1-* or *ade2-*) present in numerous strain backgrounds lead to the accumulation of a red dye in the vacuole generating bright autofluorescence. Such genetic background should be excluded, or large amount of adenine in growth medium should be included to decrease autofluorescence in the vacuole.
7. Chemical fixation (1 % formaldehyde for 15 min, directly in culture medium buffered at 100 mM final concentration with Tris–HCl, pH 8.0 to preserve fluorescence). Glycine is used to neutralize formaldehyde (125 mM final concentration).
8. Preparation of an agar pad was described for DIC observation of *C. elegans* [22]. Place the two taped slides with a clean slide between them on the bench. Place the drop of agar + SC onto the clean slide. Cover the agar with another clean slide placed on top of the three slides. Press gently so that the agar drop is flattened to a circle (the thickness of the tape spacers). Avoid getting bubbles. After the agar solidifies, separate the two slides covering the agar pad by sliding one relative to the other. The agar pad should adhere to one of the slides (usually the bottom one). The agar pad can be kept in a humid chamber for 1–2 h.
9. Automated detection based on Gaussian fitting is typically appropriate for high SNR. Other detection methods (pointpicker, spotdistance, spot tracker, Nemo; see for review [18]) might be more robust and should be explored when SNR cannot be increased.
10. Strong SNR is instrumental to decrease false-detection rate and increase the chance to accurately acquire long trajectories and accurately determine gene territories. Locus motion is analyzed using single particle tracking (SPT) techniques. SPT is based on the ability to accurately identify the center of a fluorescent spot at each time point of time-lapse imaging (2D or 3D) and connect those positions to generate trajectories.

Imaging an isolated fluorescent spot and determining its center is defined as "localization microscopy." Spatial resolution of localization microscopy is not limited by diffraction limit (200–300 nm), but by the signal-to-noise ratio (SNR) which depends on the number of photons detected and the noise generated by microscopic setup and detector (also defined as black noise). Resolution of gene trajectory, when SNR is properly controlled, can make accessible motion below 10 nm (J. Mathon, personal communication). High false-detection rate, followed by automatic reconnection will generate apparent trajectory not related to gene motion. Like the SPT algorithm, 'Nucloc' is based on "localization microscopy" and relies on identifying point-like structures with an estimated resolution of about 30 nm.

11. One important parameter (and source of error) in the detection is that the brightest GFP fluorescent spot is assumed to be the FROS tagged locus. In most of our strains, GFP is used for both TetR-GFP, recognizing tetO at labeled locus and NPC (GFP-Nup49). In normal condition, NPC have the tendency to cluster [23], which can be strongly exacerbated in NPC mutant background (*nup133*Δ *nup120*Δ). Along the same line, TetR-GFP is homogeneously distributed in the nucleus in most cells, but can form local concentration or aggregates in some cells.

12. Individual transformants should be screened for signal homogeneity of FROS tagged locus. Confocal stacks are processed after maximum intensity projection along *Z*-axis. The tagged gene should be the brightest object of the nucleus (about fivefold over the brightest NPC spot detected). After integration, FROS labeling is stable on petri dishes for months, but might change once propagated from frozen stock. When recovered from frozen stock, individual clones should be rescreened. Very high number of integrated tetO repeats will result in heterogeneous labeling and/or toxicity in the population. When testing novel growth condition (carbon source, different temperature, stresses, etc.), always use No-FROS strain as negative control.

13. SNR is often expressed in decibel (dB). dB is a dimensionless logarithmic unit which express the ratio of two values: $10 \times \log_{10}(P_1/P_0)$. Our calibration showed that SNR > 10 dB is required for accurate detection (precision < 30 nm). SNR > 10 dB requires that the signal-to-noise ratio of genemap means is over 10. In our optical setup (exposure time > 50 ms), the FROS tagged strains used delivered SNR from 20 to 90 dB. We perform accurate locus tracking for SNR 27 dB (a ratio of about 500).

14. For each frame, a detection test is applied first to screen for the presence of Gaussian peaks indicative of FROS tagged genes. This step provides only a binary detection. Next, the identified peaks need to be properly characterized by measuring relevant parameters such as intensity, subpixel position, width, and offset. They are evaluated by a multiparametric Gaussian fit on the signal intensity. Starting from the particles already detected in the current frame, deflation process subtracts their respective estimated Gaussian peaks from the raw image, and then MTT reiterated the detection and estimation processes on the resulting image until it detected only noise. Filtering tests can remove putative aberrant points.
15. For the frame at time t, MTT can iteratively reconstruct trajectories up to time t based on previous detections. Each of the detections (t) is associated with a set of parameters (including position, mean, and standard deviation of intensity). Parameters (t) are used to accurately reconnect trajectories at the next time point, $(t+1)$. Connecting individual detection, performed at each time point of the time-lapse acquisition, allows the construction of trajectories of FROS tagged loci (called "reconnection").
16. Extracted trajectories are not corrected for possible motion of the nucleus. Therefore, motion of FROS tagged gene determined in MTT is the sum of locus motion in the nucleus plus the motion of the nucleus in the cell. Ideally, the motion of the nucleus should be subtracted from the measured motion. The nuclear center can be evaluated by measuring the centroid of GFP signal (using ImageJ). When the locus has bright FROS signal, the centroid of the GFP signal might be affected by locus position (FROS tagged gene should contribute to less than 10% of total nuclear fluorescence). For short time-lapse imaging, nuclear center detection is not reliable. Retrieving apparent nuclear motion can only add noise to the experimental data. For long time-lapse imaging (>3 min), motion of the nucleus should be retrieved from the locus motion to exclude infrequent, but significant nuclear motion (>1 μm).
17. When false detections are manually excluded, the accuracy of the detection should be evaluated. Classically, we test the algorithm with simulated images and evaluate the optical setup using fluorescent beads and chemically fixed FROS tagged cells. Simulated images are useful because the absolute positioning is known ("ground truth") and the performance of the algorithm is directly testable. Simulated microscopic images of a time-lapse acquisition are generated using Gaussian approximation of point spread function (PSF) of our microscopic setup [24], with

Poisson noise to approximate photon-counting noise. Known trajectory can then be compared to the measured position. Fluorescent beads provide very high SNR imaging and are required for testing PSF of the optical setup and drift during imaging. Chemical fixation abolishes chromosome motion. Imaging of fixed cells provides an easily accessible control for tracking error, exploring possible cell drift, optical path misalignment, and tracking algorithm imprecision. Other errors, such as dynamic errors due to motion blur (apparent streaking of rapidly moving FROS in a image) during time-lapse imaging can also be directly assessed by data analysis [25].

18. To identify ballistic behavior, the plotting of traveled distance over time can be informative. However, gene motion is typically characterized by random motion and directional motion is rarely observed [26]. In statistical mechanics, the mean square displacement (MSD) is the most common measure of the spatial extent of random motion. In some way, it is often enlightening to think of the MSD as the amount of the system "explored" by the random walker.
19. Nucleus cropping is performed using a Z-projected fluorescent image and a shape criterion is included.
20. This step is very important to evaluate possible errors. Additionally, error-prone detection of nuclear envelope position (estimated by fitting an ellipsoid to the detected NPC) can be automatically discarded by rejecting excessive localization errors. The quality control step can automatically reject also loci with low fluorescent intensity, but this might bias population analysis (G1 or S phase cells have different SNR signal [27]). Therefore, we usually adapt the SNR (select individual clones with higher SNR). Alternatively, one might also omit the NPC labeling for some mutants (nuclear and nucleolar center can be extracted without NPC labeling).
21. We aligned the nuclear landmarks of different nuclei by translations that move the nuclear centers to the origin, followed by rotations around the origin that moved nucleolar centroids onto the half-axis (central axis). To facilitate visualization, we rotated all loci around the central axis and merged the results into a single plane.
22. We have observed that nuclear size and nucleolar size can be reproducibly different between independently generated clones bearing similar FROS tagged locus. Genemap should be performed in at least two independent FROS clones in such cases.

Acknowledgement

This work was supported by ATS-Nudgene and Emergence-CLEMgene of the Toulouse-IDEX. O.G. and C.N. are supported by Agence Nationale de la Recherche (ANDY). We thank I. Léger-Silvestre for thoughtful discussions and technical advices. Julien Mathon wrote the MTT-based executable files.

References

1. van Steensel B, Dekker J (2010) Genomics tools for unraveling chromosome architecture. Nat Biotechnol 28(10):1089–1095
2. Ay F, Noble WS (2015) Analysis methods for studying the 3D architecture of the genome. Genome Biol 16:183
3. Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA (2014) Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. Genes Dev 28(24):2778–2791
4. Dion V, Gasser SM (2013) Chromatin movement in the maintenance of genome stability. Cell 152(6):1355–1364
5. Huet S, Lavelle C, Ranchon H, Carrivain P, Victor JM, Bancaud A (2014) Relevance and limitations of crowding, fractal, and polymer models to describe nuclear architecture. Int Rev Cell Mol Biol 307:443–479
6. Hajjoul H, Mathon J, Ranchon H, Goiffon I, Mozziconacci J, Albert B, Carrivain P, Victor JM, Gadal O, Bystricky K, Bancaud A (2013) High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. Genome Res 23(11): 1829–1838
7. Wang R, Mozziconacci J, Bancaud A, Gadal O (2015) Principles of chromatin organization in yeast: relevance of polymer models to describe nuclear organization and dynamics. Curr Opin Cell Biol 34:54–60
8. Albert B, Mathon J, Shukla A, Saad H, Normand C, Leger-Silvestre I, Villa D, Kamgoue A, Mozziconacci J, Wong H, Zimmer C, Bhargava P, Bancaud A, Gadal O (2013) Systematic characterization of the conformation and dynamics of budding yeast chromosome XII. J Cell Biol 202(2):201–210
9. Meister P, Gehlen LR, Varela E, Kalck V, Gasser SM (2010) Visualizing yeast chromosomes and nuclear architecture. Methods Enzymol 470:535–567
10. Berger AB, Cabal GG, Fabre E, Duong T, Buc H, Nehrbass U, Olivo-Marin JC, Gadal O, Zimmer C (2008) High-resolution statistical mapping reveals gene territories in live yeast. Nat Methods 5(12):1031–1037
11. Serge A, Bertaux N, Rigneault H, Marguet D (2008) Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. Nat Methods 5(8):687–694
12. Hajjoul HM, Mathon J, Viero Y, Bancaud A (2011) Optimized micromirrors for three-dimensional single-particle tracking in living cells. Appl Phys Lett 98:243701
13. Therizols P, Duong T, Dujon B, Zimmer C, Fabre E (2010) Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. Proc Natl Acad Sci U S A 107(5):2025–2030
14. Weber SC, Thompson MA, Moerner WE, Spakowitz AJ, Theriot JA (2012) Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function. Biophys J 102(11):2443–2450
15. Backlund MP, Joyner R, Weis K, Moerner WE (2014) Correlations of three-dimensional motion of chromosomal loci in yeast revealed by the double-helix point spread function microscope. Mol Biol Cell 25(22): 3619–3629
16. Michaelis C, Ciosk R, Nasmyth K (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell 91(1):35–45
17. 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
18. Lassadi I, Bystricky K (2011) Tracking of single and multiple genomic loci in living yeast cells. Methods Mol Biol 745:499–522
19. Loiodice I, Dubarry M, Taddei A (2014) Scoring and manipulating gene position and dynamics using FROS in budding yeast. Curr Protoc Cell Biol 62: Unit 22 17 21–14

20. Dubarry M, Loiodice I, Chen CL, Thermes C, Taddei A (2011) Tight protein-DNA interactions favor gene silencing. Genes Dev 25(13): 1365–1370
21. Saad H, Gallardo F, Dalvai M, Tanguy-le-Gac N, Lane D, Bystricky K (2014) DNA dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. PLoS Genet 10(3):e1004187
22. Shaham S (2006) WormBook: methods in cell biology. The C. elegans Research Community (ed), WormBook
23. Winey M, Yarar D, Giddings TH Jr, Mastronarde DN (1997) Nuclear pore complex number and distribution throughout the Saccharomyces cerevisiae cell cycle by three-dimensional reconstruction from electron micrographs of nuclear envelopes. Mol Biol Cell 8(11):2119–2132
24. Zhang B, Zerubia J, Olivo-Marin J (2007) Gaussian approximations of fluorescence microscope point-spread function models. Appl Opt 46(10):1819
25. Backlund MP, Joyner R, Moerner WE (2015) Chromosomal locus tracking with proper accounting of static and dynamic errors. Phys Rev E Stat Nonlin Soft Matter Phys 91(6):062716
26. Chuang CH, Carpenter AE, Fuchsova B, Johnson T, de Lanerolle P, Belmont AS (2006) Long-range directional movement of an interphase chromosome site. Curr Biol 16(8): 825–831
27. Saner N, Karschau J, Natsume T, Gierlinski M, Retkute R, Hawkins M, Nieduszynski CA, Blow JJ, de Moura AP, Tanaka TU (2013) Stochastic association of neighboring replicons creates replication factories in budding yeast. J Cell Biol 202(7):1001–1012

Chapter 5

Quantitative Immunofluorescence Analysis of Nucleolus-Associated Chromatin

Stefan Dillinger and Attila Németh

Abstract

The nuclear distribution of eu- and heterochromatin is nonrandom, heterogeneous, and dynamic, which is mirrored by specific spatiotemporal arrangements of histone posttranslational modifications (PTMs). Here we describe a semiautomated method for the analysis of histone PTM localization patterns within the mammalian nucleus using confocal laser scanning microscope images of fixed, immunofluorescence stained cells as data source. The ImageJ-based process includes the segmentation of the nucleus, furthermore measurements of total fluorescence intensities, the heterogeneity of the staining, and the frequency of the brightest pixels in the region of interest (ROI). In the presented image analysis pipeline, the perinucleolar chromatin is selected as primary ROI, and the nuclear periphery as secondary ROI.

Key words Chromatin, Histone posttranslational modification, ImageJ, Nuclear architecture, Nuclear periphery, Nucleolus, Quantitative immunofluorescence

1 Introduction

Our current understanding on the functional organization of the nuclear genome is based on a variety of microscopy, genomics, and structure biology analyses. They revealed that the organization of chromatin shows specific features at different length scales ranging from the nucleosome (approx. 11 nm in diameter) to the entire nucleus (typically about 10 μm in diameter) within eukaryotic cells. Experimental evidences clearly suggest that the spatiotemporal organization of the genome follows probabilistic rules, in which chromosomal domains (0.1–10 Mb sized chromosome regions) can be defined as functional and structural units that show dynamic alterations during cell cycle and cell differentiation [1–10]. While various genomics approaches make possible the linear partitioning of the genome into chromosomal domains [11], microscopy-based analyses facilitate the three-dimensional segmentation of the nucleus in fixed or living cells. According to such investigations the chromatin can be divided basically into transcriptionally active

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_5,

euchromatin and transcriptionally inactive heterochromatin [12, 13], and the further resolution of these chromatin structures is getting better with the use of various subdiffraction resolution microscopy techniques [14].

In order to study the spatial arrangement of chromatin and its dynamics around morphological subcompartments of the nucleus in single cells, landmark proteins of the nuclear lamina, the nucleolus or centromeres can be labeled by immunofluorescence (e.g., by staining lamin B, nucleophosmin/B23, or CENP-A). Simultaneous staining of eu- and heterochromatin landmarks, such as trimethylation of lysine 4 or 9 of histone H3, and subsequent quantitative analysis of microscopy images allow the monitoring of spatial distribution of active and inactive chromatin between the individual subcompartments. Such investigations can excellently complement the observations of respective genomics or proteomics studies [15–24]. Here we describe an image analysis pipeline to monitor heterochromatin at the perinucleolar space and simultaneously at the nuclear rim in DAPI/B23/H3K9me3-stained nuclei. The method can also be applied to the investigation of the nuclear distribution of other histone posttranslational modifications or architectural chromatin proteins such as CTCF or cohesin. In fact, the chromatin features of any fluorescently labeled subnuclear compartment can be analyzed using the presented methodological approach.

The protocol is divided into three parts: the first part describes a way to measure relative proportion of total nuclear fluorescence intensities within the nucleolus-associated chromatin and at the nuclear rim, the second part deals with the evaluation of the heterogeneity of the staining, and finally the third party describes the analysis of the distribution of the 10% brightest nuclear pixels in the two ROIs. In summary, this protocol provides a detailed step-by-step description as how to quantitatively analyze different areas of the nucleus using confocal laser scanning microscope (CLSM) images of immunofluorescence stained nuclei in a semiautomated manner. The commands of a script, which was written for ImageJ, are integrated in this description and the script is included in Subheading 4.

2 Materials

1. 8-bit CLSM images of immunofluorescence stained nuclei (e.g., DAPI/B23/H3K9me3). Middle optical sections (for instance, 200 nm in thickness) of z-stacks should be selected from CLSM images based on the quality of nuclear morphology (*see* **Notes 1** and **2**). The xy pixel size should be around 80×80 nm or smaller. We use 80.25×80.25 nm.
2. ImageJ software (version 1.49s was used) including the Action Bar and Bio-Formats plug-ins (download them from imagej.nih.gov if needed).

3. Excel from Microsoft Office or equivalent software.
4. Computer workstation with Windows, Mac, or Linux OS (Mac OS 10.10.5 was used).

3 Methods

3.1 Determination of Relative Immunofluorescence Intensities in the ROIs

1. Open the selected optical section of the CLSM image file as separate channels.
2. Duplicate the channel with the DAPI signal and name it 'DAPI_mask' (*see* **Note 3**) [*Shift* + *D*].
3. Use the customized Action Bar (quant_IF) from this step on. The script should be copied into the Plugins/ActionBar folder of ImageJ (*see* **Note 4**).
4. **Get binary mask**: [*Process* > *Binary* > *Make Binary*]. Fill up holes that appear due to unequal staining [*Process* > *Binary* > *Fill holes*]. Remove bumps and dints [*Process* > *Filters* > *Gaussian Blur 3D…* (*x* = 2, *y* = 2, *z* = 2)]. Threshold the image with a value of 128 [*Shift* + *cmd* + *T*, *Apply*]. Delete unwanted background and partial nuclei manually (*see* Ronneberger et al., 2008, for suggestions if you wish to customize the settings).
5. Duplicate the channel containing the staining to be quantified and name it 'histone_channel' (*see* **Note 5**) [*Shift* + *D*].
6. To get the binary mask for nucleoli select the channel with the nucleolus staining (in our case B23), choose an appropriate mid-section and duplicate it [*Shift* + *D*]. Name the resulting image 'No_mask' (*see* **Note 6**). Perform **Gaussian filtering** [*Process* > *Filters* > *Gaussian Blur 3D…* (x = 2, y = 2, z = 2)]. Set the threshold manually to separate the nucleoli from the background and to get the outlines of the nucleoli [*Shift* + *cmd* + *T*, *Apply*]. Remove **Bumps and dints** [*Process* > *Filters* > *Gaussian Blur 3D…* (x = 2, y = 2, z = 2)]. Threshold the image with a value of 128 [*Shift* + *cmd* + *T*, *Apply*]. Delete manually unwanted background and nucleoli from all but the selected cell.
7. **Make all measurements**: In the following steps, several additional masks for the perinucleolar space and the nuclear border are created, which are applied to the previously generated mid-section images and finally fluorescence intensities are measured.

   Measure first the total nuclear fluorescence intensity to get a reference for relative quantification. Apply 'DAPI_mask' to the histone_channel [*Process* > *Image Calculator* > *AND*]. Measure the total fluorescence intensity [*cmd* + *M*] (*see* **Note 5**).

   To create a 240 nm mask for the nuclear periphery select and duplicate 'DAPI_mask' and name it 'DAPI_mask_240.' Erode 3 pixels (=240.75 nm) [*Process* > *Binary* > *Options…*; *Iterations* = *3*; *EDM output*: *Overwrite*; *Do*: *erode*] and subtract the smaller mask from the larger one [*Process* > *Image Calculator* > *Subtract*] (*see* **Note 6**). Apply this mask (Result of

DAPI_mask) to the histone_channel [*Process*> *Image Calculator*> *AND*] (Fig. 1). Measure the total fluorescence intensity of this area [*cmd*+*M*].

To mask a 240 nm rim around nucleoli duplicate the 'No_mask,' rename it as 'No_mask_240' and dilate by 3 pixels [*Process*> *Binary*> *Options…*; *Iterations*= *3*; *EDM output*: *Overwrite*; *Do*: *dilate*]. Subtract the smaller mask from the larger one [*Process*> *Image Calculator*> *Subtract*]. Apply this mask (Result of No_mask_240) to the histone_channel [*Process*> *Image Calculator*> *AND*] (Fig. 1). Measure the total fluorescence intensity of this area [*cmd*+*M*]. Transfer all values to Excel to calculate relative fluorescence intensities (*see* **Note** 7).

3.2 Determination of the Heterogeneity of Staining in the ROIs

The heterogeneity (pixel-to-pixel variation) of given staining intensities in the ROIs is calculated by dividing standard deviation of fluorescence intensity values with the mean fluorescent intensity value. The resulting value is called coefficient of variation (CV) (see also [25]).

1. Use the DAPI_mask, DAPI_mask_240, No_mask, No_mask_240, histone_channel images that were created in Subheading 3.1.
2. Calculate the heterogeneity of the staining in the entire nucleus (**Nu CV**). Inspect the histogram of the nucleus to evaluate the total staining of the nucleus in respect to over-, underexposure, width of the histogram, and normal distribution. Select DAPI_mask and add nuclear outlines to the ROI Manager [*Analyze*> *Analyze particles…*]. Select subsequently the

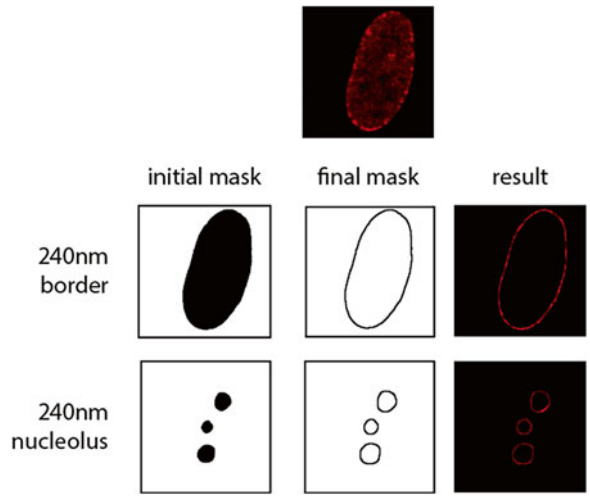

Fig. 1 Definition of the perinucleolar space and the nuclear periphery to measure fluorescence intensity values. Initial masks after thresholding (*left*), final masks with a width of 240 nm/3 px (*center*) and the result of masks applied to the histone channel (*top*) are shown (*right*)

created ROI and the histone_channel in the ROI Manager, so that the outline of the nucleus appears in the histone_channel window. Select the histone_channel. Measure [cmd + M] and transfer the values for standard deviation and mean to an Excel sheet (*see* **Note 8**). Generate a histogram [*Analyze* > *Histogram…*].

3. Calculate the heterogeneity of the staining at the nuclear periphery (**rim CV**). To get the CV from the border of the nucleus select DAPI_mask_240 and add the smaller nuclear mask to the ROI Manager [*Analyze* > *Analyze particles…*], which already contains the DAPI_mask. Then select the histone_channel, select the two ROIs (DAPI_mask and DAPI_mask_240) and run the command [*More* > *XOR*] from the ROI Manager menu. This will specify the nuclear border, i.e., the area that is included in the bigger ROI but not in the smaller one. Measure [cmd + M] and transfer standard deviation and mean values to an Excel sheet (*see* **Note 8**). Delete nuclear border ROI from the ROI Manager.
4. Calculate the heterogeneity of the staining in the perinucleolar space (**no CV**). To get the CV from the perinucleolar space, add the two nucleolar masks (No_mask, No_mask_240) to the ROI Manager [*Analyze* > *Analyze particles…*].

   Since the number of nucleoli varies, select the ROIs manually and add them to the histone_channel image using the [*More* > *XOR*] command from the ROI Manager menu. Measure [cmd + M] and transfer standard deviation and mean values to an Excel sheet (*see* **Note 8**). Delete perinucleolar space ROIs from the ROI Manager.

3.3 Determination of the Subnuclear Distribution of the 10% Brightest Pixels

To get an overview about the distribution of heterochromatin in the nucleus, the distribution of the brightest pixels of the H3K9me3 immunofluorescence staining is evaluated. The analysis of the brightest pixels delivers information about the spatial arrangement of the densest heterochromatic regions of chromosomes in the nucleus.

1. **crop nucleus**: To get an image with only the 10% brightest pixels of the nucleus apply first the DAPI_mask to the histone_channel [*Process* > *Image Calculator* > *AND*]. ().
2. Threshold the image manually to 10% [*Shift* + *cmd* + *T*, *Apply*] (*see* **Note 9**).
3. **add rim**: Add DAPI_mask and DAPI_mask_240 to the ROI Manager [*Analyze* > *Analyze particles…*]. Select the top 10% pixel image (histone_channel). Select the ROIs manually in the ROI Manager and add them to the histone_channel using the command [*More* > *XOR*] (Fig. 2). Measure the number of pixels [cmd + M] and transfer the values to an Excel sheet. Delete the nuclear border ROI from the ROI Manager.

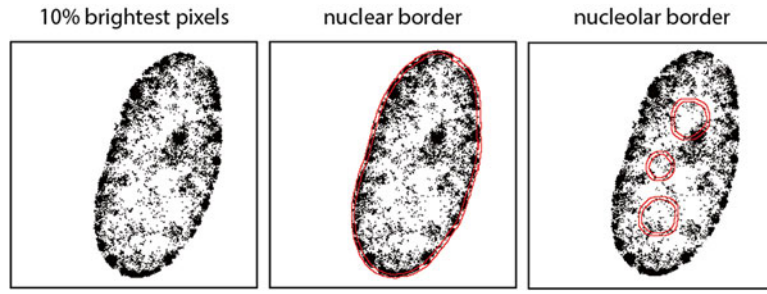

Fig. 2 Distribution of the 10 % brightest pixels. The brightest pixels (top 10 %) of H3K9me3 staining are shown on the *left*, and the ROIs are outlined in the *middle* (nuclear periphery) and on the *right* side (perinucleolar space)

4. **add no**: For counting the pixels in the perinucleolar space add No_mask and No_mask_240 to the ROI Manager [*Analyze* > *Analyze particles…*]. Due to the variable numbers of nucleoli, select the ROIs manually and add them to the histone_channel image using the [*More* > *XOR*] command from the ROI Manager menu (Fig. 2). Measure the number of pixels [cmd + M] and transfer the values to an Excel sheet. Delete the nucleolar border ROI from the ROI Manager.
5. **close all windows**: Close all windows and the canvas is clean for the analysis of the next nucleus.

4 Notes

1. In the described example, IMR90 human embryonic fibroblast cells were stained for the heterochromatin mark H3K9me3 and the nucleolar marker nucleophosmin/B23 using a standard immunofluorescence staining protocol and DNA was counterstained with DAPI in the examples shown. The H3K9me3 antibody was kindly provided by H. Kimura [26], the nucleophosmin/B23 antibody was obtained from Santa Cruz Biotechnology.
2. Image acquisition was performed with a Leica TCS SP8 confocal microscope principally following the rules proposed by Ronneberger et al. [27].
3. The exact spelling of the names is important. Buttons that can be used from the script are marked as **bold text**, commands that can be found in the regular menu of ImageJ are marked as *italic text*. Keyboard shortcuts are referred to a Mac-keyboard.
4. This Note contains the script, which should be saved as quant_IF.txt file into the ImageJ/plugins/ActionBar folder.

```
// Action Bar description file :quant_IF
run("Action
Bar”,”/plugins/ActionBar/quant_IF.txt”);
exit();

<line>
<button>
label=Get binary mask
arg=<macro>
selectWindow("DAPI_mask");
setOption("BlackBackground", false);
run("Make Binary");
run("Fill Holes");
run("Gaussian Blur 3D…", "x=2 y=2 z=2");
//run("Threshold…");
//setThreshold(128, 255);
setOption("BlackBackground", false);
run("Convert to Mask");
</macro>

<button>
label=gaussian filtering
arg=<macro>
run("Gaussian Blur 3D…", "x=2 y=2 z=2");
run("Gaussian Blur 3D…", "x=2 y=2 z=2");
</macro>

<button>
label=bumps&dints
arg=<macro>
run("Gaussian Blur 3D…", "x=2 y=2 z=2");
//run("Threshold…");
run("Convert to Mask");
run("Fill Holes");
</macro>
</line>

<line>
<button>
label=make all measurements
arg=<macro>
selectWindow("histone_channel");
run("8-bit");
imageCalculator("AND create", "histone_channel", "DAPI_mask");
selectWindow("Result of histone_channel");
run("Measure");
selectWindow("Result of histone_channel");
close();
selectWindow("DAPI_mask");
run("Duplicate…", "title=DAPI_mask_240");
run("Options…", "iterations=3 count=1 pad edm=Overwrite do=Erode");
imageCalculator("Subtract create", "DAPI_mask","DAPI_mask_240");
imageCalculator("AND create", "histone_channel","Result of DAPI_mask");
run("Measure");
selectWindow("Result of histone_channel");
close();
selectWindow("Result of DAPI_mask");
```

```
close();
selectWindow("No_mask");
run("Duplicate…", "title=No_mask_240");
run("Options…", "iterations=3 count=1 pad edm=Overwrite do=Dilate");
imageCalculator("Subtract create", "No_mask_240","No_mask");
selectWindow("Result of No_mask_240");
imageCalculator("AND create", "histone_channel","Result of No_mask_240");
selectWindow("Result of histone_channel");
run("Measure");
selectWindow("Result of No_mask_240");
close();
selectWindow("Result of histone_channel");
close();
</macro>
</line>

<line>
<button>
label=Nu CV
arg=<macro>
selectWindow("DAPI_mask");
setOption("BlackBackground", false);
run("Make Binary");
run("Fill Holes");
run("Gaussian Blur 3D…", "x=2 y=2 z=2");
//run("Threshold…");
//setThreshold(128, 255);
setOption("BlackBackground", false);
run("Convert to Mask");
run("Analyze Particles…", "display exclude summarize add");
selectWindow("histone_channel");
roiManager("Select", 0);
run("Measure");
run("Histogram");
</macro>

<button>
label=rim CV
arg=<macro>
selectWindow("DAPI_mask_240");
run("Analyze Particles…", "display exclude summarize add");
selectWindow("histone_channel");
roiManager("Select", 0);
roiManager("Select", newArray(0,1));
roiManager("XOR");
run("Measure");
</macro>

<button>
label=no CV
arg=<macro>
selectWindow("No_mask");
run("Analyze Particles…", "display exclude summarize add");
selectWindow("No_mask_240");
run("Analyze Particles…", "display exclude summarize add");
</macro>
</line>
```

```
<line>
<button>
label=crop nucleus
arg=<macro>
imageCalculator("AND create", "histone_channel", "DAPI_mask");
selectWindow("Result of histone_channel");
selectWindow("histone_channel");
close();
selectWindow("DAPI_mask");
run("Analyze Particles…", "display exclude summarize add");
roiManager("Select", 0);
selectWindow("Result of histone_channel");
roiManager("Select", 0);
</macro>

<button>
label=add rim
arg=<macro>
selectWindow("DAPI_mask_240");
run("Analyze Particles…", "display exclude summarize add");
selectWindow("Result of histone_channel");
roiManager("Select", 0);
roiManager("Select", newArray(0,1));
roiManager("XOR");
selectWindow("Result of histone_channel");
run("Measure");
</macro>

<button>
label=add no
arg=<macro>
selectWindow("No_mask");
run("Analyze Particles…", "display exclude summarize add");
selectWindow("No_mask_240");
run("Analyze Particles…", "display exclude summarize add");
</macro>

<line>
<button>
label=close all windows
arg=<macro>
selectWindow("No_mask");
close();
selectWindow("No_mask_240");
close();
selectWindow("DAPI_mask");
close();
selectWindow("DAPI_mask_240");
close();
selectWindow("Result of histone_channel");
close();
selectWindow("DAPI_channel");
close();
</macro>
</line>
```

5. Parameters that are measured via [cmd + M] can be defined in the menu: *Analyze > Set Measurements…* Here '*RawIntDen*' (*Raw Integrated density*) is used to determine the total fluorescence intensity. This command sums up all brightness values of the pixels in the selected area.
6. Larger areas can be created by using different parameters. Areas can be increased by intervals depending on the pixel size, which was 80.25 nm in our case.
7. Values were copied into an Excel sheet for raw data storage and further calculations. Graphs can be created in R using the R-package *beeswarm*.
8. Standard deviation and mean values were transferred into an Excel file and CV was calculated there.
9. The histogram has 255 bins where pixels are assigned. Due to this fact one cannot threshold the image exactly at 10%. Therefore, the value just above 10% was chosen.

Acknowledgements

This work was supported by the DFG SFB960 program.

References

1. de Wit E, De Laat W (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26:11–24. doi:10.1101/gad.179804.111
2. Bickmore WA (2013) The spatial organization of the human genome. Annu Rev Genomics Hum Genet 14:67–84. doi:10.1146/annurev-genom-091212-153515
3. Bickmore WA, van Steensel B (2013) Genome architecture: domain organization of interphase chromosomes. Cell 152:1270–1284. doi:10.1016/j.cell.2013.02.001
4. Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 14:390–403. doi:10.1038/nrg3454
5. Rivera CM, Ren B (2013) Mapping human epigenomes. Cell 155:39–55. doi:10.1016/j.cell.2013.09.011
6. Tanay A, Cavalli G (2013) Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization. Curr Opin Genet Dev 23:197–203. doi:10.1016/j.gde.2012.12.009
7. Sexton T, Cavalli G (2015) The role of chromosome domains in shaping the functional genome. Cell 160:1049–1059. doi:10.1016/j.cell.2015.02.040
8. Furlan-Magaril M, Várnai C, Nagano T, Fraser P (2015) 3D genome architecture from populations to single cells. Curr Opin Genet Dev 31:36–41. doi:10.1016/j.gde.2015.04.004
9. Naumova N, Imakaev M, Fudenberg G et al (2013) Organization of the mitotic chromosome. Science 342:948–953. doi:10.1126/science.1236083
10. Phillips-Cremins JE, Sauria MEG, Sanyal A et al (2013) Architectural protein subclasses shape 3d organization of genomes during lineage commitment. Cell 153:1281–1295. doi:10.1016/j.cell.2013.04.053
11. van Steensel B, Dekker J (2010) Genomics tools for unraveling chromosome architecture. Nat Biotechnol 28:1089–1095. doi:10.1038/nbt.1680
12. Cremer T, Cremer M, Hübner B et al (2015) The 4D nucleome: evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. FEBS Lett 589:2931–2943. doi:10.1016/j.febslet.2015.05.037
13. Even-Faitelson L, Hassan-Zadeh V, Baghestani Z, Bazett-Jones DP (2016) Coming to terms

with chromatin structure. Chromosoma 125:95. doi:10.1007/s00412-015-0534-9

14. Lakadamyali M, Cosma MP (2015) Advanced microscopy methods for visualizing chromatin structure. FEBS Lett 589:3023–3030. doi:10.1016/j.febslet.2015.04.012
15. Guelen L, Pagie L, Brasset E et al (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453:948–951. doi:10.1038/nature06947
16. Németh A, Conesa A, Santoyo-Lopez J et al (2010) Initial genomics of the human nucleolus. PLoS Genet 6:e1000889. doi:10.1371/journal.pgen.1000889
17. van Koningsbruggen S, Gierlinski M, Schofield P et al (2010) High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol Biol Cell 21:3735–3748. doi:10.1091/mbc.E10-06-0508
18. Wijchers PJ, Geeven G, Eyres M et al (2015) Characterization and dynamics of pericentromere-associated domains in mice. Genome Res 25:958–969. doi:10.1101/gr.186643.114
19. Bai B, Yegnasubramanian S, Wheelan SJ, Laiho M (2014) RNA-Seq of the nucleolus reveals abundant SNORD44-derived small RNAs. PLoS One 9:e107519. doi:10.1371/journal.pone.0107519.s004
20. Bai B, Liu H, Laiho M (2014) Small RNA expression and deep sequencing analyses of the nucleolus reveal the presence of nucleolus-associated microRNAs. FEBS Open Bio 4:441–449. doi:10.1016/j.fob.2014.04.010
21. Caudron-Herger M, Pankert T, Seiler J et al (2015) Alu element-containing RNAs maintain nucleolar structure and function. EMBO J 34:2758, doi: 10.15252/embj.201591458
22. Andersen JS, Lam YW, Leung AKL et al (2005) Nucleolar proteome dynamics. Nature 433:77–83. doi:10.1038/nature03207
23. Moore HM, Bai B, Boisvert F-M et al (2011) Quantitative proteomics and dynamic imaging of the nucleolus reveal distinct responses to UV and ionizing radiation. Mol Cell Proteomics 10:M111.009241. doi:10.1074/mcp.M111.009241
24. Dillinger S, Garea AV, Deutzmann R, Németh A (2014) Analysis of histone posttranslational modifications from nucleolus-associated chromatin by mass spectrometry. Methods Mol Biol 1094:277–293. doi:10.1007/978-1-62703-706-8_22
25. Zhang W, Li J, Suzuki K et al (2015) Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. Science 348:1160–1163. doi:10.1126/science.aaa1356
26. Hayashi-Takanaka Y, Yamagata K, Wakayama T et al (2011) Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. Nucleic Acids Res 39:6475–6488. doi:10.1093/nar/gkr343
27. Ronneberger O, Baddeley D, Scheipl F et al (2008) Spatial quantitative analysis of fluorescently labeled nuclear structures: problems, methods, pitfalls. Chromosome Res 16:523–562. doi:10.1007/s10577-008-1236-4

Chapter 6

Visualization of the Nucleolus in Living Cells with Cell-Penetrating Fluorescent Peptides

Robert M. Martin, Henry D. Herce, Anne K. Ludwig, and M. Cristina Cardoso

Abstract

The nucleolus is the hallmark of nuclear compartmentalization and has been shown to exert multiple roles in cellular metabolism besides its main function as the place of ribosomal RNA synthesis and assembly of ribosomes. The nucleolus plays also a major role in nuclear organization as the largest compartment within the nucleus. The prominent structure of the nucleolus can be detected using contrast light microscopy providing an approximate localization of the nucleolus, but this approach does not allow to determine accurately the three-dimensional structure of the nucleolus in cells and tissues. Immunofluorescence staining with antibodies specific to nucleolar proteins albeit very useful is time consuming, normally antibodies recognize their epitopes only within a small range of species and is applicable only in fixed cells. Here, we present a simple method to selectively and accurately label this ubiquitous subnuclear compartment in living cells of a large range of species using a fluorescently labeled cell-penetrating peptide.

Key words Cell permeable nucleolar marker, Cell-penetrating peptide, Fluorescence microscopy, Living cells, Nucleolus, Poly-arginine

1 Introduction

In 1831, a Scottish botanist Robert Brown described for the first time the "nuclei" in plant cells [1]. A few years later, in 1836, Gabriel Gustav Valentin working with neuronal cells described within nuclei a prominent substructure later termed "nucleolus" [2]. It was only one century later, in the 1960s, that findings derived from electron microscopic autoradiography of [^{3}H]uridine incorporation established the major function of the nucleolus in rRNA synthesis, rRNA processing, and ribosome biogenesis [3, 4]. Several other tasks have been more recently added to the functional portfolio of the nucleolus (reviewed in [5, 6]).

The nucleolus is a self-organizing structure not delimited by membranes that gets disassembled and reassembled around the rDNA loci at every cell division cycle ([7]; reviewed in [8]).

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_6,

Transcription of ribosomal genes, which make up the fibrillar centers, takes place surrounding this structure and creates the so-called dense fibrillar component [9, 10]. Further processing of the precursor RNA and multiple proteins gives rise to ribosome precursor particles, which form the granular component (reviewed in [11]). Over 300 nucleolar proteins have been identified by proteomics [12]. The first nucleolar targeting sequences were identified in viral proteins ([13, 14]; reviewed in [15]) and further shown to be able to cross cellular membranes [16, 17]. Recently, we performed a systematic analysis of peptide sequences sufficient for nucleolar targeting [18]. We found that six or more arginines are sufficient to label the nucleolus by binding to RNA. Hence, these peptide sequences can label the nucleolus in all species tested (insect, fish, mouse, human, and very likely also yeast). Importantly, in particular nona- and deca-arginine peptides are similarly highly active in cell membrane permeation [19].

In this chapter, we describe how adding these peptides conjugated to a fluorescent label directly to the cell culture medium without the need for transient or stable transfections of the cells provides a live cell fluorescent nucleolar marker in a variety of species [20]. The nucleolar marker is membrane permeable and reaches its target within minutes. To prevent proteolytic degradation the peptide can be synthetized with D-amino acids. The cell-penetrating-peptide-based marker provides a method for the noninvasive marking of the nucleolus in living cells, allowing the three-dimensional reconstruction of the nucleolus, which is not possible with contrast microscopy techniques. Besides that, it can easily be used in combination with different fluorophores like fluorescent proteins or DNA dyes to correlate different aspects of nuclear structures with the nucleolus. In addition, nucleolar labeling is also preserved during fixation and staining of the cells. Furthermore, it does not affect cellular viability, proliferation, and rDNA transcription.

2 Materials

Prepare all solutions at room temperature unless otherwise noted. Solutions used to culture and prepare cells should be warmed to 37 °C prior to use. All solutions and materials in contact with live cells must be sterile and have to be handled inside a tissue culture biosafety cabinet. Disposal of reagents, solutions, and cell material has to be carried out following the specific regulations in effect.

2.1 Cells and Materials for Cell Culture

1. Human HeLa cervix epithelium adenocarcinoma cells.
2. Adult rat ventricular cardiomyocytes.
3. Mouse C2C12 myoblast cells.
4. Pac2 zebrafish fibroblast cells (*Danio rerio*).
5. Sf9 insect ovary cells (*Spodoptera frugiperda*).

6. High-glucose (4.5 g/L) Dulbecco's Modified Essential Medium (DMEM).
7. Leibovitz L-15 medium.
8. EX-CELL® 420 Insect Serum Free medium.
9. Glutamine.
10. Gentamicin.
11. Penicillin/Streptomycin 100×.
12. Fetal calf serum (FCS). Aliquots of 50 mL are stored at –20 °C.
13. Standard medium for HeLa cell culture: DMEM (**item 6**) was supplemented with 5 mM L-glutamine (**item 9**), 5 µg/mL gentamycin (**item 10**), and 10% FCS (**item 12**). Prepare a 500 mL bottle under sterile conditions and store at 4 °C.
14. Standard medium for adult rat ventricular cardiomyocytes was 10 mM HEPES buffer with 0.3 mM Ca^{2+} and 0.5% bovine serum albumin stored at 4 °C.
15. Standard medium for C2C12 cell culture: DMEM (**item 6**) was supplemented with 5 mM L-glutamine (**item 9**), 5 µg/mL gentamycin (**item 10**), and 20% FCS (**item 12**). Prepare a 500 mL bottle under sterile conditions and store at 4 °C.
16. Standard medium for Pac2 cells: Leibovitz L-15 medium (**item 7**) was supplemented with 15% FCS (**item 12**) and 1% penicillin–streptomycin (**item 11**) from the 100× stock solution and stored at 4 °C.
17. Standard medium for Sf9 cells: EX-CELL® 420 Insect Serum Free medium (**item 8**) was supplemented with 5 mM L-glutamine (**item 9**) and 10% FCS (**item 12**) and stored at 4 °C.
18. Phosphate buffered saline (PBS) supplemented or not with 0.5 mM EDTA (PBS/EDTA).
19. HEPES buffer 1 M, pH 7.2–7.5.
20. 0.25% trypsin in PBS/EDTA prepared from trypsin powder.

2.2 Peptides

Deca-arginine peptides containing amino-terminal fluorescein label (FITC-R_{10}) were synthesized using D-amino acids by Peptide Specialty Laboratories (Heidelberg, Germany). After synthesis the peptides were HPLC purified and their molecular weight was confirmed by mass spectrometry.

Ultrapure water is used to resuspend the lyophilized peptides.

2.3 Chamber Systems and Coverslips for Live Cell Labeling of Nucleoli

1. µ-dish 35 mm low with Ibidi standard bottom (Ibidi®).
2. Chambered coverslip µ-slide 8 well (Ibidi®).
3. Chambered coverglass 4 and 8 well (Thermo Fisher Scientific, Nunc™ Lab-Tek™).
4. Coverslips, 12 mm diameter.
5. Parafilm®.

2.4 Microscopy

2.4.1 Confocal Laser Scanning Microscope

For live cell microscopy any suitable confocal microscope system can be used. We use a Zeiss LSM510Meta confocal setup mounted on an Axiovert 200 M inverted microscope equipped with a 63× phase contrast plan-apochromat oil objective NA1.4. During all acquisitions, the main beam splitter was a HFT UV/488/543/633 and the parameters for the detection of FITC were as follows:

Excitation wavelength: 488 nm.

Emission filter: BP500-530 nm.

2.4.2 Environmental Chamber and Stage Incubation System

For live cell microscopy any suitable microscope incubation system can be used. The microscope is housed in an environmental cage incubator, which is connected to a heating system providing a constant temperature of 37 °C. In addition, a chambered stage incubation system is used that provides a temperature of 37 °C and humidified 5 % CO_2 atm (all incubation components from Okolab, Ottaviano, Italy). Cells grown in chambered coverslip systems are kept inside the stage incubation system during imaging to maintain stable environmental conditions.

2.5 Image Analysis

For the analysis of microscopic images any suitable software package can be applied. We used ImageJ (http://imagej.nih.gov/ij/) to measure, e.g., relative fluorescence intensities in nucleoli compared to the nucleoplasm.

3 Methods

3.1 Cell Culture

3.1.1 Maintaining Mammalian Cells in Culture

1. HeLa and C2C12 cells (**items 1** and **3** in Subheading 2.1) are grown in p100 dishes in the cell culture incubator at 37 °C and 5 % CO_2 atm, until they reach 70–80 % confluence. Adult rat ventricular cardiomyocytes primary cells (item 2 in Subheading 2.1) cannot be maintained in culture and are used the day they are isolated.
2. Cells are grown in standard medium as indicated in Subheading 2.1.
3. To split cells, aspirate medium and rinse with 1× PBS/EDTA prewarmed to 37 °C.
4. Directly apply 1.0 mL of Trypsin–EDTA (**item 20** in Subheading 2.1) solution prewarmed to 37 °C to the cell layer and tilt the dish gently. Leave flask in the incubator for approximately 1–3 min.
5. Hit the side of the dish with the palm of one hand several times to detach the cells, avoid spilling of fluid to the side and lid of the dish.

6. When all cells have detached, add 9.0 mL of growth medium to the flask and resuspend the cells by gentle pipetting up and down several times. Avoid creating bubbles and foam. Transfer between 1.0 (1:10) and 2.0 (1:5) mL to a new cell culture dish.
7. Fill up to a final volume of 10 mL with standard medium.
8. Place the dish in a cell culture incubator and examine the cells daily using a cell culture microscope.

3.1.2 Maintaining Fish Cells in Culture

1. Pac2 cells (**item 4** in Subheading 2.1) are grown in p25 flasks in a cell culture incubator at 28 °C without CO_2 enriched atmosphere.
2. Pac2 cells are grown in standard medium (**item 16** in Subheading 2.1).
3. To split cells, aspirate medium and rinse twice with 1× PBS/EDTA (**item 18** in Subheading 2.1).
4. Directly apply sufficient trypsin–EDTA solution (**item 20** in Subheading 2.1) to cover the cell layer, tilt the dish gently, and incubate for 5 min at room temperature.
5. Dilute the detached cells with standard medium and pipette up and down several times to completely detach all cells and disperse cell clots.
6. Transfer 20% of the volume to a new flask and fill up with standard medium and place in the incubator at 28 °C.

3.1.3 Maintaining Insect Cells in Culture

1. Sf9 cells (**item 5** in Subheading 2.1) are maintained in p25 flasks as suspension culture in an orbital shaker with ambient atmosphere at 28 °C.
2. For labeling experiments with microscopic observation, cells are grown in an incubator with ambient atmosphere at 28 °C.
3. Sf9 cells are grown in standard medium (**item 17** in Subheading 2.1).
4. To subculture cells, resuspend cells gently by using a plastic pipette and transfer 1.0 mL to a new flask, fill up with standard medium and place in the orbital shaker at 28 °C.

3.1.4 Preparation of Peptides for the Labeling of Live Cells

1. Peptides are delivered lyophilized and therefore have to be resuspended.
2. Add sterile deionized ultrapure water to the lyophilized peptides to prepare peptide stock solutions at a final concentration of 1 mM, vortex to completely dissolve and homogenize the peptide solution.
3. Spin down peptide solution and prepare aliquots of 50 µL in sterile Eppendorf tubes and store at −20 °C until further use.

3.2 Preparation of Cells and Peptides for Nucleolar Labeling and Imaging

3.2.1 Culturing Cells in Live Cell Chamber Systems

1. Take live cell chambers out of sterile packaging and place into an empty sterile p100 dish (*see* **Note 1**).
2. Wash, trypsinize, and resuspend HeLa and C2C12 cells as described in Subheading 3.1.1.
3. Wash, trypsinize, and resuspend Pac2 cells as described in Subheading 3.1.2.
4. Use directly the suspension culture of Sf9 cells as described in Subheading 3.1.3, pipette up and down to disperse cells.
5. Seed equal volumes of cell suspension into each well of a sterile live cell chamber system at 60–70% confluence.
6. Typically 100 μL of cell suspension were seeded into one well of an 8-well LabTek or Ibidi chamber and correspondingly 200 μL were seeded into 4 well chambers. These amounts correspond to a split ratio of around 1: 2.5 taking into account the dish and chamber surface areas with cells grown to 70–80% confluence. Amounts of cells seeded have to be adapted to the respective cell type that will be used.
7. Fill up chambers up to half of the volume with growth medium and close the lid. To avoid contamination on the way to the incubator, close additionally the p100 dish containing the live cell chamber.
8. Human HeLa, mouse C2C12, and Pac2 fish cells are grown under the conditions described in Subheadings 3.1.1 and 3.1.2, respectively. Sf9 insect cells, however, are grown in an incubator with ambient atmosphere at 28 °C without shaking to allow attachment of the cells to the chamber surface.
9. Cells are grown over night and used the next day for nucleolar labeling with the transducible peptide marker (*see* **Note 2**).

3.2.2 Preparation of Coverslips for Fixation of Cells After Labeling the Nucleoli

1. Place 12 mm coverslips in a dedicated holder inside a closed container and immerse in pure 100% ethanol for several hours.
2. Use fine tip forceps to place the desired number of coverslips into an empty sterile cell culture dish.
3. Leave the dish with the lid half open inside the laminar flow cabinet until all traces of ethanol evaporated.
4. Immediately before starting to split cells to seed onto coverslips, use 1× PBS and fill carefully into the dish with the coverslips until all coverslips are covered. Avoid coverslips to adhere to each other by applying PBS slowly and carefully at the side of the dish without moving the coverslips.
5. Aspirate PBS and continue with washing, trypsinizing, and resuspending cells as described in Subheadings 3.1.1, 3.1.2, and 3.1.3, respectively.

6. Seed the required volume of resuspended cells to reach 70% confluence and fill up with growth medium. Avoid disturbance to prevent coverslips floating around.
7. Place the dish inside the cell culture incubator at 37 °C and 5% CO_2 until use, typically on the next day.

3.2.3 Labeling of the Nucleoli in Living Cells Using a Cell-Penetrating Peptide Marker (See Fig. 1)

1. Dilute deca-arginine peptides in DMEM without FCS to a final concentration of 10 μM just before the start of the labeling experiments and keep at 4 °C until use (*see* **Notes 3** and **4**).
2. For imaging experiments place cells grown in chambered coverslips or microscopy dishes inside the stage incubation system, fix using metal clips or similar devices to prevent movements and leave for at least 30 min to adapt to the environment and observe cells by phase contrast microscopy.
3. Remove all growth medium from the first well to be labeled and immediately apply with great care the peptide in DMEM solution (*see* **Note 5**).
4. Close the chamber lid and the stage incubation system. Adjust the microscope objective to focus the cells. It is possible to observe cells at this stage by microscopy; however, excess labeled peptide and the high concentration of fluorophores in the medium will cause very high intensities when observing the peptide fluorescence.
5. Incubate the cells for 30 min to 1 h with the peptides in DMEM. Aspirate all peptide solution, wash cells one time in DMEM without FCS or alternatively in 1× PBS to remove excess fluorescent peptides, aspirate the washing solution, and apply the standard growth medium to the cells.
6. Close the chambered coverslip and the stage incubation system, focus the cells, and start the experimental observations and image acquisition.
7. For cells grown on coverslips, aspirate medium carefully from the dish and apply the peptides in DMEM at the side of the dish, not directly on the coverslips (*see* **Note 6**). After removal of the peptide in DMEM solution use forceps to place the coverslips with cells turned upward on Parafilm in a humidified chamber for further treatment.
8. Perform the washing steps with 1× PBS on individual coverslips in the humidified chamber and proceed with a protocol for cell fixation, permeabilization, and immunostaining (*see* **Note 7**).
9. Alternatively, the application of peptides in DMEM, as well as the washing step can be performed in a laminar flow biosafety cabinet. In this case, cells are kept in the cell culture incubator for the peptide uptake incubation. Microscopic observation can start after labeling or any time later (*see* **Note 8**).

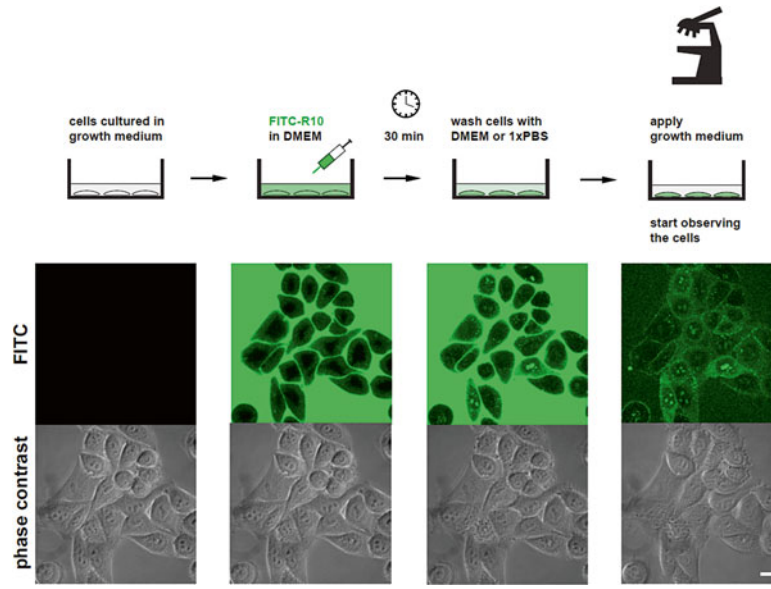

Fig. 1 Application of a fluorescent cell-penetrating-peptide-based nucleolar marker. The *upper row* shows schematically the steps for the labeling of nucleoli in live cells with the cell-penetrating peptide marker. Cells are cultivated in microscopy dishes and after removal of the growth medium incubated with the peptide marker diluted in DMEM. After an incubation time of about 30 min, the peptide solution is removed and cells are washed with fresh DMEM or PBS to remove excess fluorescent peptides. Finally the cells are supplied with standard growth medium followed by immediate microscopic observation of the fluorescently labeled nucleoli. The *lower panel* shows exemplary images of a culture of human HeLa cells at the different stages of the labeling process depicted in the scheme earlier. The *upper image row* shows FITC fluorescence of the cell-penetrating peptide marker with the corresponding phase contrast image below. Note the lack of signals inside the cell after application of the fluorescent peptide in the second step, while after 30 min of incubation in step 3, nucleolar labeling is already visible. After washing away excess peptide, the fluorescently labeled nucleoli are distinctively visible in the vast majority of cells against the darker nucleoplasmic surrounding. Scale bar 10 μm

10. For cells in suspension culture, we suggest mild centrifugation steps to remove growth medium, resuspension of cells in DMEM with diluted peptides followed by mild centrifugation for peptide removal and washing steps.
11. Additional labeling of cells with cell permeable dyes like DRAQ5 [21] can be performed before or after peptide labeling of nucleoli (*see* **Note 9**).

3.3 Imaging (See Fig. 2)

1. Place the dish containing the cells treated with the nucleolar marker on the microscope stage. Close the incubation chamber and wait 30–60 min to allow the sample to acclimatize to the chamber temperature before starting with image acquisition. This will avoid focus drift while imaging.

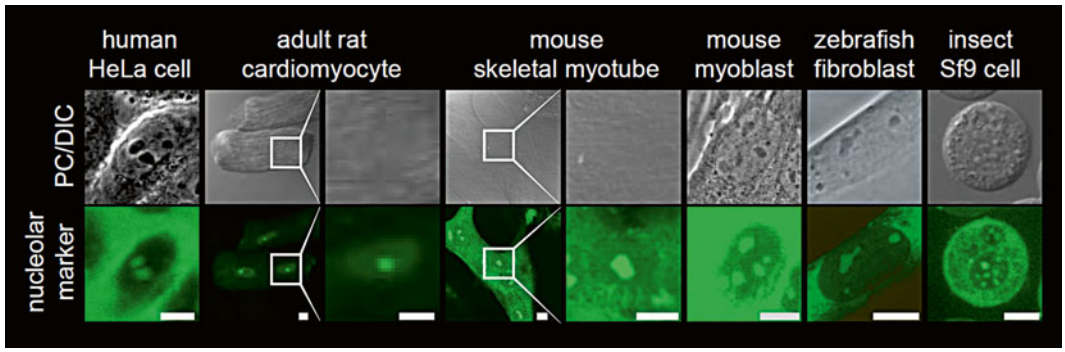

Fig. 2 Intracellular distribution of nucleolar markers in living cells of various species and origin. Shown are mid-optical sections of cultured cell lines and terminally differentiated primary cells as indicated earlier. For visualization of nucleoli in situ, cells were treated for 30 min with medium containing the nucleolar marker, washed in PBS, and incubated in medium for live cell imaging on a confocal microscope equipped with a climatization chamber to maintain a constant temperature of 37 °C, 5 % CO_2, and 60 % humidity. The *lower image row* shows FITC fluorescence of the cell-penetrating peptide marker with the corresponding phase/differential interference contrast (PC/DIC) image below. Scale bar 10 μm

2. Focus the cells using the transmitted light channel (i.e., differential interference contrast (DIC), phase contrast, etc.).
3. Adjust the microscope settings, in particular, the 488 nm laser wavelength intensity and exposure time to optimize the visualization of the nucleoli (minimizing overexposure of the sample). To capture the entire volume of the nucleolus it is required to image optical sections that span approximately −4 μm to 7 μm from the focal plane (for HeLa cells, *see* **Note 10**).
4. Image the cells.
5. Export the images as tiff files for further processing using ImageJ.

3.4 Image Analysis (See Fig. 3)

Segmentation of Nucleoli

1. Open the image stack (or single image) in ImageJ.
2. Select the cell of interest using the "rectangular" selection option in the tools menu bar.
3. Crop ("Image" menu > "Crop"), Fig. 3a.
4. Separate the stack into single images ("Image" menu > "Stacks" > "Stacks to Images"), Fig. 3b.
5. Close without saving the images that do not contain any fraction of the nucleus. The marker distributes strongly in the cytosol and nucleoli and is weakly distributed over the nucleoplasm. This can be used to easily recognize the nucleus without the need of any extra nuclear marker.

For Each Stack Image

6. Select the nucleus using the "Freehand selections" in the tools menu bar, Fig. 3c.

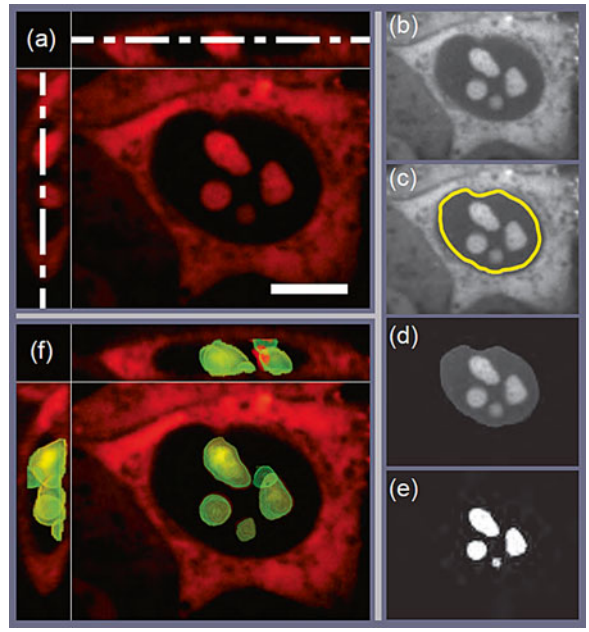

Fig. 3 Image analysis steps for segmentation of nucleoli. (**a**) Image stack of a single HeLa cell visualized using Volocity. *xy* as well as *xz* and *yz* views of the cell are shown. (**b**) Example image from z-stack for segmentation. (**c**) Nucleus selection (*yellow line*). (**d**) Clearance of signal outside the nucleus. (**e**) Binary image of the nucleoli. (**f**) Overlay of image stack in (**a**) and a mesh representation of the surface of the 3D rendered nucleoli. Scale bar 10 μm

7. Clear the marker signal outside the nucleus ("Edit" menu > "Clear Outside"), Fig. 3d.
8. Convert the image to a binary form to isolate the nucleoli region ("Process" menu > "Binary" > "Make Binary" and use the "Otsu" filter (*see* **Note 11**), Fig. 3e).
9. For further processing each image can be used individually as a mask for the nucleolus or the images can be collected into a single stack ("Image" menu > "Stacks" > "Images to Stacks").
10. Save resulting individual images or image stack as tiff (click selected image and "File" menu > "Save as" > "tiff").
11. The 3D stack can be visualized using the ImageJ plug-in "3D Viewer" ("Plugins" menu > 3D > "Volume Viewer"), select view angle and options and save as an image "Save View," Fig. 3f.

4 Notes

1. We place the chambered coverslip systems and microscopy dishes inside p100 cell culture dishes for better handling, easy transport, and sterility between laminar flow biosafety cabinet, incubator, and microscope incubator.

2. For primary cells, e.g., freshly isolated cardiomyocytes it is recommended to perform the peptide labeling and washing steps as soon as possible after preparation of the sample due to their very limited lifespan [20]. In the case of other primary cell cultures with longer survival time, it might be possible to delay the labeling after sample preparation.
3. The efficiency of uptake of cell-penetrating peptide markers is a function of peptide length and concentration but is also cell type specific. A concentration of 10 μM FITC labeled deca-L-arginine was sufficient to label a range of different cell types from different species (*see* Fig. 2 and [18]). However, for certain cell types the concentration of FITC-R_{10} might be elevated up to 50 μM to achieve efficient nucleolar labeling with a low to moderate rate of toxicity [22]. Higher concentrations of R_{10} peptide might result in high frequency of damaged or dead cells.
4. For the peptide transduction we prepare a volume of peptides diluted in DMEM sufficient to completely cover all cells in the respective number of chambers to be labeled. Typically in a single well of a 4-well chamber it is sufficient to apply 200 μL of diluted peptide in DMEM to cover all cells completely and provide enough volume to prevent dry spots in the center of the well due to the fluid meniscus. Accordingly, a volume of 100 μL is sufficient in one well of an 8-well chamber.
5. When applying the peptide diluted in DMEM to the cells, place the tip of the pipette into one corner of the well, close to the bottom but not touching it and slowly release the fluid to immerse the cells. Avoid drops of liquid falling on the cells to prevent detaching of cells. When the chamber bottom surface is covered with liquid, add the rest of the volume slightly faster but still with great care. Perform washing steps and the final application of growth medium in the same way.
6. For increased efficiency of peptide uptake into cells to be fixed, we incubated cells with the peptide up to 1 h.
7. Membrane permeable peptides are not recommended for labeling of nucleoli postfixation [20].
8. The persistence of the nucleolar labeling in living cells was successfully observed up to 24 h after the labeling event.
9. For additional labeling of cellular components by transient transfection of cells with, e.g., plasmids encoding proteins tagged with red fluorescent proteins, we suggest to perform the transfection the day before the nucleolar labeling with the peptides.
10. This range depends significantly on the specific cell type and stage of the cell cycle.
11. Most filters (e.g., “MaxEntropy,” “Mean,” etc.) can be used here with very similar results.

Acknowledgements

We thank all present and past members of the laboratory for their contributions over the years. Robert M. Martin is supported by a fellowship of the Fundação para a Ciência e Tecnologia, Portugal (SFRH/BPD/66611/2009). The laboratory of M. Cristina Cardoso is supported by grants of the German Research Foundation (DFG) and the Federal Ministry of Education and Research (BMBF).

References

1. Brown R (1831) Observations on the organs and mode of fecundation in Orchideae and Asclepiadae. Trans Linnean Soc Lond 16(3): 709–737
2. Valentin G (1836) Repertorium für Anatomie und Physiologie, vol 1. Verlag von Veit unc Comp, Berlin
3. Brown DD, Gurdon JB (1964) Absence of ribosomal Rna synthesis in the anucleolate mutant of Xenopus Laevis. Proc Natl Acad Sci U S A 51:139–146
4. Granboulan N, Granboulan P (1965) Cytochimic ultrastructurale du nucleole. II. Etude des sites de synthese du RNA dans le nucleole et le noyau. Exp Cell Res 38:604–619
5. Pederson T (1998) The plurifunctional nucleolus. Nucleic Acids Res 26(17):3871–3876
6. Pederson T (2011) The nucleolus. Cold Spring Harb Perspect Biol 3(3):a000638. doi:10.1101/cshperspect.
7. Dundr M, Misteli T, Olson MO (2000) The dynamics of postmitotic reassembly of the nucleolus. J Cell Biol 150(3):433–446
8. Hernandez-Verdun D (2011) Assembly and disassembly of the nucleolus during the cell cycle. Nucleus 2(3):189–194. doi:10.4161/nucl.2.3.16246
9. Cheutin T, O'Donohue MF, Beorchia A, Vandelaer M, Kaplan H, Defever B, Ploton D, Thiry M (2002) Three-dimensional organization of active rRNA genes within the nucleolus. J Cell Sci 115(Pt 16):3297–3307
10. Derenzini M, Pasquinelli G, O'Donohue MF, Ploton D, Thiry M (2006) Structural and functional organization of ribosomal genes within the mammalian cell nucleolus. J Histochem Cytochem 54(2):131–145
11. Scheer U, Hock R (1999) Structure and function of the nucleolus. Curr Opin Cell Biol 11(3):385–390
12. Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI, Mann M (2005) Nucleolar proteome dynamics. Nature 433(7021):77–83
13. Dang CV, Lee WM (1989) Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70, and HIV tat proteins. J Biol Chem 264(30):18019–18023
14. Hatanaka M (1990) Discovery of the nucleolar targeting signal. Bioessays 12(3):143–148
15. Leonhardt H, Cardoso MC (1995) Targeting and association of proteins with functional domains in the nucleus: the insoluble solution. In Structural and Functional Organization of the Nuclear Matrix. Int Rev Cytol 162B:303–335
16. Frankel AD, Pabo CO (1988) Cellular uptake of the tat protein from human immunodeficiency virus. Cell 55(6):1189–1193
17. Green M, Loewenstein PM (1988) Autonomous functional domains of chemically synthesized human immunodeficiency virus tat transactivator protein. Cell 55(6):1179–1188
18. Martin RM, Ter-Avetisyan G, Herce HD, Ludwig AK, Lattig-Tunnemann G, Cardoso MC (2015) Principles of protein targeting to the nucleolus. Nucleus 6(4):314–325. doi:10.1080/19491034.2015.1079680
19. Herce HD, Garcia AE, Cardoso MC (2014) Fundamental molecular mechanism for the cellular uptake of guanidinium-rich molecules. J Am Chem Soc 136(50):17459–17467. doi:10.1021/ja507790z
20. Martin RM, Tunnemann G, Leonhardt H, Cardoso MC (2007) Nucleolar marker for living cells. Histochem Cell Biol 127(3):243–251
21. Martin RM, Leonhardt H, Cardoso MC (2005) DNA labeling in living cells. Cytometry A 67(1):45–52
22. Tünnemann G, Ter-Avetisyan G, Martin RM, Stöckl M, Herrmann A, Cardoso MC (2008) Live-cell analysis of cell penetration ability and toxicity of oligo-arginines. J Pept Sci 14(4):469–476. doi:10.1002/psc.968

Part II

Analysis of Ribosomal RNA Transcription and Processing

Chapter 7

Purification of Crystallization-Grade RNA Polymerase I from *S. cerevisiae*

Christoph Engel

Abstract

Purification of RNA polymerase (Pol) I is essential for functional as well as for structural studies. The product needs to be extremely pure in order to exclude secondary effects, e.g., caused by copurified nucleic acids in subsequent experiments. For this purpose, the method presented here was originally introduced nearly a decade ago but underwent constant optimization [1]. The polymerase is extracted from its endogenous source, since no overexpression system for the entire 590 kDa, 14-subunit complex is available thus far. Following yeast cultivation, a number of standard protein purification techniques are applied and combined to a robust but elaborate procedure that takes 3 days. In brief, a yeast strain with histidine-tagged RNA polymerase I is fermented, cells are broken by bead beating, and cell debris is removed by a two-step centrifugation. The lysate is then dialyzed, the Pol-I-containing pellet resuspended, and polymerase I enriched by a His-trap affinity step, followed by sequential purification via anion and cation exchange and a final size exclusion chromatography.

Key words Protein purification, RNA polymerase I, Transcription, *Saccharomyces cerevisiae* fermentation, Affinity and size exclusion chromatography

1 Introduction

The 14-subunit RNA polymerase I transcribes the 35S ribosomal RNA (rRNA) precursor in yeast and is distinct from the other eukaryotic RNA polymerases Pol II and Pol III [2, 3]. By providing rRNAs which form the central building blocks for all ribosomes, Pol I contributes up to 60% of cellular transcription levels during exponential growth [4]. Recently, Pol I came into the spotlight of numerous investigations because its crystal structure was solved in two different forms [5, 6] and in a third one shortly after [7]. Whereas these crystal structures answered many questions, they raised a similar amount of new ones. Therefore, a high number of follow-up experiments of structural as well as biochemical nature are imminent. For most of these in vitro assays, extremely pure material is required. Here we describe a method for the purification of crystallization grade Pol I

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_7,

from a *Saccharomyces cerevisiae* strain that carries a 10× histidine tag on the C-terminus of its largest subunit A190.

Each step of the purification is easily performed. The combination of classical biochemical techniques with affinity and size exclusion chromatography (SEC) steps exploits the main characteristics of the Pol I complex. Its solubility, global as well as local electrostatic charge distribution, and the overall size of the enzyme determine its specific reaction to the steps employed in this procedure. A deca-histidine tag was chosen in order to increase Ni-Nitrilotriacetic acid (NTA) binding affinity which allows for a more stringent washing and therefore improves contaminant removal at this stage [5]. A single FLAG tag at the largest Pol I subunit A190 furthermore allows for an efficient tracking of Pol-I-containing fractions using immunoblotting. Despite its stringency, Pol I does not dissociate during the procedure. While the heterodimeric subunit complex A49/A34.5 can be lost from Pol I [1, 8], the product of the technique presented here is not only entirely complete but also fully active in RNA production and cleavage [5].

Recent unpublished results furthermore demonstrate that the protocol is stable enough to be applied for the purification of Pol I from other organisms, such as the fission yeast *Schizosaccharomyces pombe*, with little changes. During the preparation of this work, other purification protocols for Pol I were published [9, 10]. It might be worth thinking about the intended use for the purified polymerase. Whereas other techniques might be quicker or easier, the protocol presented here yields polymerase of great purity and was used to determine the crystal structure of Pol I at the highest resolution [5].

2 Materials

Since Pol I purification can be rather stressful in practice, it is recommended to prepare all media, salt stocks, and buffers that are used at any point of the protocol well in advance. This holds especially true for the first steps until the His-Trap procedure is finished. During buffer preparation, it is advisable to (a) filter all buffers (e.g., using a 0.2 μm filter and a vacuum pump), (b) readjust the pH after all components have been added, (c) only add dithiothreitol (DTT), beta-mercapto-ethanol, and/or protease inhibitor just before use, and (d) store all buffers at 4 °C. The presence of general laboratory equipment such as pipets, tubes, beakers, flask, shakers, centrifuges, and a basic set of reagents is required and will not be specifically mentioned.

2.1 Cell Growth and Fermentation

1. YPD medium: 2% (w/v) Bacto Peptone, 2% (w/v) D(+)glucose monohydrate, 1% Bacto Yeast Extract; dissolve in water and autoclave.
2. YPD for a 200 L fermenter: 4 kg D(+)glucose, 3 kg Bacto Yeast Extract, 4 kg Bacto Peptone.

3. Antifoam C Emulsion 1:10 (Sigma). Add 200 mL to a 200 L fermenter after autoclaving.
4. Ampicillin Na-salt. Dissolve in water at 100 mg/mL.
5. Tetracyclin hydrochloride. Dissolve in 80% ethanol at 12.5 mg/mL.
6. Fermenter: Whereas any fermenter with a volume of 100–300 L can be applied, we use an INFORS-HT Techfors fermenter with cell-separation system.
7. For freezing and storage of yeast pellets, single use 250 mL conical tubes have proven reliable (Corning).
8. Protease inhibitor stock for several purifications (100-fold PI): Dissolve 28.4 mg Leupeptin, 137 mg Pepstatin, 17 g phenylmethylsulfonylfluoride (PMSF), 33 g benzamidine hydrochloride in 1 L of pure ethanol.
9. Lysis/Freezing buffer: 150 mM HEPES pH 7.8, 60 mM magnesium chloride, 20% v/v glycerol, 5 mM DTT, 1 × PI.

2.2 Breaking the Cells and Removal of Debris

1. Ammonium sulfate: weigh 198.21 g and dissolve in 500 mL Millipore water to obtain a 3 M stock.
2. Beta-mercapto-ethanol at 14.3 M.
3. Glass beads 0.5 mm (Biospec).
4. Bead beater (Biospec).
5. Crude salt for ice mix, 25 kg (e.g., from AGRAVIS Streusalz).
6. PMSF and benzamidine: Weigh 4.35 g of PMSF and 3.0 g of benzamidine hydrochloride, combine and dissolve in 250 mL of 80% ethanol. Use a 100-fold dilution to obtain a 1 mM concentration for each component.
7. Dialysis tubes 12–14 kDa MWCO (Spectra).
8. Dilution buffer: 100 mM HEPES pH 7.8, 20 mM magnesium chloride, 400 mM ammonium sulfate.
9. Twofold Dialysis buffer: 100 mM potassium acetate, 40 mM HEPES pH 7.8, 2 mM magnesium chloride, 20% v/v glycerol. Beta-mercapto-ethanol and protease inhibitor are added separately and will be at 10 mM and 1× concentration, respectively, in the final onefold dialysis buffer (*see* Subheadings 3.2, **steps 9** and **10**).
10. Resuspension/Wash 1 buffer: 1.5 M potassium acetate, 20 mM HEPES pH 7.8, 1 mM magnesium chloride, 10 mM imidazole, 10% v/v glycerol, 10 mM beta-mercapto-ethanol, 0.5 × PI.

2.3 Affinity- and Size Exclusion Chromatography

1. Ni-NTA Agarose (Qiagen).
2. Gravity flow column casings "Eco-Pac" (Bio-Rad).
3. Amicon Ultra Concentrators 100 kDa MWCO (Millipore).

4. MonoQ 10/100 GL column (GE Healthcare).
5. MonoS 5/50 GL column (GE Healthcare).
6. Superose 6 10/300 column (GE Healthcare).
7. Aekta FPLC system: Any system which can operate the above-mentioned columns can be used. The most recent version that offers all necessary options at time of publication is the "Aekta pure 25 M" system by GE Healthcare.
8. Ni-NTA Wash buffer 2: 300 mM potassium acetate, 20 mM HEPES pH 7.8, 1 mM magnesium chloride, 25 mM imidazole, 10% v/v glycerol, 10 mM beta-mercapto-ethanol.
9. Ni-NTA Elution buffer: 300 mM potassium acetate, 20 mM HEPES pH 7.8, 1 mM magnesium chloride, 200 mM imidazole, 10% v/v glycerol, 10 mM beta-mercapto-ethanol.
10. MonoQ buffer A: 20 mM HEPES pH 7.8, 1 mM magnesium chloride, 10% v/v glycerol, 5 mM DTT.
11. MonoQ buffer B: 2 M potassium acetate, 20 mM HEPES pH 7.8, 1 mM magnesium chloride, 10% v/v glycerol, 5 mM DTT.
12. SEC buffer: 60 mM ammonium sulfate, 5 mM HEPES pH 7.8, 1 mM magnesium chloride, 10 µM zinc chloride, 5 mM DTT.

3 Methods

3.1 Cell Growth and Fermentation

1. Thaw a cryo-stock aliquot of the *S. cerevisiae* CB010 strain with FLAG/10×His tagged A190 [5], streak it on a YPD-agar plate and incubate at 30 °C for 48 h (*see* **Note 1**).
2. In the morning, pick one or more large colonies and inoculate a 50 mL culture of YPD supplemented with 50 µg/mL Ampicillin (1:2000 dilution of the stock) and 12.5 µg/mL Tetracycline (1:1000 dilution of the stock). Grow the culture at 30 °C for 10–14 h while shaking at 130 rpm.
3. Inoculate a 500 mL culture (YPD supplemented with antibiotics) with 40 mL of the preculture and grow overnight at 30 °C while shaking at 130 rpm.
4. On the next morning, inoculate three 2 L cultures (YPD with antibiotics) with 150 mL of the preculture and grow over 10–14 h at 30 °C while shaking at 130 rpm.
5. While the final precultures are growing, prepare a 200 L fermenter including YPD medium and autoclaving.
6. Measure OD_{600} and inoculate the 200 L fermenter with the appropriate amount of cells to reach a starting OD_{600} of 0.10–0.15. Supplement with 50 µg/mL Ampicillin and 12.5 µg/mL Tetracycline as well as 200 mL Antifoam C Emulsion (1:10) (*see* **Notes 2** and **3**).

7. Operate the fermenter at 30 °C, 250 rpm stirring, and an air influx of 80 NL/h. Monitoring of pH values with a probe and constant pH adjustment are not necessary.
8. Grow the cells overnight until an OD_{600} of 3–4 is reached (this will take around 10 h). Harvest the cells with a flow-through centrifuge or a filter device (*see* **Note 4**).
9. Resuspend the yeast pellet in lysis buffer (1 mL for each 2 g of pellet) on ice. Aliquot the cells into 225 mL portions and flash-freeze in liquid nitrogen for 5–10 min. Pellets can be stored at −80 °C until use (*see* **Notes 5** and **6**).

3.2 Breaking the Cells and Removal of Debris

1. Thaw two pellets of 225 mL 10×His-tagged Pol I (*see* Subheading 3.1) in warm water. Vortex in intervals to ensure complete solubilization (*see* **Note 7**).
2. Add 35 mL of 3 M ammonium sulfate and an additional 1 mL of 100× protease inhibitor (EDTA free). At this point, move to a 4 °C cold room and perform all steps on ice.
3. Fill 200 mL of glass beads into a bead-beater chamber and mix them with one complete pellet. Close the chamber with a well-sealed stirrer (attached to a lid) and make sure no air is left inside (*see* **Note 8**).
4. Mount the chamber on a bead beater and distribute an ice/salt mixture over the outside of the bucket before you start beating. Using an interval of 30 s beating followed by a 90 s idle period for cooling is recommended. Perform bead beating for a total of 90 min (*see* **Notes 9** and **10**).
5. Before opening the beater bucket, discard the water/salt mixture on the outside and clean well. Open the bucket and decant the entire glass bead/broken cell mixture in a sieve to filter off the beads. Collect the flow-through and wash the beads with 50 mL of dilution buffer for each pellet (*see* **Note 11**).
6. Combine both flow-through- and wash-fractions and centrifuge at 10,000 × *g* and 4 °C for 30 min. Decant the supernatant into a fresh beaker.
7. Ultracentrifuge the supernatant for 90 min at 30,000 × *g* and 4 °C with a swinging bucket rotor.
8. Open the tubes and remove the upper (fat-) layer, e.g., with a vacuum pump. Aspire the cleared supernatant with a 25 mL pipet and transfer it into a fresh beaker. Do not disturb the pellet which might be rather soft and not clearly defined. Discard the pellet.
9. Prepare three 2 L beakers for dialysis by filling them with 1 L of water and adding 1 L of 2× Dialysis Buffer under stirring. Add freshly prepared PMSF and benzamidine (dissolved in 80% ethanol) to a final concentration of 1 mM each.

10. Presoak six dialysis tubes in Millipore water for 30 min and fill them with 50 mL of ultracentrifuge supernatant each. Close the tubes and insert two filled ones into each 2 L dialysis beaker. Add 1.4 mL of beta-mercapto-ethanol under gentile stirring to reach a total concentration of 10 mM. Incubate over night at 4 °C.
11. Resolve the dialysis by removing the tubes from the beakers and opening them on the bottom (*see* **Note 12**).
12. Combine all six individual dialyzed samples and centrifuge at 18,500 × *g* and 4 °C for 60 min.
13. Carefully decant the supernatant and discard it. Resolubilize the pellet with 30 mL Resuspension Buffer until no debris is visible any longer. Transfer the entire sample into a 50 mL Falcon tube (*see* **Note 13**).
14. Fill the tube to 40 mL with Resuspension Buffer and incubate on a slowly rotating wheel at 4 °C for 15 min.

3.3 Affinity- and Size Exclusion Chromatography

1. Prepare three empty cases of 20 mL gravity flow columns at 4 °C and distribute 16 mL (slurry) Ni-NTA beads (8 mL bed volume) among them. Wash the beads with 2 column volumes (CV) of water followed by 3 CV of Resuspension Buffer.
2. Remove the Ni-NTA beads and combine them with the resuspended dialysis pellet from Subheading 3.2 in a single tube when no debris is visible any longer (*see* **Note 14**).
3. Incubate on a gently rotating wheel for 4 h at 4 °C.
4. Now distribute the entire volume over the three gravity-flow columns and let the buffer flow through.
5. Subsequently wash with 5 CV Resuspension Buffer followed by 5 CV of Wash Buffer 2.
6. Elute Pol I with 5 CV of Ni-NTA Elution Buffer.
7. Centrifuge the eluate for 15 min at 10,000 × *g* and 4 °C and prepare a 10/100 MonoQ column using an Aekta FPLC system.
8. Equilibrate your column to the potassium acetate concentration used in the Ni-NTA elution buffer (0.3 M potassium acetate, 15% MonoQ buffer B). Apply the complete volume of sample, e.g., using a 50 mL Superloop.
9. Wash out unbound sample (2 CV) and run a gradient from 0.3 to 2.0 M potassium acetate (15–100% MonoQ buffer B). Collect all fractions and run an SDS-PAGE. Coomassie (or similar) staining will show at which peak Pol I elutes, the two large bands of subunits A190 and A135 should be clearly visible whereas other bands might be harder to identify. An absorption peak at a conductivity of 51.5 mS/cm should contain your Pol I (*see* **Notes 15** and **16** and Fig. 1).

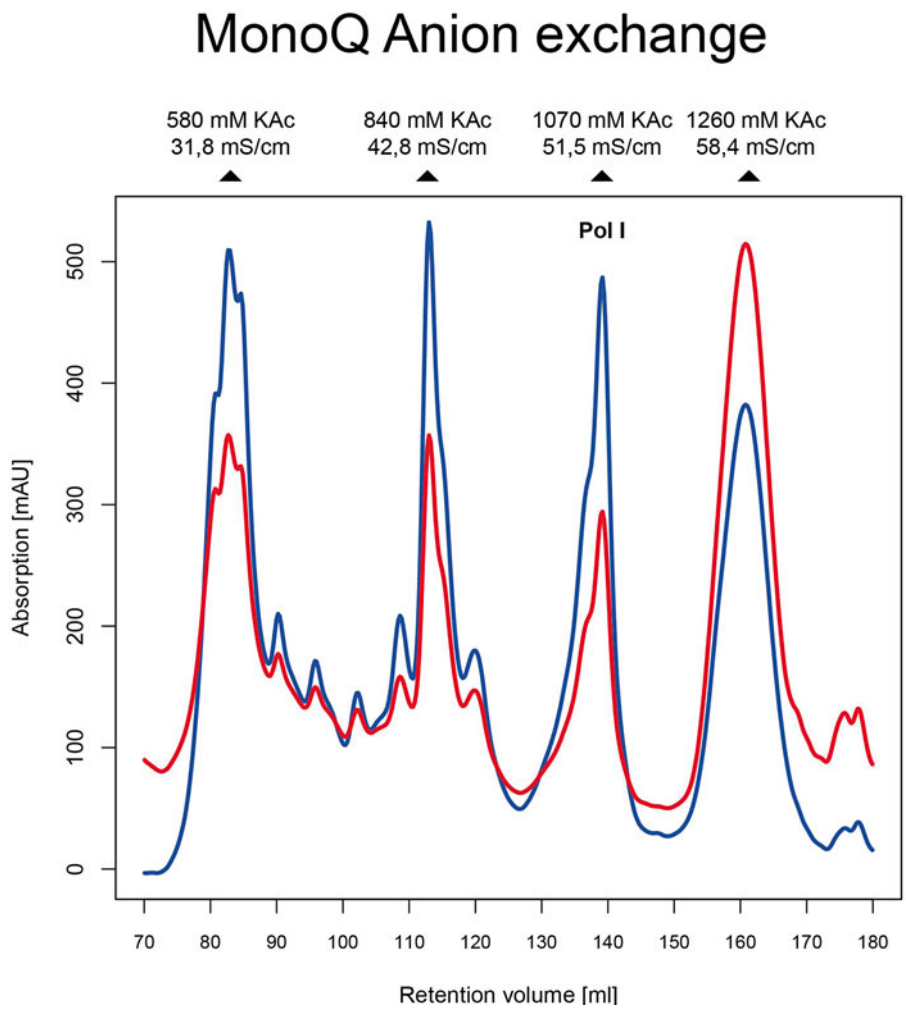

Fig. 1 MonoQ Anion exchange chromatography profile from a 10/100 MonoQ GL column. Absorption at 280 nm (A280) is displayed in *blue* and absorption at 260 nm (A260) in *red*. Elution parameters are indicated above the respective peaks: Potassium acetate concentration and conductivity. At this stage, numerous contaminating peaks are still present and the Pol I containing peak is marked

10. Pool the entire peak and dilute to a salt concentration of 0.1 M potassium acetate using MonoQ Buffer A.
11. Load the entire sample on a 5/50 MonoS column and wash out unbound protein with 2 CV of 0.2 M potassium acetate buffer (10% MonoQ buffer B). Run the following three-step gradient and collect fractions of 0.5 mL: (1) 2 CV to 17.5% buffer B (0.35 M), (2) a 3 CV plateau 17.5% buffer B, (3) a 17.5 CV gradient to 35% buffer B (0.7 M). Pol I elutes in a large peak at 0.47 M potassium acetate which corresponds to a conductivity of 26.9 mS/cm (*see* Fig. 2).

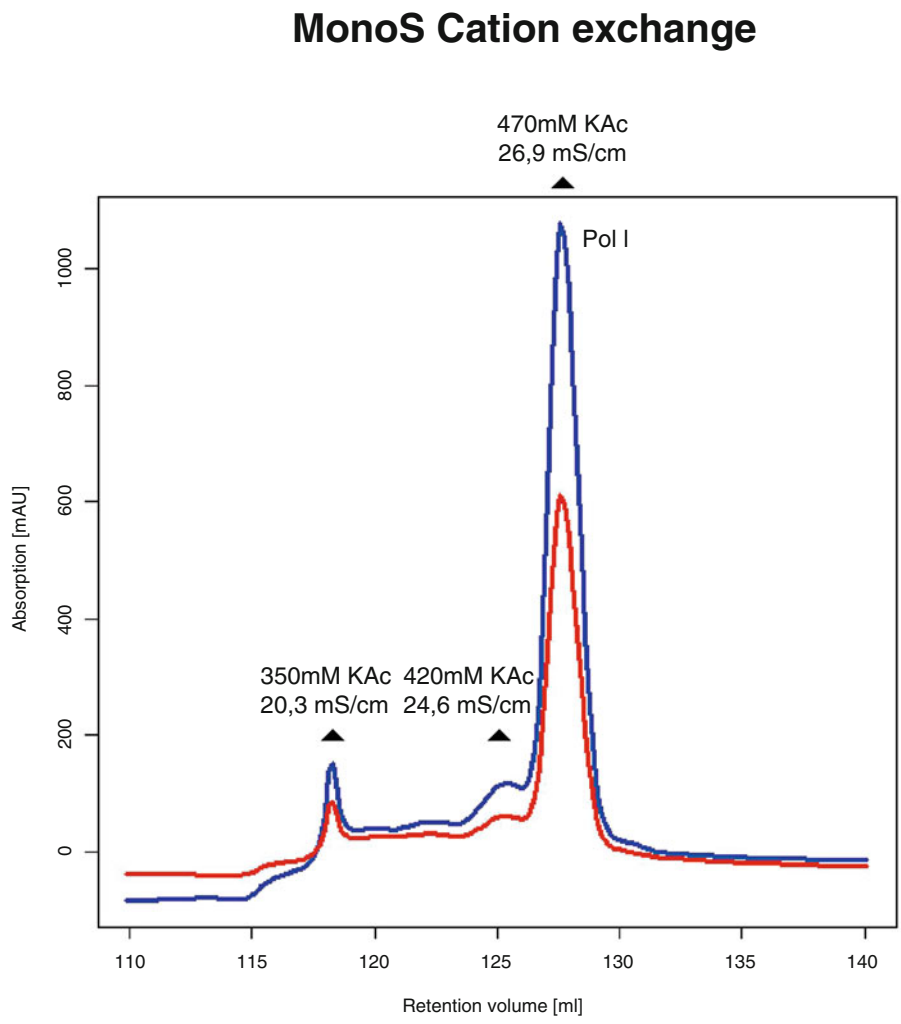

Fig. 2 MonoS Cation exchange chromatography profile from a 5/50 MonoS GL column. Absorption at 280 nm (A280) is displayed in *blue* and absorption at 260 nm (A260) in *red*. Elution parameters are indicated above the respective peaks: Potassium acetate concentration and conductivity. At this stage, purity is remarkably increased compared to the MonoQ step (*see* Fig. 1) and last contaminating peaks are removed. The Pol I containing peak is indicated

12. Pool the MonoS peak and concentrate to 0.5 mL using a 100 kDa cutoff concentrator.
13. Equilibrate a 10/300 Superose 6 column with SEC Buffer and inject the entire sample. Collect fractions of 0.5 mL (*see* **Note 17**).
14. Check all relevant fractions on a SDS-PAGE. Crystallization grade Pol I elutes in a single peak at a retention volume of around 11.9 mL (*see* Figs. 3 and 4).
15. Use a 100 kDa cutoff concentrator until the desired concentration is reached. Flash freeze aliquots at liquid nitrogen and store them at −80 °C until use (*see* **Notes 18** and **19**).

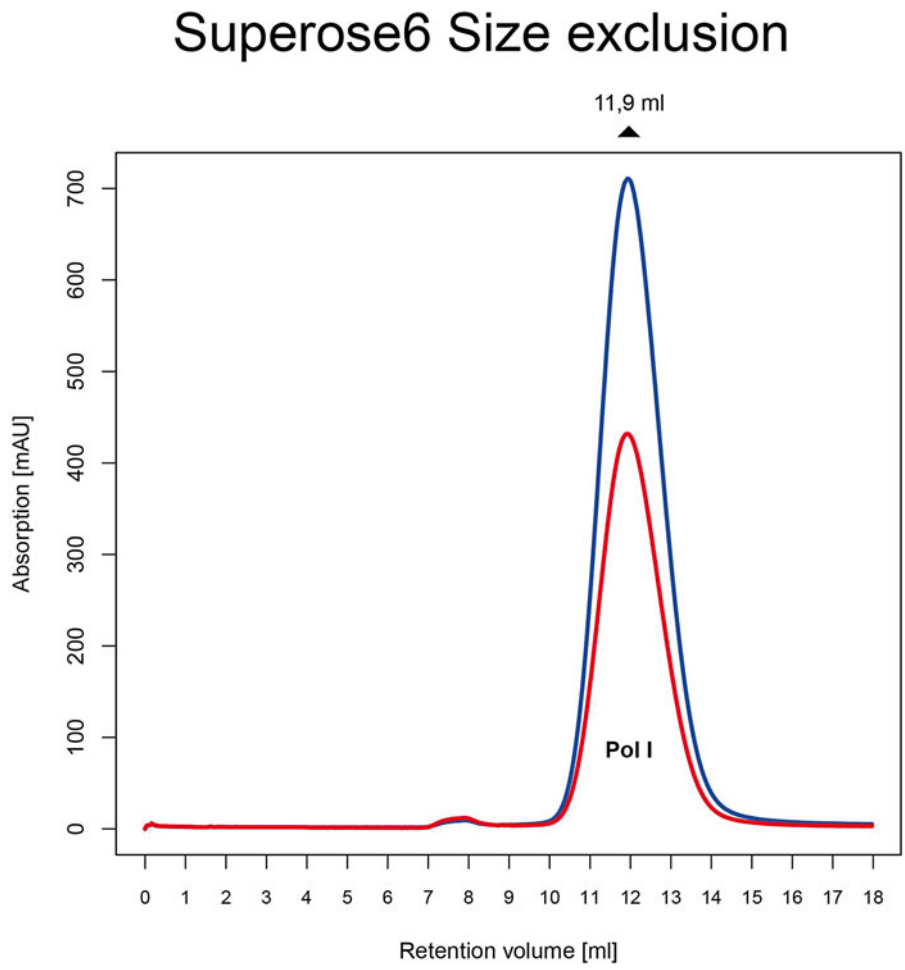

Fig. 3 Size exclusion chromatography profile from a 10/300 Superose 6 column. Absorption at 280 nm (A280) is displayed in *blue* and absorption at 260 nm (A260) in *red*. A small portion of aggregates is removed and ~99 % purity achieved. Compare Fig. 4 for a Coomassie stained SDS-PAGE of the marked peak

4 Notes

1. In general and throughout the entire procedure, all steps should be carried out as precise and as quickly as possible. Longer waiting times can result in partial polymerase degradation and/or lower binding efficiencies in the following steps.
2. Before fermentation it is recommended to visually inspect your yeast cells with a standard light microscope in order to make sure no bacterial contamination is introduced into your large-scale fermentation. The CB010 strain is deficient in some proteases and cells tend to cluster more than wild type. This is expected and does not influence the experiment.

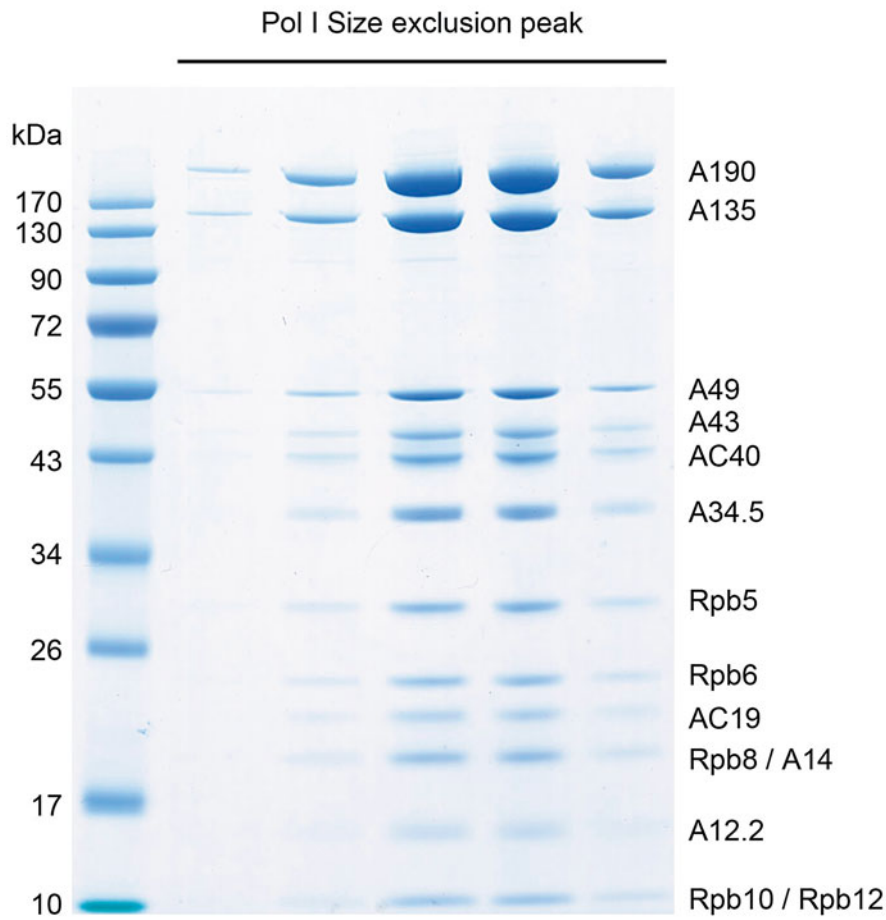

Fig. 4 Coomassie-stained SDS-PAGE (12 % Acrylamide) of the final SEC peak (compare Fig. 3). A molecular weight standard is indicated. All 14 subunits of Pol I are visible and their identity was confirmed by mass spectrometry

3. At the end of the fermentation, make sure that the cells do not overgrow. A low OD_{600} ensures that cells have not transitioned into stationary phase and thereby optimal rRNA synthesis rates. A centrifuge or filter system should be used for harvesting that can process the high volume of the fermenter efficiently. If no such centrifuge is at hand, it may be better to stop fermentation at an OD_{600} of as low as 2.0.
4. From a 200 L fermentation, the yeast pellet should have a weight of around 1000 g.
5. Make sure that the entire yeast pellet is dissolved in lysis buffer before aliquoting.
6. Freezing as much as 225 mL cell suspension in the 20 % glycerol buffer may take a while. The pellets should stay in liquid nitrogen until the boiling fades (5–10 min). Resupply nitrogen in case the evaporation is too high.

7. While thawing the yeast pellets, take care that the tubes are not getting too hot by inverting and vortexing them in regular intervals.
8. Before closing a bead-beater bucket, mix the beads and the suspended cells well. It can be helpful to add some mL of Dilution Buffer to a bucket so that it is entirely filled before closing it with the lid and its integrated stirrer. This will lead to a slight overflow of the bucket during closing, but ensures that no air is trapped inside, which could lead to increased foaming (and thereby denaturation of protein) during lysis.
9. Be sure that the buckets are entirely closed and cool them during the entire bead-beating procedure. Cooling the bucket is necessary to counteract overheating and is best achieved with a mixture of ice with salt. For this, simply use commercial salt (e.g., used for roadworks) and mix it with ice to achieve temperatures of −8 to −10 °C. It may be beneficial to exchange this salt/ice mixture after 30–45 min and check for overheating of the beaters during the lysis.
10. Running bead beaters in 30 s on–90 s off intervals is helpful to prevent overheating. A programmable plug-in for the power supply does the job very well.
11. For filtering the beads from the lysate, we prepare a thin net (fly-screen works well) and span it on the top of a plastic container. Now the container can be cut open on its bottom, placed onto a glass beaker, and filled with the lysate that needs filtering.
12. After the overnight dialysis is stopped, prepare 100 mL Resuspension Buffer by freshly adding PMSF and benzamidine to a total concentration of 0.5 mM each. Also add 70 µL of beta-mercapto-ethanol to obtain a total concentration of 10 mM. Generally, add protease inhibitors and reducing agents such as DTT and beta-mercapto-ethanol just before the use of a buffer.
13. Do not vortex while solubilizing the dialysis pellet even though the process might take several minutes of repeated pipetting.
14. For incubation of the dialyzed and resolubilized sample with Ni-NTA beads, it is best to use a high amount of Ni-NTA beads since a lot of contaminants are still present at this stage and might influence the efficiency of Pol I binding. Furthermore, a low total volume (50 mL in one tube) increases the local concentration and might improve binding. Ni-NTA beads can be stripped, stringently washed, and reused.
15. While running the MonoQ column in succession of the Ni-NTA affinity step, I recommend dividing your gradient into three parts in order to increase resolution in the interesting segment: (1) A steep gradient from 0.3 to 0.5 M potassium

acetate (15–25% MonoQ Buffer B) over 1 CV, (2) a shallow gradient from 0.5 to 1.4 M (25–70%), (3) a final steep gradient from 1.4 to 2.0 M potassium acetate (70–100%). The polymerase will elute at 1.07 M potassium acetate.

16. After the MonoQ column has finished, it is safe to keep Pol I in the elution buffer over night. I recommend running the MonoS column on the next day directly followed by the final SEC.
17. The final SEC run may be omitted, in case the prepared polymerase will later on be used for interaction experiments which include another gel filtration.
18. Starting with two yeast pellets (corresponds to ~300 g cells) the final yield can be up to 1.2 mg. Upscaling is feasible for up to four pellets but comes along with logistical short passes (e.g., four ultracentrifuges are needed at the same time).
19. If this purification is performed for the first time, I recommend taking samples at all stages and running an SDS-PAGE which can be analyzed by western blotting or (in later stages) directly stained with Coomassie. Furthermore, a functional test can be carried out in the form of a simple fluorescence-based elongation essay as described [1, 5].

Acknowledgements

I thank Patrick Cramer for supervising and supporting all research and method development. I acknowledge the contribution of all past and present members of the Pol I subgroup to a constant evolution of the protocol, especially Claus Kuhn and Stefan Jennebach. Thanks also to Carlo Bäjen, Tobias Gubbey, Sarah Sainsbury, Andrea Stutz and Youwei Xu. This work was funded by the Boehringer Ingelheim Fonds and the Max Planck Society.

References

1. Kuhn CD, Geiger SR, Baumli S, Gartmann M, Gerber J, Jennebach S, Mielke T, Tschochner H, Beckmann R, Cramer P (2007) Functional architecture of RNA polymerase I. Cell 131(7):1260–1272. doi:10.1016/j.cell.2007.10.051
2. Roeder RG, Rutter WJ (1969) Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature 224(5216):234–237
3. Vannini A, Cramer P (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell 45(4):439–446. doi:10.1016/j.molcel.2012.01.023
4. Warner JR (1999) The economics of ribosome biosynthesis in yeast. Trends Biochem Sci 24(11):437–440
5. Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P (2013) RNA polymerase I structure and transcription regulation. Nature 502(7473):650–655. doi:10.1038/nature12712
6. Fernandez-Tornero C, Moreno-Morcillo M, Rashid UJ, Taylor NM, Ruiz FM, Gruene T, Legrand P, Steuerwald U, Muller CW (2013) Crystal structure of the 14-subunit RNA polymerase I. Nature 502(7473):644–649. doi:10.1038/nature12636

7. Kostrewa D, Kuhn CD, Engel C, Cramer P (2015) An alternative RNA polymerase I structure reveals a dimer hinge. Acta Crystallogr D Biol Crystallogr 71(Pt 9):1850–1855. doi:10.1107/S1399004715012651
8. Geiger SR, Lorenzen K, Schreieck A, Hanecker P, Kostrewa D, Heck AJ, Cramer P (2010) RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. Mol Cell 39(4):583–594. doi:10.1016/j.molcel.2010.07.028
9. Moreno-Morcillo M, Taylor NM, Gruene T, Legrand P, Rashid UJ, Ruiz FM, Steuerwald U, Muller CW, Fernandez-Tornero C (2014) Solving the RNA polymerase I structural puzzle. Acta Crystallogr D Biol Crystallogr 70(Pt 10):2570–2582. doi:10.1107/S1399004714015788
10. Appling FD, Schneider DA (2015) Purification of active RNA polymerase I from yeast. Methods Mol Biol 1276:281–289. doi:10.1007/978-1-4939-2392-2_16

Chapter 8

Analysis of *S. cerevisiae* RNA Polymerase I Transcription In Vitro

Michael Pilsl, Philipp E. Merkl, Philipp Milkereit, Joachim Griesenbeck, and Herbert Tschochner

Abstract

RNA polymerase I (Pol I) activity is crucial to provide cells with sufficient amounts of ribosomal RNA (rRNA). Synthesis of rRNA takes place in the nucleolus, is tightly regulated and is coordinated with synthesis and assembly of ribosomal proteins, finally resulting in the formation of mature ribosomes. Many studies on Pol I mechanisms and regulation in the model organism *S. cerevisiae* were performed using either complex in vitro systems reconstituted from more or less purified fractions or genetic analyses. While providing many valuable insights these strategies did not always discriminate between direct and indirect effects in transcription initiation and termination, when mutated forms of Pol I subunits or transcription factors were investigated. Therefore, a well-defined minimal system was developed which allows to reconstitute highly efficient promoter-dependent Pol I initiation and termination of transcription. Transcription can be initiated at a minimal promoter only in the presence of recombinant core factor and extensively purified initiation competent Pol I. Addition of recombinant termination factors triggers transcriptional pausing and release of the ternary transcription complex. This minimal system represents a valuable tool to investigate the direct impact of (lethal) mutations in components of the initiation and termination complexes on the mechanism and regulation of rRNA synthesis.

Key words RNA polymerase I, In vitro transcription, *Saccharomyces cerevisiae*

1 Introduction

Cell free extracts, cell fractionation, and reconstitution of transcription in vitro served as important prerequisites to identify and characterize Pol I transcription initiation and termination [1–7]. Transcription termination in vitro was mainly studied using tailed templates on which purified RNA polymerase starts RNA synthesis at 3′ extended templates [1, 7–9]. This allowed investigating mutations in both DNA cis elements and in transacting factors involved in termination, but represents a rather unphysiological situation. In these reactions, transcription bubble formation might be compromised leading to persistent RNA–DNA hybrids behind the elongating polymerase [10].

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_8,

A ternary transcription complex reconstituted with DNA, RNA, and purified RNA polymerase I might mimic a transcription bubble [11, 12], but it is unclear how well this experimental system represents the *bona fide* transcription process. Reconstitution of efficient transcription at the yeast Pol I promoter required the addition of four initiation factors, CF (core factor) consisting of Rrn6, Rrn7 and Rrn11, UAF (upstream activating factor) consisting of Rrn5, Rrn9, Rrn10, UAF30 and the histones H3 and H4, the TATA binding protein TBP, and the Pol I-associated factor Rrn3 [3]. Whereas several initiation factors (CF, Rrn3, TBP) could be recombinantly expressed and purified to obvious homogeneity, strong transcription initiation could only be achieved adding purified UAF fractions and fractions containing Pol I or Pol I-Rrn3 enriched from yeast cell extracts [5, 6]. Here, we present a protocol to reconstitute efficient transcription initiation in vitro using only highly purified Pol I–Rrn3 complexes from yeast whole cell extracts and overstoichiometric amounts of CF expressed and purified from baculovirus-infected insect cells. The same system can be used to investigate Pol I termination after addition of highly purified recombinant termination factor Nsi1.

2 Materials

2.1 Preparation of Immobilized Templates

1. Streptavidin-coupled magnetic beads, Dynabeads.
2. BcMag Separator (Bioclone Inc., MS-04).
3. Coupling buffer 5 mM Tris/HCl pH 7.5, 0.5 mM EDTA, and 1 M NaCl.
4. TE buffer 10 mM Tris/HCl pH 7.5, 1 mM EDTA.

2.2 Purification of Transcription Termination Factor Nsi1

1. Branson sonifier 250 macrotip.
2. Anti-flag® M2 Affinity Agarose (Sigma).
3. MobiCol column (MoBiTec).
4. TAP 100 buffer (50 mM HEPES/KOH pH 7.8, 100 mM KCl, 5 mM $Mg(OAc)_2$, 0.1% Tween-20).
5. TAP 300 buffer (50 mM HEPES/KOH pH 7.8, 300 mM KCl, 5 mM $Mg(OAc)_2$, 0.1% Tween-20).
6. 100× protease inhibitor stock solution (PIs): 200 mM benzamidine, 100 mM PMSF in ethanol.

2.3 Purification of Transcription Initiation Factor CF

1. Branson sonifier 250 macrotip.
2. Anti-flag® M2 Affinity Agarose (Sigma).
3. MobiCol column (MoBiTec).
4. TAP 100 buffer (50 mM HEPES/KOH pH 7.8, 100 mM KCl, 5 mM $Mg(OAc)_2$, 0.1% Tween-20).

5. TAP 200 buffer (50 mM HEPES/KOH pH 7.8, 200 mM KCl, 5 mM $Mg(OAc)_2$, 0.1% Tween-20).
6. 100× protease inhibitor stock solution (PIs): 200 mM benzamidine, 100 mM PMSF in ethanol.

2.4 Purification of Initiation Competent Pol I–Rrn3 Complexes

1. Yeast strain BSY420-RRN3-TEV-ProtA-His7 (ade2-1, can1-100, his3-200, leu2-3,112 trp1-1, ura3-1, RRN3-TEV-ProtA-His7).
2. Plasmid Ycplac111-GAL-Rrn3-TEV-ProtA-His_7 (LEU2).
3. YPD medium 2% w/v Bacto™ Peptone, 1% w/v Bacto™ Yeast Extract, 2% w/v glucose.
4. YPR medium: YP medium using 2% w/v raffinose as sugar.
5. SCR-Leu: 0.67% w/v YNB + Nitrogen, 0.063% w/v CSM-His-Leu-Trp, 2% w/v raffinose + 20 mg/L L-histidine + 50 mg/L L-tryptophan.
6. Yeast lysis buffer: 0.15 M HEPES/KOH pH 7.8, 40% glycerol, 0.4 M $(NH_4)_2SO_4$, 60 mM $MgCl_2$, 3 mM DTT.
7. 100× protease inhibitor stock solution (PIs): 200 mM benzamidine, 100 mM PMSF in ethanol.
8. Bead beating chamber 400 mL stainless steel (Biospec).
9. Buffer A: 20% glycerol, 20 mM HEPES/KOH pH 7.8, 10 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT, 1× PIs.
10. Buffer A90: buffer A including 90 mM KCl.
11. Buffer A350: buffer A including 350 mM KCl.
12. Buffer B: 20% glycerol, 20 mM HEPES/KOH pH 7.8, 2 mM $MgCl_2$, 1 mM DTT, 0.15% NP40.
13. Buffer B200: buffer B including 200 mM potassium acetate, but no NP40.
14. Buffer B600: buffer B including 600 mM potassium acetate.
15. Buffer B1500: buffer B including 1500 mM potassium acetate.
16. IgG (rabbit serum, I5006-100MG, Sigma) -coupled magnetic beads (1 mm BcMag, FC-102, Bioclone). A detailed protocol of the preparation of the beads has been described in [13].

2.5 Promoter-Dependent Transcription Assay

1. Transcription buffer 2×: 40 mM HEPES/KOH pH 7.8, 20 mM $MgCl_2$, 10 mM EGTA, 5 mM DTT, 300 mM potassium acetate, 0.4 mM ATP, 0.4 mM GTP, 0.4 mM UTP, 0.02 mM CTP, 0.3 μCi ^{32}P-CTP.
2. Proteinase K buffer: 10 mM Tris/HCl pH 7.5, 5 mM EDTA, 0.3 M NaCl, 0.6% SDS, 0.5 mg/mL proteinase K.
3. 6% polyacrylamide gel (180 × 180 × 0.5 mm) containing 7 M urea and 1× TBE.

3 Methods

3.1 Preparation of Immobilized Templates

1. PCR reactions are performed using one biotinylated primer around 20 nucleotides downstream of the yeast terminator site which hybridizes to 3′end of the transcription template and one primer binding around 20 nucleotides upstream of the Pol I promoter (Fig. 1a) [7].
2. 10 µg of biotinylated PCR fragment are incubated with 1 mg streptavidin-coupled magnetic beads in 200 µL coupling buffer for 30 min at room temperature. The suspension is washed three times with 200 µL of the same buffer and three times with 200 µL 20 mM HEPES/KOH pH 7.8 using a magnetic support (BcMag Separator) to separate immobilized DNA from supernatant. The beads are resuspended in

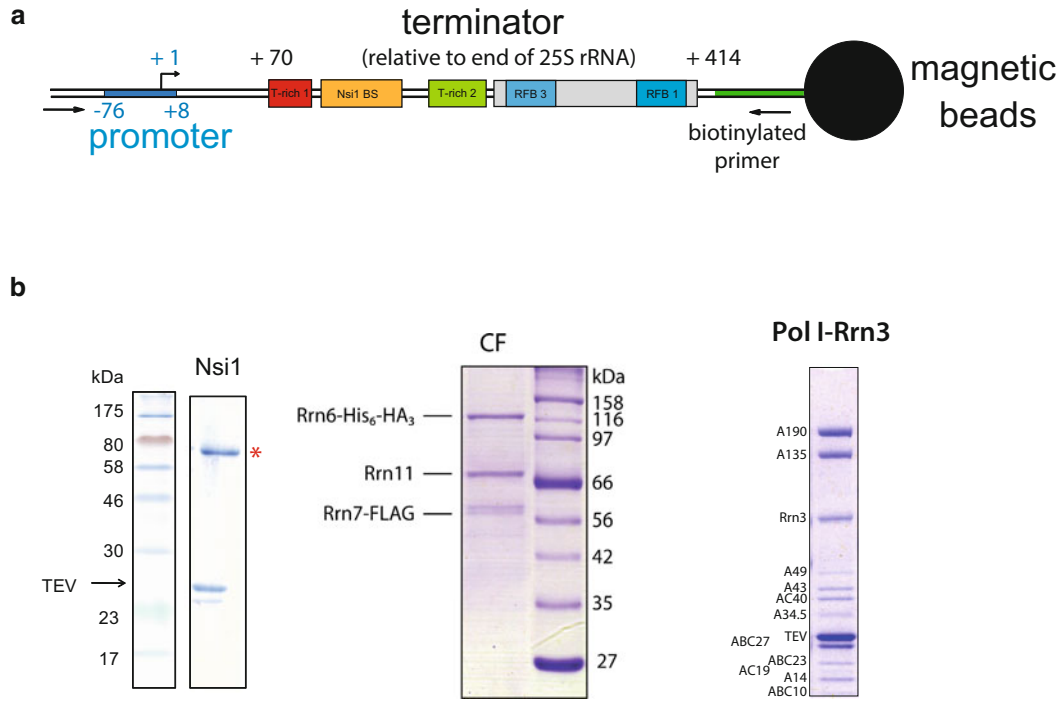

Fig. 1 Overview about the components of the in vitro transcription reaction. (**a**) Schematic representation of the immobilized DNA template for analysis of promoter-dependent Pol I transcription initiation and termination. The promoter region contains the core element to which CF and Pol I–Rrn3 are recruited (+1 indicates the transcription start site). The terminator element starts 70 nt downstream of the 3′end of the 25S rRNA (+70), contains two T-rich elements, the Nsi1- and multiple Fob1-binding sites within the replication fork barrier (RFB, framed by the elements RFB3 and RFB1), and ends at position +414 downstream of the 3′end of the 25S rRNA. The positions of the (biotinylated) primers used for PCR amplification of the DNA fragment are indicated as arrows. (**b**) Coomassie stained SDS PAGE of purified protein fractions which are sufficient to analyze promoter-dependent Pol I initiation and termination in vitro. Sizes of molecular weight marker proteins are given for the panels at the *left* and in the *middle* of the figure. The positions of protein bands and their respective name in the lanes containing the purified factors are indicated

100 µL TE buffer. Use 0.5–1 µL (50–100 ng) for transcription reactions.

3.2 Purification of Transcription Termination Factor Nsi1

1. Cloning of the NSI1 gene for expression as fusion protein with N-terminal FLAG-tag and recognition site for TEV protease in baculovirus-infected SF21 insect cells has been described in [7].
2. 50×10^6 infected SF21 cells are resuspended in 40 mL ice cold TAP 300 buffer. All subsequent steps are performed on ice or at 4 °C. The cell suspension is sonicated in a 50 mL reaction tube using a Branson sonifier 250 macrotip at output 5 for 30 s followed by a pause of 30 s. This cycle is repeated five times.
3. Cell debris is sedimented at $4000 \times g$ for 15 min.
4. The clear lysate is added to 50 µL (100 µL slurry) anti-flag® M2 Affinity Agarose which was equilibrated three times with 5 mL TAP 300 buffer.
5. The suspension is incubated on a rotating wheel at 4 °C for 2 h.
6. Beads are sedimented at $130 \times g$ for 5 min at 4 °C, the supernatant is removed, and the beads are washed three times with 10 mL TAP 100 buffer.
7. Beads are resuspended in 100 µL TAP 100 buffer, transferred to a 200 µL reaction tube, and incubated in the presence of 20.8 µg TEV protease for 2 h at 16 °C under gentle shaking.
8. After centrifugation at $16{,}000 \times g$ for 5 min at 4 °C, the supernatant is loaded on a MobiCol column (MoBiTec) and spun through the column at $16{,}000 \times g$ for 5 min in an empty reaction tube to remove remaining beads. Samples taken during the procedure are analyzed by SDS PAGE to monitor the purification of Nsi1 and the protein concentration in the final elution fraction is determined. Nsi1 concentration is adjusted to 1.5 pmol/µL in TAP 100 buffer (*see* Fig. 1b).

3.3 Purification of Transcription Initiation Factor CF

1. Cloning of the genes coding for CF subunits for coexpression in baculovirus-infected SF21 insect cells has been described in [7]. Purification is performed using a FLAG-tag fused to the C-terminus of the CF subunit Rrn7 (*see* **Note 1**).
2. 50×10^6 cells are resuspended in 30 mL ice-cold TAP100 buffer including 1× PIs and lysed using a Branson Sonifier 250 (output 5, duty cycle 40 %, 30 s pulse, 30 s cooling, six repeats). All subsequent steps are performed on ice or at 4 °C.
3. Cell debris is sedimented at $4000 \times g$ for 15 min.
4. The cleared cell lysate is incubated with 200 µL equilibrated anti-flag® M2 Affinity Agarose (*see* Subheading 3.2, **step 4**) for 2 h at 4 °C while inverting.
5. The beads are washed three times with 10 mL TAP100 buffer and TAP200 buffer.

6. Bound proteins are eluted adding 50–100 µL TAP200 buffer to the beads including 0.2 mg/mL FLAG peptide. Incubate for 3 h at 4 °C while inverting. The whole suspension is loaded on a disposable spin column (MobiCol) and the eluted proteins are separated from the agarose by centrifugation for 1 min at 3000 × *g* and 4 °C. The protein concentration is adjusted to 0.5–1 pmol/µL. 1–2 µL are used for one transcription reaction (40–80 nM end concentration) (*see* Fig. 1b).

3.4 Purification of Initiation Competent Pol I–Rrn3 Complexes

The Pol I–Rrn3 complex is purified from strain BSY420 Rrn3-ProtA, which overexpresses Rrn3 fused to a TEV cleavage site linked to a protein A-His_7-tag under the control of a GAL1/10 promoter on plasmid YCplac111-GAL-Rrn3-ProtA.

1. Strain BSY420 Rrn3-ProtA harboring plasmid YCplac111-GAL-Rrn3-ProtA is cultivated in SCR-Leu medium (500 mL preculture).
2. A 20 L fermenter is inoculated to an OD_{600} ~0.05 and the strain is grown overnight (doubling time approx. 250 min) in YPR medium until the culture reaches an OD_{600} of 1–2.
3. Rrn3 overexpression is induced for 3 h adding 2% w/v galactose.
4. The cells are harvested by centrifugation at 2000 × *g* for 2 min. Cells are washed twice with ice cold water and resuspended in yeast lysis buffer including 1× PIs (1 mL buffer per 1 g of cell paste).
5. The cell suspension is frozen in liquid nitrogen and stored at −80 °C.
6. Cells are thawed on ice. All subsequent steps are performed at 4 °C.
7. 250–300 mL cell suspension are placed in a 400 mL stainless steel bead beating chamber (Biospec) and glass beads (0.5 mm diameter) are added to fill the chamber.
8. Cells are disrupted using 10 cycles of bead beating for 30 s and cooling for 90 s. The temperature of the surrounding ice/salt bath should be kept at about −5 to −10 °C.
9. Glass beads and cell debris are removed by centrifugation for 10 min at 4000 × *g*.
10. The supernatant is centrifuged at 100,000 × *g* for 90 min.
11. The clear supernatant is carefully removed (about 300 mL). Avoid to include the turbid supernatant.
12. The clear supernatant is dialyzed against buffer A to reach the conductivity of buffer A90.
13. The dialyzed sample is loaded on a DEAE-Sepharose column (13 × 5.2 cm), equilibrated with buffer A90, washed with 1 L buffer A90 and proteins are eluted with buffer A350.

14. The protein-containing peak fractions (between 200 and 300 mL) are dialyzed against buffer B containing 1× PIs overnight. The volume of the dialysis buffer should be at least 10× the volume of the protein peak fraction.
15. The dialyzed sample is centrifuged at 40,000 × *g* for 30 min at 4 °C. The pellet which contains the initiation competent Pol I is resuspended in 2 mL buffer B600. The protein concentration is adjusted to 2.5–5 mg/mL.
16. The protein solution is centrifuged at 40,000 × *g* for 10 min to remove insoluble protein aggregates.
17. Equal protein amounts (10–20 mg) are incubated with 200 µL of IgG-coupled magnetic beads (1:1 slurry bead volume (100 µL); equilibrated three times with 500 µL buffer B600) for 2 h on a rotating wheel.
18. The beads are washed four times with 1 mL buffer B1500 and then three times with 1 mL buffer B200 without NP40 and PIs.
19. The beads are resuspended in 100 µL buffer B200 supplemented with 11.7 µg TEV protease and incubated for 2 h at 16 °C under shaking with 800 rpm in a thermomixer.
20. The supernatant is collected, the beads are washed with 2 × 50 µL buffer B200 and both wash steps are added to the eluted fraction.
21. 5–10% of the elution fraction are analyzed in an SDS PAGE. The protein concentration is adjusted to 0.05–0.1 pmol/µL (*see* Fig. 1b). The concentration of the Pol I–Rrn3 complex in the transcription reactions should be 4 nM.

3.5 Promoter-Dependent Transcription Assay

1. 1.5 mL reaction tubes (Sarstedt safety seal) are placed on ice. 0.5–1 µL immobilized template is added. This corresponds to 50–100 ng DNA, yielding a final concentration of 5–10 nM per transcription reaction (25 µL reaction volume).
2. 1 µL Nsi1 (1.5 pmol/µL; final concentration 40–80 nM) is added, if transcription should be terminated at the termination site and incubated 10 min on ice.
3. 1–2 µL CF (0.5–1 pmol/µL; final concentration 20–40 nM) and 1–3 µL Pol I–Rrn3 (final concentration 4–12 nM) are added to each tube.
4. 20 mM HEPES/KOH pH 7.8 are added to a final volume of 12.5 µL.
5. Transcription is started adding 12.5 µL transcription buffer 2×.
6. The samples are incubated at 24 °C for 30 min at 400 rpm in a thermomixer.
7. The reaction tubes are placed near a magnet and the supernatant is carefully transferred into a fresh 1.5 mL reaction tube.

8. 200 µL Proteinase K buffer are added to the supernatant. This fraction should contain transcripts which were released from the template DNA coupled to the beads.
9. 200 µL Proteinase K buffer are added to the magnetic beads pellet. These fractions contain nonreleased transcripts associated with stalled ternary transcription complexes.
10. The samples are incubated at 30 °C for 15 min at 400 rpm in a thermomixer.
11. 700 µL ethanol p.a. are added and mixed. Nucleic acids are precipitated at −20 °C over night or for 30 min at −80 °C.
12. The samples are centrifuged for 10 min at 12,000 × *g*. The supernatant is removed. The precipitate is washed with 0.15 mL 70% ethanol. After centrifugation, the supernatant is removed and the pellets are dried at 95 °C for 2 min.
13. RNA in the pellet is dissolved in 12 µL 80% formamide, 0.1× TBE, 0.02% bromophenol blue, and 0.02% xylene cyanol.

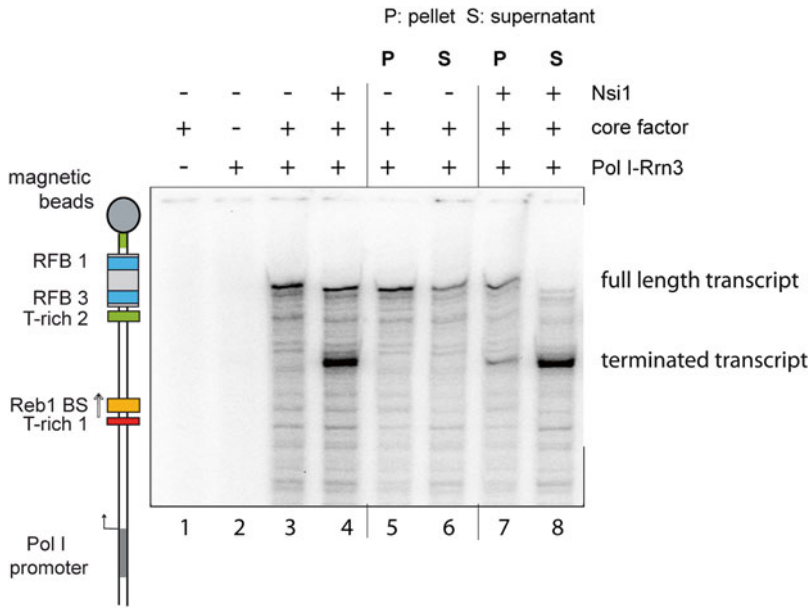

Fig. 2 Analysis of radioactively labeled RNAs synthesized by a minimal promoter-dependent Pol I transcription system in vitro. Binding of Nsi1 to its cognate binding site is sufficient to halt transcription and release transcripts in a promoter-dependent termination assay. Biotinylated templates which carry the biotinylation downstream of the Pol I promoter were immobilized on streptavidin-coupled magnetic beads. When termination is analyzed, Nsi1 (40–80 nM end concentration) is added to the DNA prior to transcription. Transcription was started by addition of RNA Pol I–Rrn3 complex (4 nM end concentration) and core factor (50 nM) and allowed to proceed for 30 min at room temperature. Samples were processed as a whole or fractionated in pellet (P) and supernatant (S) before RNA extraction. Presence of Nsi1, core factor, and Pol I–Rrn3 in the reactions is indicated. A cartoon on the *left* shows a schematic representation of the template used for transcription

14. Samples are heated for 2 min under vigorous shaking at 95 °C and briefly centrifuged.
15. Samples are loaded on a 6% polyacrylamide gel containing 7 M urea and 1× TBE. RNAs are separated applying 25 W for 30–40 min (for optimal resolution prerun the gel applying 25 W for 10–20 min). Xylene cyanol migrates 2/3 of the gel size.
16. Gel is dried after 10 min rinsing in water for 30 min at 80 °C using a vacuum dryer.
17. Radiolabeled transcripts are visualized using a PhosphoImager (Fig. 2).
18. *See* also **Notes 2–5**.

4 Notes

1. Purification of active CF can also be achieved using the protocol of Schneider and coworkers [6].
2. Transcription initiation in vitro does neither require the Pol I initiation factors TBP nor UAF, if CF is used in a fivefold excess over Rrn3–Pol I.
3. Ensure that the immobilized template and the protein factors are in suspension during transcription.
4. Salt optimum is between 100 and 300 mM potassium acetate for Pol I transcription. Higher or lower salt concentrations hamper Pol I in initiation.
5. Drying of the precipitated RNA for 2 min is important to remove all ethanol. However, longer heating is not advisable since RNA might get overdried and cannot be dissolved.

Acknowledgement

This work was supported by grants of the DFG (SFB 960). P. E. M. was partly supported by a fellowship of the German National Academic Foundation.

References

1. Lang WH, Morrow BE, Ju Q et al (1994) A model for transcription termination by RNA polymerase I. Cell 79(3):527–534
2. Schultz MC, Choe SY, Reeder RH (1991) Specific initiation by RNA polymerase I in a whole-cell extract from yeast. Proc Natl Acad Sci U S A 88(3):1004–1008
3. Keener J, Josaitis C, Dodd J et al (1998) Reconstitution of yeast RNA polymerase I transcription *in vitro* from purified factors. J Biol Chem 273:33795–33802
4. Milkereit P, Schultz P, Tschochner H (1997) Resolution of RNA polymerase I into dimers and monomers and their function in transcription. Biol Chem 378:1433–1443
5. Milkereit P, Tschochner H (1998) A specialized form of RNA polymerase I, essential for initiation and growth dependent regulation of

rRNA synthesis, is disrupted during transcription. EMBO J 17:3692–3703

6. Bedwell GJ, Appling FD, Anderson SJ et al (2012) Efficient transcription by RNA polymerase I using recombinant core factor. Gene 492(1):94–99
7. Merkl P, Perez-Fernandez J, Pilsl M et al (2014) Binding of the termination factor nsi1 to its cognate DNA site is sufficient to terminate RNA polymerase I transcription in vitro and to induce termination in vivo. Mol Cell Biol 34(20):3817–3827
8. Jeong SW, Lang WH, Reeder RH (1995) The release element of the yeast polymerase I transcription terminator can function independently of Reb1p. Mol Cell Biol 15(11):5929–5936
9. Lang WH, Reeder RH (1995) Transcription termination of RNA polymerase I due to a T-rich element interacting with Reb1p. Proc Natl Acad Sci U S A 92(21):9781–9785
10. Kadesch TR, Chamberlin MJ (1982) Studies of in vitro transcription by calf thymus RNA polymerase II using a novel duplex DNA template. J Biol Chem 257(9):5286–5295
11. Kuhn CD, Geiger SR, Baumli S et al (2007) Functional architecture of RNA polymerase I. Cell 131(7):1260–1272
12. Geiger SR, Lorenzen K, Schreieck A et al (2010) RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. Mol Cell 39(4):583–594
13. Hamperl S, Brown CR, Perez-Fernandez J et al (2014) Purification of specific chromatin domains from single-copy gene loci in Saccharomyces cerevisiae. Methods Mol Biol 1094:329–341

Chapter 9

Preparation of Chromatin Templates to Study RNA Polymerase I Transcription In Vitro

Gernot Längst

Abstract

Cellular DNA is packaged into chromatin, which is the substrate of all DNA-dependent processes in eukaryotes. The regulation of chromatin requires specialized enzyme activities to allow the access of sequence-specific binding proteins and RNA polymerases. In order to dissect chromatin-dependent features of transcription regulation in detail, in vitro systems to generate defined chromatin templates for transcription are required. I present a protocol that allows the assembly of nucleosomes on ribosomal RNA (rRNA) minigenes by salt gradient dialysis and subsequent sucrose gradient centrifugation. This procedure yields high nucleosome occupancy and high dynamic response in subsequent transcriptional analysis. It provides an invaluable tool to study rRNA gene transcription, as transcription on free DNA is clearly different from the more in vivo-like transcription on reconstituted chromatin templates.

Key words Chromatin assembly, In vitro reconstitution, Salt gradient dialysis, In vitro transcription, Ribosomal DNA

1 Introduction

It is well established that the chromatin structure is a key regulator of eukaryotic gene expression. The assembly of nucleosomes on DNA efficiently represses transcription by all three eukaryotic RNA polymerases, as the histone octameres on the gene promoters mask the binding sites for the transcription factors [1–4]. Activation in chromatin is mainly due to the antirepression of nucleosomal inhibition, repositioning the octamere such that it is compatible with the binding of the transcription initiation factors. Specific binding of pioneering factors and the activity of chromatin remodeling enzymes are required to allow efficient transcription activation and in the case of the mammalian rRNA gene, the Transcription Termination Factor I was shown to exhibit such activity [4–6]. Mouse ribosomal RNA genes are regulated by the epigenetic modification of DNA and histones, as well as by the specific positioning of the promoter-bound nucleosome that defines the ON- and OFF-states of the rRNA gene promoter [7].

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_9,

The TTF-I protein in mammals and proteins with similar functions in other taxonomic classes generally bind to more than one binding site at the terminator of the rRNA gene and one additional site at the gene promoter [7]. The termination activity of TTF-I and its chromatin-specific functions provide the unique possibility to study the transcriptional regulation of RNA Polymerase I on circular DNA plasmids, harboring so-called rRNA minigenes. A minigene constitutes of the rRNA gene promoter, a truncated rRNA gene sequence, and the respective termination region [5]. Recombinant TTF-I protein allows efficient transcription termination, generating a transcript of a defined length that can be quantified in the in vitro transcription experiments. The advantage of using the circular DNA lies in its beneficial properties during chromatin assembly, giving rise to nucleosome arrays of high occupancy and quality. Using such templates for in vitro transcription assays revealed a striking difference between free DNA and chromatin. On chromatin templates, but not on free rDNA, clustered binding sites at the terminator promote cooperative binding of TTF-I, loading TTF-I to the downstream terminators before it binds to the rDNA promoter. Clustered binding sites reconstituted into nucleosomes increase the binding affinity of transcription factors in chromatin, thus influencing the timing and strength of DNA-dependent processes [6] (Fig. 1).

In vitro transcription assays on a circular minigene containing the rDNA promoter fused a partial terminator containing the first five TTF-I binding sites revealed striking differences on free DNA and chromatin. On free DNA both read-through transcripts and a

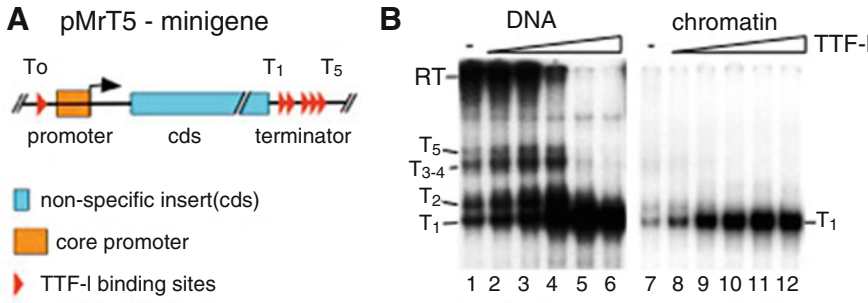

Fig. 1 Multiple termination sites are required for efficient transcription activation in chromatin and to form a defined terminated transcript. (**a**) Overview to the rDNA minigene pMrT5 harboring the mouse rRNA gene promoter including the TTF-I binding site T_0, a nonspecific insert as coding sequence, and the first five TTF-I binding sites (T_1–T_5) of the murine terminator region. (**b**) Analysis was performed comparatively on free DNA (*lanes 1–6*) or in vitro assembled chromatin (*lanes 7–12*), either in the absence (*lanes 1* and *7*) or presence of increasing levels of recombinant TTF-I (*lanes 2–6* and *8–12*). The radioactively labeled transcripts were separated by PAA gel electrophoresis and detected by autoradiography. Read-through transcripts (RT) and transcripts terminated at terminator sites ($T_{1–5}$) are indicated

heterogeneous population of transcripts randomly terminated at any of the TTF-I binding sites were synthesized due to subsaturating TTF-I levels in the nuclear extract (Fig. 1, lane 1). In the presence of increasing concentrations of recombinant TTF-I, the amount of transcripts stopped at terminator T_1 progressively increased, due to the saturation of the terminator sites (Fig. 1, lanes 1–6). A strikingly different result was obtained on rDNA templates assembled into chromatin (Fig. 1, lanes 7–12). Consistent with Pol I transcription on chromatin requiring binding of TTF-I to the promoter-proximal terminator T_0 and ATP-dependent chromatin remodeling [4, 5], transcription was repressed in the absence of TTF-I (no read-through transcripts (RT); Fig. 1, lane 7). The addition of TTF-I relieved transcriptional repression, however yielding only a single RNA terminated at T_1 (lanes 8–12). Already the lowest TTF-I concentration results in specific termination of the transcription at T_1. Besides the dramatic differences of the transcription results on free DNA and chromatin, the result suggests that transcription in chromatin is only initiated when the termination sites are fully occupied by TTF-I. We were able to show that on chromatin templates TTF-I binds with higher binding affinity to the terminator reconstituted in chromatin than to the single T_0 site at the promoter. Thereby defining the order of events during transcription initiation, with first binding to the terminators, second binding to the promoter and activating transcription, and third forming a DNA loop between the promoter and terminator for augmented transcription. The in vitro chromatin studies were nicely recapitulated by the in vivo analysis of TTF-I binding site function [6].

Classic chromatin assembly protocols make use of the fact that histones dissociate from DNA at salt concentration above 1.4 M NaCl and the reversibility of this process. Salt dialysis protocols, where histones and DNA are mixed at 2 M NaCl concentrations and then slowly dialyzed to low NaCl concentrations (50 mM NaCl), form high quality arrays of nucleosomes, minimizing the formation of nonspecific insoluble aggregates [8, 9]. The use of circular, supercoiled DNA enables highly efficient nucleosome assembly during salt dialysis, as each nucleosome accommodates one negative supercoil and the process is cooperative on circular and linear DNA [10]. This procedure yields histone octameres, densely packed on dsDNA but lacking the natural spacing between adjacent nucleosomal particles [8]. Generally, in vitro reconstitution of genomic DNA by salt gradient dialysis does not recapitulate the majority of in vivo nucleosome positions (=defined position of a particular nucleosome relative to a given base pair) [11]. However, like the “Sea urchin 5S rDNA” [12] and the “satellite DNA” [13], the rRNA gene contains a rather structural promoter with a high propensity of DNA bending that is capable to position nucleosome similar to the occupancy found in vivo [14–16].

Chromatin assembly systems based on extracts, like the *Drosophila* embryo extract [17] or purified histone chaperones with ATP-dependent remodeling enzymes are available [18] to generate spaced nucleosomes, but they usually do not achieve nucleosome positioning in the absence of sequence-specific binders [5, 19]. In the case studied here nucleosome spacing does not matter, as the nuclear extracts used for in vitro transcription contain the nucleosome spacing activities and chromatin modifying activities to allow efficient transcription from the salt dialyzed DNA templates. A comparison between chromatin templates reconstituted by salt dialysis from purified/recombinant histones revealed no functional difference in chromatin-dependent transcriptional regulation [20].

The chromatin assembly protocol consists of a titration of histone to DNA ratios and a subsequent MNase digestion analysis to determine the quality of the reconstituted chromatin (Fig. 2). Chromatin is incubated for increasing time with MNase, to progressively release nucleosomal particles. Digestion conditions have to be determined experimentally and should be established such that at highest incubation time only mono- and di-nucleosomes are visible and at the lowest digestion time still low levels of nondigested plasmid DNA can be observed (Fig. 3a). Different histone:DNA ratios give rise to characteristic digestion patterns (Fig. 2). In the case of substoichiometric assembly conditions, regions of free DNA exist that give rise to subnucleosomal DNA fragments, migrating below 147 bp (Fig. 3a, lanes 1–3). Generally only a discrete mono-nucleosomal DNA fragment can be observed at these conditions. With increasing histone:DNA ratios the levels of subnucleosomal DNA decrease and more of the di-, tri-, and

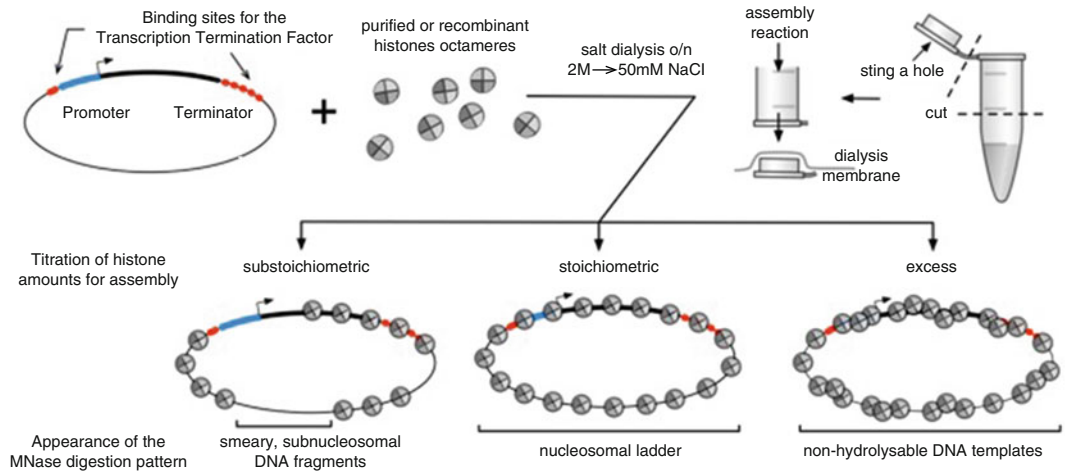

Fig. 2 Overview to the minigene constructs and the salt dialysis method for nucleosome reconstitutions. The assembly of the mini-dialysis chambers is shown on the *right*. The *lower part* gives the expected results of analytical MNase digestions dependent on the histone:DNA ratio

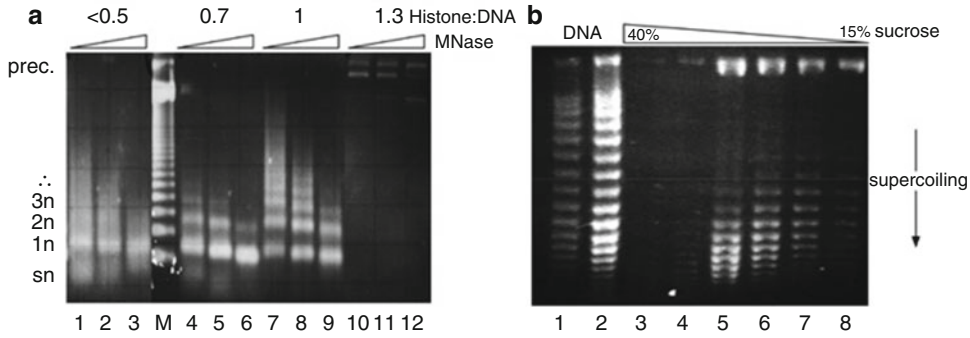

Fig. 3 Analytical MNase digestion of reconstituted chromatin. (**a**). Chromatin assemblies with the indicated histone:DNA ratios were digested with MNase for 30–270 s. Purified DNA was visualized by agarose gel electrophoresis and ethidium bromide staining. The regular fragment ladder indicative of the nucleosomal array is marked (1n–3n). DNA precipitates (prec.) and subnucleosomal DNA (sn) arising at either high or low histone:DNA ratios are indicated. (**b**) Supercoiling assay. Nucleosomes were assembled by salt dialysis and purified in a sucrose gradient. Individual fractions were incubated with topoisomerase I, and the topoisomer distribution of the purified DNA was visualized on agarose gels containing chloroquine. Partially relaxed supercoiled DNA (*lanes 1* and *2*) and fractions with decreasing nucleosome density (*lanes 3–8*) are shown. *Lane 5* shows the nucleosomal fraction that should be used for transcription experiments

additional multimeres of the nucleosomal DNA can be observed (Fig. 3a, lanes 4–9). Conditions with optimal histone:DNA ratios are characterized by the absence of subnucleosomal DNA and a crisp appearance of the nucleosomal ladder (Fig. 3a, lanes 7–9). Overtitration of the histone levels results in the appearance of nonspecific precipitates that cannot be hydrolyzed by MNase (Fig. 3a, lanes 10–12). Having set up the optimal histone:DNA ratios a large-scale assembly reaction is performed and subsequently purified with a sucrose density gradient to enrich the chromatin templates with high nucleosome occupancy. Finally, the chromatin templates can be stored in siliconized tubes for up to 1 year at 4 °C, without significant changes in quality.

2 Materials

2.1 Chromatin Assembly by Salt Dialysis

1. Template DNA. The rDNA minigene of interest in the context of a plasmid backbone. Prepare DNA by using a commercial plasmid preparation kit according to manufacturer's directions. Store DNA preparation in TE buffer (10 mM Tris–HCl pH 8.0, 1 mM EDTA).
2. Histones. Purified histones—from chicken erythrocyte blood or *Drosophila* embryos [21]—or recombinant, refolded histone octameres, expressed in bacteria [22]. Histones are typically stored in 1 M NaCl, 50 % v/v glycerol, 5 mM DTT, at –20 °C.

3. BSA 10 mg/mL in ddH_2O.
4. High salt buffer: 10 mM Tris–HCl pH 7.6, 2 M NaCl, 1 mM EDTA, 0.05% w/v Igepal CA630.
5. Low salt buffer: 10 mM Tris–HCl pH 7.6, 50 mM NaCl, 1 mM EDTA, 0.05% w/v Igepal CA630 (prepare as 20× stock; 1 L).
6. 5 M NaCl.
7. β-mercaptoethanol.
8. Bovine serum albumin (BSA) 20 mg/mL in water.
9. Siliconized 1.5 mL tubes.
10. Dialysis tubing (MWCO 3.5 kDa).
11. Floater for 1.5 mL tubes.
12. Drawn out, U-shaped Pasteur pipets. The Pasteur pipets are first drawn out, after heating them with a Bunsen burner. Subsequently, heated pipets are bent twice to allow the removal of air bubbles below the mini-dialysis chamber floater.
13. Two 3 L plastic beakers.
14. Magnetic stirrer and large (e.g., 4 cm) stir bar.
15. Peristaltic pump.

2.2 Chromatin Digestion with MNase

1. MNase (20 U/μL): Resuspend lyophilized MNase in Ex50 buffer (10 mM Tris–HCl pH 7.6, 50 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM EGTA, 10% v/v glycerol, 1 mM DTT, 0.2 mM PMSF).
2. Ex-80 buffer: 10 mM Tris–HCl pH 7.6, 80 mM KCl, 10% v/v glycerol, 1.5 mM $MgCl_2$, 1 mM DTT.
3. 1 M $CaCl_2$.
4. Stop buffer: 100 mM EDTA, 4% w/v sodium dodecyl sulfate.
5. 10 mg/mL proteinase K in ddH_2O.
6. 10 mg/mL glycogen in ddH_2O.
7. 10 mg/mL BSA in ddH_2O.
8. 100% ethanol (−20 °C).
9. 70% ethanol (−20 °C).
10. 7.5 M ammonium acetate.
11. TBE- or TAE-buffered agarose gel (1.3% w/v).
12. Orange G loading dye: 50% w/v glycerol, 10 mM Tris–HCl, pH 8.0, 0.25% w/v orange G.
13. Siliconized reaction tubes (1.5 mL).
14. Benchtop centrifuge.

2.3 Sucrose Density Centrifugation

1. Ex-80 buffer containing 15 and 30% sucrose (w/v).
2. Chloroquine 100 mM in ddH_2O.

3. Tris–glycine running buffer: 50 mM Tris, 380 mM glycine, 3.3 μM chloroquine.
4. 1.2% w/v agarose gel in Tris–glycine running buffer.
5. Ultracentrifuge with SW41 rotor or similar.
6. Gradient Master (Biocomp) or similar.
7. 10 mg/mL BSA in ddH_2O.
8. Siliconized reaction tubes (1.5 mL).

3 Methods

3.1 Nucleosome Reconstitution by Salt Gradient Dialysis

3.1.1 Preparing the Mini-Dialysis Chambers

1. The mini-dialysis chamber uses a siliconized 1.5 mL tube (*see* **Note 1**), having a hole in the lid and the bottom part removed (Fig. 2). The dialysis membrane is squeezed between the lid and the tube, with the lid now forming the bottom part of the reaction chamber. To prepare the mini-dialysis chamber, close the lid and pinch a hole into the lid with sharp scissors. Scrape out the plastic, only leaving the edges of the lid. Remove the lid from the 1.5 mL tube.
2. Use a "dog nail cutter" to cut off the end of a siliconized 1.5 mL tube.
3. Cut off about 2 cm squares of dialysis tubing and place in a small beaker filled with 1× Low salt buffer for about 10 min. Use just one layer of dialysis membrane per reaction chamber.
4. Place the prepared lid top down onto a saran wrap and place the dialysis membrane on top of the lid. Press the cut siliconized tube with its top over the dialysis membrane locking the membrane between the lid and the tube (*see* **Note 2**). Remove excess of dialysis membrane, fix the mini-dialysis chamber in a floater, and place it into a beaker with High salt buffer.

3.1.2 Salt Gradient Dialysis Reconstitution of Chromatin

1. The volume in the mini-dialysis chamber should not be below 40 μL and the DNA concentration should be above 100 ng/μL to obtain reproducible assembly reactions. Samples should be prepared in siliconized tubes, with the plasmid DNA, 200 ng/μL BSA, variable amounts of histones and filling up with High salt buffer to the final volume (*see* **Note 3**). Mix gently.
2. Pour 300 mL of High salt buffer containing 1 mM β-mercaptoethanol into the 3 L beaker. Position the beaker on a magnetic stirrer and put the floater with mini-dialysis chambers in the beaker.
3. Remove air bubbles below the mini-dialysis chamber with the bent Pasteur pipet.

4. Pipet samples into the floating dialysis chambers and incubate in High salt buffer for 30 min before starting the peristaltic pump.
5. Adjust the speed of the magnetic stirrer underneath to allow slow mixing without creating air bubbles.
6. Set up the peristaltic pump, and adjust the velocity to about 200 mL/h. Prepare 3 L of Low salt buffer containing 1 mM β-mercaptoethanol that will be pumped into the High salt buffer containing beaker over night. Fix the tubing of the peristaltic pump with tape.
7. At the end of buffer dilution, transfer the floater to a beaker with 300 mL of Low buffer containing 1 mM β-mercaptoethanol. Remove air bubbles with the Pasteur pipet and dialyze for 1 h.
8. Transfer the samples into siliconized tubes and determine the volume with the pipet. The volume generally increases up to 20%. Dialyzed chromatin samples are stored at 4 °C and should not be frozen.

3.2 MNase Digestion and Analysis of the Reconstituted Chromatin

1. The MNase digestion requires about 1 μg of reconstituted chromatin. For each assembly reaction, prepare one siliconized tube with 1 μg of chromatin, 1 μL of BSA, and Ex-80 buffer to a final volume of 50 μL.
2. For each chromatin assembly reaction, prepare three 1.5 mL reaction tubes containing 4 μL of Stop buffer.
3. To start the MNase digestion add 50 μL of MNase Mix (Ex-80, 6 mM $CaCl_2$, 200 ng/μL BSA, 20 U MNase), mix gently and remove 30 μL of the reaction after 30, 90, and 270 s. Transfer the reaction to the reaction tubes with the Stop buffer. Change the pipet tips after each transfer (*see* **Note 4**).
4. DNA is deproteinized by the addition of 1 μL of Proteinase K and incubation at 50 °C for 1 h.
5. For DNA precipitation, 1 μL of glycogen and 27 μL of 7.5 M ammonium acetate are added. The sample is thoroughly mixed, 160 μL of cold ethanol are added, mixed, and incubated for 10 min on ice.
6. DNA is precipitated with a benchtop centrifuge, washed with 70% ethanol, and dissolved in 10 μL of TE.
7. The purified DNA samples are electrophoresed in 1.3% agarose gels with Orange G as loading dye and documented.

3.3 Sucrose Gradient Purification of Chromatin and Supercoiling Analysis

1. A linear sucrose gradient is formed in SW41 tubes using the Gradient Master (Biocomp), according to the manufacturer's description.
2. A large-scale chromatin reconstitution (more than 20 μg are recommended) at optimal histone:DNA ratio is performed for further purification by sucrose density centrifugation.

3. The chromatin sample is layered on top of the gradient and overlayed with mineral oil.
4. The samples are placed in the SW41 rotor and centrifuged over night with 270,000 × *g* for 14 h, without brake.
5. Prepare about 20 siliconized tubes to collect the fractionated samples and add 5 μL of BSA in each tube.
6. Samples are collected by pinching a hole through the bottom of the tube and collecting 500 μL fractions in siliconized tubes.
7. Pipet 20 μL of each fraction into a new siliconized tube, add 20 μL of Ex-80 buffer and 10 U of Topoisomerase I. Incubate for 1 h at 30 °C.
8. Ad 4 μL Stop buffer and 1 μL of proteinase K and incubate for 1 h at 50 °C.
9. Precipitate the DNA and dissolve it in 10 μL of TE.
10. Load samples on 1.2 % Tris–glycine agarose gels with 3.3 μM chloroquine. The running buffer must also contain 3.3 μM chloroquine.
11. Run the gel for 24 h at 5 V/cm.
12. Perform ethidium bromide staining and document the 1D supercoiling analysis.
13. Dialyze the fractions containing the highest nucleosome occupancy and determine the DNA concentrations.
14. Chromatin fractions should be stored at 4 °C.

4 Notes

1. Siliconized tubes have to be used to minimize sticking of histones to tube walls.
2. It is important that the membrane is not crumbled, as this may result in leaky chambers. Make sure that the membranes do not dry out.
3. It is recommended to perform an initial wide-range titration of histone to DNA ratios with, for example, ratios of 0.6, 0.8, 1, and 1.2, using 4 μg of DNA. Then to do a large-scale assembly reaction with the optimal histone:DNA ratio (not exceeding a DNA concentration of 1 mg/mL). Ideally a mass ratio of 1:1 should give the best results, but stickiness of histones and difficulties in the quantification of histone concentrations may result in optimal assembly conditions that differ from the ratio of 1. The analytical digestion with MNase is required to determine the optimal histone:DNA ratio.
4. The analytical MNase digestion is best to determine the quality of the reconstituted chromatin. It is important to titrate the

MNase and to adjust the digestion times to visualize the full kinetics of DNA hydrolyzation. Short time points reveal how many nucleosomes were aligned next to each other on average. The higher the number of nucleosomal multimers that can be identified the better is the assembly. On the other hand, a smeary appearance of the hydrolysis products indicates an overtitration of the histones, which will result in nondigestable DNA at one point. Low histone:DNA ratios result in the appearance of subnucleosomal sized DNA fragments corresponding to free DNA on the reconstituted templates. In addition, the digestion reveals only the mono-nucleosomal DNA bands but no or little multimers (Fig. 3).

Acknowledgements

Work in our laboratory is funded by the German Research Community (DFG, grants within the SFB960)

References

1. Knezetic JA, Luse DS (1986) The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell 45:95–104
2. Lorch Y, LaPointe JW, Kornberg RD (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell 49:203–210
3. Almouzni G, Méchali M, Wolffe AP (1990) Competition between transcription complex assembly and chromatin assembly on replicating DNA. EMBO J 9:573–582
4. Längst G, Blank TA, Becker PB, Grummt I (1997) RNA polymerase I transcription on nucleosomal templates: the transcription termination factor TTF-I induces chromatin remodeling and relieves transcriptional repression. EMBO J 16:760–768. doi:10.1093/emboj/16.4.760
5. Längst G, Becker PB, Grummt I (1998) TTF-I determines the chromatin architecture of the active rDNA promoter. EMBO J 17:3135–3145. doi:10.1093/emboj/17.11.3135
6. Diermeier S, Grummt I, Németh A et al (2013) Chromatin-specific regulation of mammalian rDNA transcription by clustered TTF-I binding sites. PLoS Genet 9:e1003786. doi:10.1371/journal.pgen.1003786
7. Grummt I, Längst G (2013) Epigenetic control of RNA polymerase I transcription in mammalian cells. Biochim Biophys Acta 1829:393–404. doi:10.1016/j.bbagrm.2012.10.004
8. Oudet P, Gross-Bellard M, Chambon P (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4:281–300
9. Axel R, Melchior W, Sollner-Webb B, Felsenfeld G (1974) Specific sites of interaction between histones and DNA in chromatin. Proc Natl Acad Sci U S A 71:4101–4105
10. Patterton HG, von Holt C (1993) Negative supercoiling and nucleosome cores. I. The effect of negative supercoiling on the efficiency of nucleosome core formation in vitro. J Mol Biol 229:623–636. doi:10.1006/jmbi.1993.1068
11. Kaplan N, Moore IK, Moore IK et al (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458:362–366. doi:10.1038/nature07667
12. Simpson RT, Stafford DW (1983) Structural features of a phased nucleosome core particle. Proc Natl Acad Sci U S A 80:51–55
13. Neubauer B, Linxweiler W, Hörz W (1986) DNA engineering shows that nucleosome phasing on the African green monkey alpha-satellite is the result of multiple additive histone-DNA interactions. J Mol Biol 190:639–645
14. Felle M, Exler J, Merkl R et al (2010) DNA sequence encoded repression of rRNA gene transcription in chromatin. Nucleic Acids Res 38:5304–5314. doi:10.1093/nar/gkq263
15. Schroth GP, Siino JS, Cooney CA et al (1992) Intrinsically bent DNA flanks both sides of an RNA polymerase I transcription start site. Both

regions display novel electrophoretic mobility. J Biol Chem 267:9958–9964

16. Marilley M, Pasero P (1996) Common DNA structural features exhibited by eukaryotic ribosomal gene promoters. Nucleic Acids Res 24:2204–2211
17. Becker PB, Wu C (1992) Cell-free system for assembly of transcriptionally repressed chromatin from Drosophila embryos. Mol Cell Biol 12:2241–2249
18. Lusser A, Urwin DL, Kadonaga JT (2005) Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. Nat Struct Mol Biol 12:160–166. doi:10.1038/nsmb884
19. Pazin MJ, Bhargava P, Geiduschek EP, Kadonaga JT (1997) Nucleosome mobility and the maintenance of nucleosome positioning. Science 276:809–812
20. Strohner R, Németh A, Nightingale KP et al (2004) Recruitment of the nucleolar remodeling complex NoRC establishes ribosomal DNA silencing in chromatin. Mol Cell Biol 24:1791–1798. doi:10.1128/MCB.24.4.1791-1798.2004
21. Simon RH, Felsenfeld G (1979) A new procedure for purifying histone pairs H2A + H2B and H3 + H4 from chromatin using hydroxylapatite. Nucleic Acids Res 6:689–696
22. Luger K, Rechsteiner TJ, Richmond TJ (1999) Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol 119:1–16. doi:10.1385/1-59259-681-9:1

Chapter 10

A Nonradioactive Assay to Measure Production and Processing of Ribosomal RNA by 4sU-Tagging

Kaspar Burger and Dirk Eick

Abstract

In vivo metabolic pulse labeling is a classical approach to assess production and processing of ribosomal RNA (rRNA). However, conventional labeling techniques can be indirect and require work with radioactivity. Here, we describe in detail a protocol for in vivo metabolic labeling, purification, and readout of nascent rRNA by 4-thiouridine (4sU). We propose 4sU labeling as standard nonradioactive technique for the analysis of rRNA metabolism during ribosome biogenesis.

Key words 4-thiouridine, Nascent RNA, Metabolic pulse labeling, rRNA synthesis, Ribosome biogenesis, Click chemistry

1 Introduction

Measurement of RNA levels is central in gene expression analysis. With various assays available, choice of method depends on the species and abundance of RNA. In some cases, conventional assays for RNA quantitation, such as northern blotting, real-time quantitative PCR (RT-qPCR), or array-based approaches may reflect production, processing, and turnover in RNA metabolism insufficiently. Poor discrimination of nascent RNA from steady-state levels may be misleading. In vivo metabolic pulse labeling is an elegant approach to overcome this obstacle. Classical approaches to label nascent RNA with radioactive tracers such as uracil (^{3}H or ^{14}C), methyl-methionine (^{3}H), or phosphoric acid (^{32}P) allow autoradiography analysis of highly abundant transcripts such as ribosomal RNAs (rRNAs) or transfer RNAs (tRNAs) [1, 2]. More recently, nonradioactive uridine analogs such as 5-bromouridine (5-BrU) or 4-thiouridine (4sU) were described to label and quantify less abundant transcripts including nascent premessenger RNA (pre-mRNA). However, issues with cell permeability may restrict 5-BrU labeling to limited cell types and tissues. 4sU is a thiol group containing photo-reactive nucleoside with high cell permeability that

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_10,

can be introduced into eukaryotic cells and allow nondisruptive metabolic labeling of nascent RNA within minutes [3]. In brief, 4sU is incorporated in nascent RNA during transcription and can be separated from unlabeled, steady-state RNA using click chemistry. After isolation of total RNA, 4sU-labeled RNA is specifically biotinylated in vitro (4sU-tagging) and purified from nonlabeled RNA fraction on streptavidin-coated beads [4]. 4sU-tagged RNA can be sequenced and analyzed genome-wide at nucleotide resolution and thus provides a powerful alternative to assays analyzing genome-wide transcription such as NET-seq or GRO-seq [5–8]. 4sU-labeling is also required for PAR/HITS-CLIP to characterize RNA–protein interactions genome wide [9, 10].

4sU is not a biologically inert nucleoside. Success and relevance of 4sU-tagging strongly depend on usage of subtoxic 4sU concentrations and reasonably short incubation times. For example, excessive incorporation of 4sU into pre-rRNA strongly impairs production and processing on rRNA in human cells [11]. However, optimal 4sU labeling conditions vary depending on RNA species and cell types [5, 12]. Our previous work showed that 4sU-tagging can be used to measure production of the 47/45S rRNA precursor and stepwise processing to intermediate 32S rRNA and mature 28S/18S rRNAs in human cells [13]. Based on previous reports [14, 15], we here introduce an in-depth protocol for successful labeling, tagging, purification of nascent RNA, and its readout exemplified by production and processing of rRNA.

2 Materials

1. Cell culture reagents and plastic ware.
2. 4sU stock solution: 50 mM 4sU in sterile, RNAse-free H_2O (or 1× PBS). Store in small aliquots at −20 °C. *See* **Notes 1** and **2**.
3. Uridine stock solution: 50 mM uridine in sterile, RNAse-free H_2O (or 1× PBS). Store in small aliquots at −20 °C.
4. [^{32}P]-ortho-phosphate (optional).
5. 1× PBS: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4. Adjust to pH 7.2, autoclave, and store at room temperature (RT).
6. RNase-free H_2O: add 0.1% (v/v) DEPC to H_2O. Incubate at 37 °C overnight, autoclave, and store at RT.
7. RNA extraction kit (e.g., peqGOLD Total RNA Kit (PeqLab)). Optional: reagents for enhanced phase separation (e.g., peqGOLD PhaseTrap reagent (PeqLab)).
8. Absolute ethanol (96–100%).
9. 10× biotinylation buffer: 100 mM Tris/HCl pH 7.4, 10 mM EDTA. Store at 4 °C.

10. 1× TE buffer: 10 mM Tris/HCl pH 8.0, 1 mM EDTA.
11. 1× μMACS washing buffer: 100 mM Tris/HCl pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween. Store at 4 °C.
12. μMACS Streptavidin kit (Miltenyi).
13. Magnetic stand μMACS Separator (Miltenyi).
14. 10× MOPS buffer: 20 mM MOPS, 50 mM NaAc, 1 mM EDTA. Adjust to pH 7.0. Buffer is light sensitive, store in the dark at RT.
15. 37% formaldehyde.
16. 1% denaturing agarose-gel: 4 g agarose, 300 mL RNAse-free H_2O, 40 mL 10× MOPS buffer, 67 mL formaldehyde (37%).
17. 2× RNA loading dye (10 mL): 5 mL formamide (50%), 1.5 mL formaldehyde (15%), 1 mL 10× MOPS buffer (1×), 1 mL bromophenolblue (1%, dissolved in RNAse-free H_2O), 10 μL EtBr (10 mg/mL stock, dissolved in RNase-free H_2O), 1.5 mL RNase-free-H_2O. Buffer is light sensitive, store at −20 °C.
18. Spectrophotometer (e.g., Nanodrop 1000), agarose-gel electrophoresis equipment, ultra violet (UV)-light-transilluminator and PhosphorImager (e.g., Fuji FLA-5100) devices and imaging software (e.g., AIDA Version 3.20.116 (Raytest), ImageJ 1.34s (NIH)) to visualize and quantify RNA.

3 Methods

The protocol is optimized to measure production and processing of rRNA in adherent human cell lines (Fig. 1a). We describe in detail individual steps for labeling, tagging, purification, and readout of nascent RNA (Fig. 1b).

3.1 Metabolic Pulse (Double) Labeling of Nascent RNA with 4sU (and ^{32}P)

1. Seed U2OS cells on a 10 cm petri dish 24 h prior to labeling and culture to 80% confluence. Use one plate per time point and include a negative control without 4sU. We suggest the following time points for analysis of rRNA synthesis: 0, 15, 30, 45, 60, 120, and 180′ (minutes) (*see* **Notes 3** and **4**).
2. (*optional*) Add [^{32}P]-ortho-phosphate (^{32}P) directly to cell culture medium (15 μCi/mL) for 1 h prior to 4sU labeling to spike rRNA with a radioactive tracer (*see* **Notes 5** and **6**).
3. Just before labeling, defrost an aliquot of 4sU and dilute 4sU stock solution (50 mM) with fresh, prewarmed culture medium to 10 μM final concentration (2 μL/10 mL medium) and mix (*see* **Note 2**). Include a sample treated with 10 μM uridine as negative control.
4. Aspirate cell culture medium and replace with an identical volume of 4sU-containing medium. Incubate cells at 37 °C for various time points.

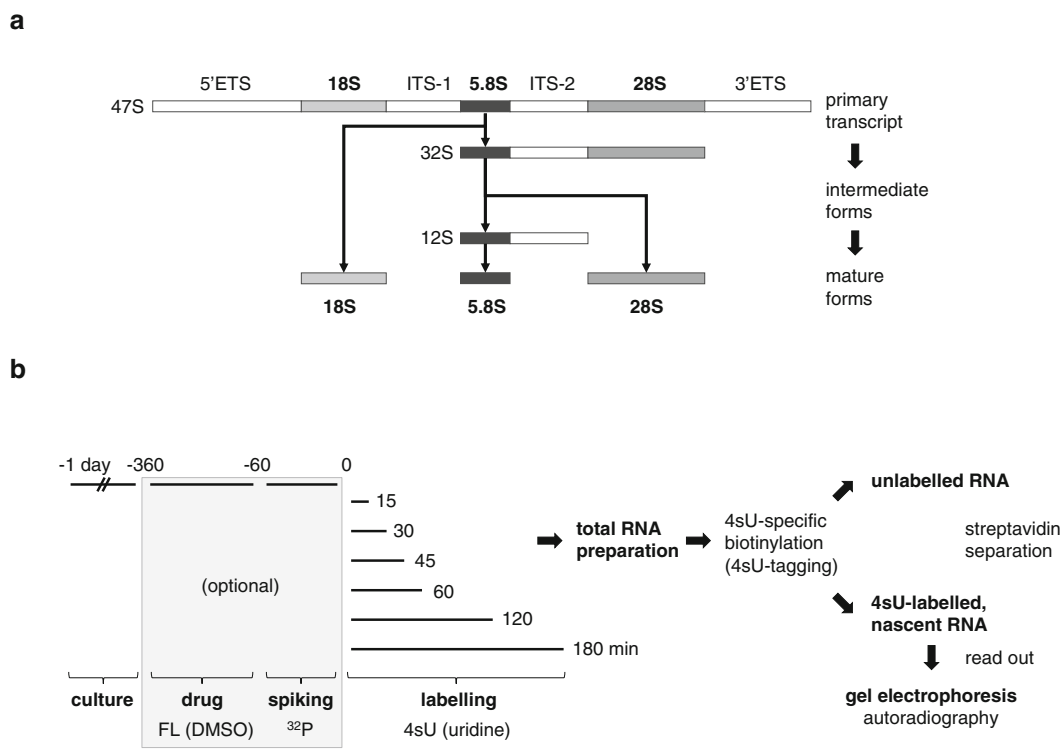

Fig. 1 Analysis of rRNA synthesis by 4-thiouridine (4sU) labeling. (**a**) Transcription and processing of rRNA. A polycistronic transcription unit encodes a 47S precursor rRNA (primary transcript), containing 5′ and 3′ external transcribed spaces (5′ETS, 3′ETS); internal transcribed spacers (ITS-1, ITS-2); and 18S, 5.8S, and 28S rRNAs. A cascade of endo- and exonucleolytic cleavage and processing steps generates mature rRNAs from intermediates. (**b**) In vivo metabolic double labeling workflow for analysis of production and processing of nascent rRNA. Prior to 4sU labeling, cells can be transfected with short-interfering RNAs or treated with small molecule inhibitors such as flavopiridol (FL). Add radioactive tracer such as [^{32}P]-orthophosphate (^{32}P) 1 h prior to 4sU to spike nascent rRNA. Incubate cells with dimethyl sulfoxide (DMSO) and uridine for controls. After labeling, total RNA is prepared, biotinylated in vitro (4sU-tagging), and separated into unlabeled and labeled RNA fractions on streptavidin-coated beads. rRNA forms from both fractions can be analyzed by agarose-gel electrophoresis and autoradiography

5. One minute before reaching each time point, completely remove labeling medium and immediately add cell lysis reagent provided in RNA extraction kits. Use the appropriate amount of lysis buffer per 10 cm culture dish and incubate for 5′ at room temperature to allow complete lysis (*see* **Notes 7** and **8**). Aim for help from a colleague, if you process several replicates in parallel.
6. Clear the crude lysate from the bulk of DNA and debris e.g., using a DNA removing column (*See* **Note 9**).

Pause point Lysate can be stored at −20 °C. (If RNA is spiked with a ^{32}P tracer, sample starts to undergo autoradiolysis after several days).

3.2 Preparation of Total RNA

1. Defrost frozen lysate and perform the next steps at room temperature. Add an identical volume of 70% ethanol and mix by pipetting up and down until precipitates disappear. Load 750 µL of sample to an RNA binding column sitting on a 2 mL collection tube and spin (1′, 12,000 rcf). Discard flow through. Repeat until complete sample volume is loaded on column (*see* **Note 10**).
2. Add 500 µL high salt wash buffer from RNA extraction kit (e.g., RNA wash buffer I from PeqGOLD total RNA kit) to an RNA binding column, spin (1′, 12,000 rcf) and discard flow through. Repeat with 600 µL low salt wash buffer (e.g., RNA wash buffer II from PeqGOLD total RNA kit) (*see* **Note 11**).
3. Put RNA column into new 2 mL collection tube and spin empty (3′, 12,000 rcf) to remove residual low salt wash buffer completely from column (*see* **Note 12**).
4. Put RNA column into clean 1.5 mL reaction tube and pipette 75 µL prewarmed, RNase-free H_2O (55 °C) directly and dropwise to column. Incubate for 10′ and spin (3′, 12,000 rcf). The flow through contains eluted total RNA (*see* **Note 13**).
5. Use a spectrophotometer to measure concentration and purity of total RNA. Concentration should range from 1000 to 2000 ng/µL. Thirty micrograms of total RNA is sufficient for 4sU-tagging of rRNA.

Pause point: RNA sample can be stored at –20 °C (*see* **Note 14**).

3.3 Biotinylation of 4sU-Labeled Total RNA (4sU-Tagging)

EZ-link biotin-HPDP is pyridyldithiol activated and allows thiol-specific biotinylation of 4sU-labeled RNA by formation of reversible disulfide bonds with the –SH group of 4sU. Perform this step at room temperature.

1. Use same amount of purified total RNA (30–100 µg) in a final volume of 600 µL for all samples. Example: biotinylation of 60 µg total RNA (1200 ng/µL). Mix 50 µL total RNA with 60 µL 10× biotinylation buffer (1 µL/µg RNA), 120 EZ-link biotin-HPDP (2 µL/µg RNA), and 370 µL RNAse-free H_2O. Incubate at room temperature for 1.5 h with rotation (*see* **Note 15**).
2. Add an equal volume of chloroform/isoamylalcohol (24:1) (600 µL), mix vigorously and incubate for 3′ until phases begin to separate and bubbles disappear. Spin sample together with empty Phase Trap Gel Heavy Lock tubes (5′, 13,000 rcf). Transfer upper aqueous phase of sample into new 1.5 mL tube and mix with one volume of chloroform/isoamylalcohol (24:1). Mix vigorously and incubate with RNA solution for 3′. Transfer sample into prespun Phase Trap Gel Heavy Lock tubes and spin again (5′, 13,000 rcf). The upper phase contains the RNA sample (*see* **Note 16**).

3. Transfer RNA sample into a new 1.5 mL tube, precipitate RNA by addition of 1/10 of reaction volume 5 M NaCl (60 µL) and an equal volume of absolute isopropanol (600 µL), mix vigorously and spin (20′, 13,000 rcf). Precipitated RNA is visible as pale pellet. Remove supernatant and add 1 mL ethanol. A white pellet should now be clearly visible. Mix vigorously until the pellet is floating and spin again (5′, 8000 rcf). Remove ethanol in supernatant, short-spin and remove remaining ethanol. Dry RNA pellet at 37 °C for 5′ until pellet color becomes transparent. Resuspend RNA pellet in 100 µL RNase-free H_2O and incubate for 10′ at 37 °C with rotation (*see* **Note 17**).

Pause point: RNA sample can be stored at –20 °C.

3.4 Purification of Labeled, 4sU-Tagged from Unlabeled RNA

At this stage, the RNA sample contains both biotinylated, 4sU-tagged RNA (nascent RNA) and nontagged RNA (steady-state). Immunoprecipitation of biotinylated RNA using streptavidin-coated magnetic beads separates RNA fractions to specifically assess production and processing of nascent RNA levels. Do not process more than 12 samples at a time.

1. Place µ-Magnetic-Activated-Cell-Sorting (µMACS, size M) column into µMACS Separator magnetic stand. Warm µMACS washing buffer to 65 °C and load 900 µL to prerun and equilibrate column. Discard flow through (*see* **Note 18**).
2. Incubate 100 µL biotinylated RNA sample for 10′ at 65 °C and place on ice immediately for 5′. Add 75 µL streptavidin-coated magnetic beads (µMACS Streptavidin kit) to sample and incubate for 15′ with rotation (*see* **Notes 19** and **20**).
3. Load sample to µMACS column. The separator binds 4sU-tagged RNA. The flow through contains unlabeled RNA. Discard flow through (*see* **Note 21**).
4. Wash column five times with 800 µL µMACS washing buffer. Discard flow through (*see* **Note 22**).
5. Upon completion of last wash, place a clean 2 mL collection tube containing 700 µL lysis buffer from RNA extraction kit (e.g., lysis buffer T from PeqGOLD total RNA kit) underneath the column. Add 100 µL elution buffer (100 mM DTT) to column and elute for 3′. Repeat elution with 100 µL elution buffer (100 mM DTT). The 2 mL collection tube contains 900 µL eluted nascent RNA (*see* **Note 23**).

   Pause point: RNA sample can be stored at –20 °C.

3.5 Recovery of Nascent (and Steady-State) RNA

1. Add 500 µL absolute ethanol (96–100%) to 900 µL eluted sample and mix by pipetting.
2. Recover labeled, nascent RNA as described in Subheading 3.2. Elute nascent RNA in 35 µL RNase-free H_2O.

3. (optional) Recover unlabeled RNA as described in Subheading 3.3 (**step 3**). Resuspend unlabeled RNA in 35 μL RNase-free H_2O.

   Pause point: RNA sample can be stored at –20 °C.

3.6 Agarose-gel Electrophoresis

Various assays exist to readout 4sU-labeled RNA, depending on RNA species and abundance. To assess time-dependent production and processing of rRNA forms, nascent 4sU-labeled RNA is separated by agarose-gel electrophoresis and stained with ethidium bromide (EtBr) (Fig. 2a).

1. Heat 4.5 g agarose in 300 mL RNAse-free H_2O to boiling. Cool agarose solution to 65 °C and add 40 mL 10× MOPS buffer (RT) and 67 mL formaldehyde solution (37%) (RT) and mix thoroughly before pouring. Use 1× MOPS running buffer (*see* **Note 24**).

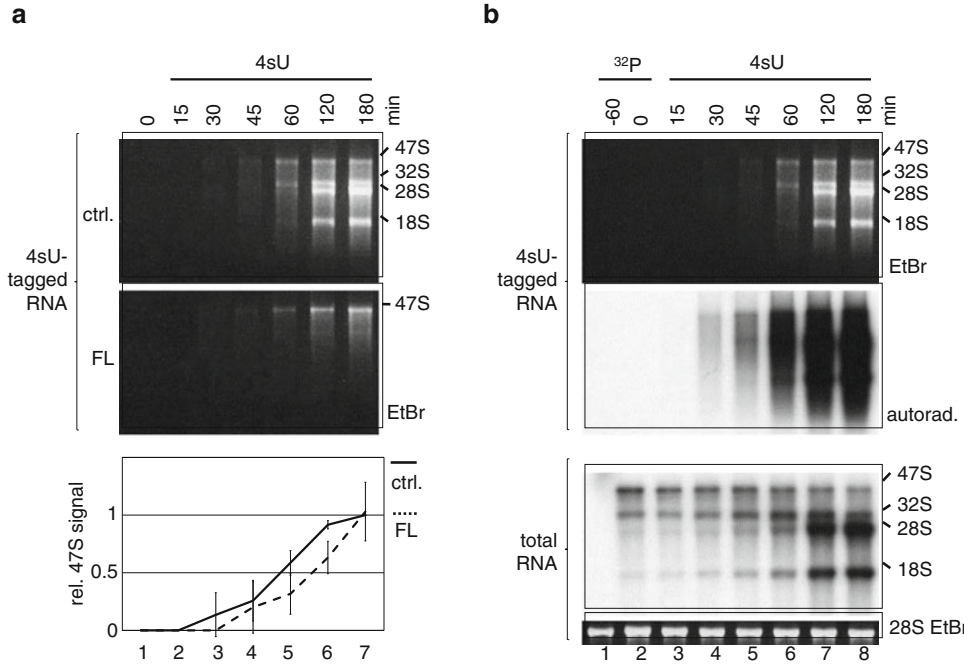

Fig. 2 Analysis of nascent rRNA synthesis by 4sU-tagging. (**a**) Flavopiridol (FL) inhibits production and processing of 47S rRNA. *Upper panel*: nascent, 4sU-labeled RNA (10 μM) is separated by agarose-gel electrophoresis. Time-dependent production and processing is visualized by staining with ethidium bromide (EtBr) and impaired in presence of flavopiridol (800 nM, 6 h). *Lower panel*: time-dependent increase of 47S rRNA. Signals are normalized to time point 180 min. ctrl.: 0.1% dimethyl sulfoxide (DMSO). (**b**) Double labeling of nascent RNA with [^{32}P]-orthophosphate (^{32}P, 15 μCi/mL) and 4sU (10 μM). *Upper panel*: double labeling confirms purification of nascent RNA with signals from EtBr staining and autoradiography increasing concomitantly. *Lower panel*: conventional analysis of rRNA synthesis by autoradiography of double-labeled total RNA. 28S EtBr: loading control. Similar results were obtained using unlabeled RNA upon streptavidin separation of 4sU-tagged RNA (data not shown)

2. Mix purified 4sU-labeled RNA with 2× RNA loading dye (1:1), incubate at 55 °C for 12′, load and run gel until blue running front is separated approximately 5 cm from well (*see* **Note 25**).
3. Take a gel photo under UV light and readout rRNA forms (*see* **Note 26**).

3.7 Autoradiography

Autoradiography confirms successful purification of nascent RNA after double labeling with 4sU and ^{32}P as tracer. Both total and 4sU-labeled nascent RNA can be assessed (Fig. 2b).

1. Separate total or nascent RNA sample by agarose-gel electrophoresis as described in Subheading 3.6.
2. Put the gel in a gel dryer on top of a filter paper and cover with plastic foil. Dry the gel for 5 h at 80 °C using a vacuum pump.
3. Put dried gel into an autoradiography cassette and cover gel with an autoradiography detection film. Expose film at −80 °C, defrost cassette, and visualize signals using a developing machine. Expose film for several time points (*see* **Note 27**).

4 Notes

1. Prepare all reagents with laboratory-grade chemicals, use sterile filtered, RNase-free H_2O, clean reaction tubes, and filtered tips.
2. 4sU can be dissolved in sterile, RNAse-free H_2O (or 1× PBS). To generate a yellow 50 mM stock solution, add 7.68 mL solvent to 100 mg lyophilized 4sU powder and store in small aliquots at −20 °C. 4sU is light sensitive and aliquots should be defrosted only once.
3. 4sU can be used with standard cell culture media such as Dulbecco's modified Eagle's Medium (DMEM) or Roswell Park Memorial Institute Medium (RPMI) 1640, including 10% fetal bovine serum, 1% glutamine, and 1% penicillin/streptomycin.
4. Cells should be monitored for possible contaminations before 4sU labeling. For instance, absence of mycoplasma should be confirmed by PCR-based assays to avoid incorporation of label into and generation of 23S/16S rRNA signals in metabolic pulse labeling by highly active prokaryotic RNA metabolism.
5. Prior to 4sU labeling, cells can be manipulated in various ways, such as transfection of short-interfering RNAs or pretreatment with drugs.
6. Replace standard medium with phosphate-free culture medium (DMEP-P, 10% FBS-P) to increase ^{32}P incorporation rate in rRNA. All steps require work under radiation-protected conditions in case of double labeling with a radioactive tracer.
7. This protocol describes preparation of total RNA using a commercial kit. Use sufficient amounts of lysis buffer T for effective

lysis. Cells can also be lysed using Trizol reagent (Ambion life technologies) or acid phenol/chloroform. If so, total RNA is extracted using standard phenol/chloroform extraction and isopropanol precipitation protocol [14, 16].

8. For lysis of 4sU labeled suspension cells, spin down cell suspension (4′, 1200 rcf), remove supernatant, and add lysis reagent to pellet. Resuspend pellet and incubate for 10′ at room temperature.
9. DNA removing columns can be reloaded several times. However, DNA removing columns tend to clog if lysis is performed with insufficient amounts of lysis buffer or high amount of starting material. If so, remove clogged lysate, add additional lysis buffer, and load to a fresh DNA removing column.
10. RNA columns can be repeatedly loaded several times until maximum binding capacity of approximately 100 μg is reached. This protocol requires five to six loading cycles.
11. Repeat washing with low salt wash buffer to remove salt thoroughly, in case downstream assays require low salt conditions.
12. This step is crucial for quantitative elution of RNA from column, since residual ethanol interferes with eluent.
13. Warm eluent (55 °C) increases RNA yield. A second round of elution can be performed when expecting high RNA yield (>50 μg).
14. Radiolysis can significantly reduce RNA half-life when performing double labeling with radioactive tracer.
15. Biotinylation of highly concentrated RNA may result in reaction volumes smaller than 600 μL. If so, fill up to 600 μL with 1× TE buffer to reduce loss of biotinylated material in subsequent precipitation step. Store 10× biotinylation buffer in 1.2 mL aliquots at 4 °C and EZ-link biotin-HPDP stock solution (1 mg/mL dissolved in dimethylformamide (DMF)) at −20 °C.
16. Two subsequent chloroform extractions are required to completely remove residual biotin-HPDP. The initial volume of RNA is reduced to about 80%, since biotin-HPDP solvent (DMF) is also removed.
17. Complete removal of ethanol and drying of pellet is crucial to quantitatively dissolve RNA in RNase-free H_2O. Do not overdry the RNA pellet to avoid refractory resuspension.
18. Gently press on the top of the column with your thumb to initiate flow through the column.
19. Heating of RNA sample resolves secondary structures and promotes binding of biotinylated RNA to streptavidin-coated beads with high affinity. Do not put in >100 μg RNA to avoid background binding. Prolong incubation time to increase RNA yield.
20. Briefly shake streptavidin-coated beads to equilibrate. Use cut filter tips pipette beads.

21. Unlabeled RNA is not required to measure nascent rRNA production and processing. Collect flow through for comparison of nascent and steady-state RNA levels.
22. Use 65 °C warm μMACS washing buffer for washes 1–3 and room temperature buffer for washes 4 and 5. Constantly reload washing buffer to prevent the μMACS column from running dry.
23. DTT elutes 4sU-labeled nascent RNA by reduction of disulfide bonds between 4sU and biotin-HPDP. Reduced biotin-HPDP remains bound to the μMACS column by streptavidin-coated magnetic beads. Use freshly prepared 65 °C warm DTT to increase RNA yield.
24. Formaldehyde promotes denaturation of RNA secondary structures. Pour gel under fume hood, cover with plastic foil, and allow polymerization at 4 °C for 1 h. MOPS buffer is light sensitive, do not recycle MOPS buffer. Cover gel with aluminum foil, if exposed to light. Do not add EtBr to the gel, as EtBr cross-stains formaldehyde under UV light. EtBr is included in 2× RNA loading dye.
25. Incubation at 55 °C promotes denaturation of RNA secondary structures. When processing several samples, adjust different RNA sample volumes with RNAse-free H_2O to same volume. Use identical incubation times and sample volumes to ensure equal separation.
26. Reduce running time, in case rRNA forms are blurry and increase running time, if rRNA forms are not separated sufficiently. Signals can be quantified using ImageJ software.
27. Exposure time can vary depending on detection system, activity of tracer, incorporation rate, quality and quantity of RNA, and exposure temperature. Use reflector screens to enhance signal and decrease exposure time. Signals can be quantified using a Phosphorimager (Fuji) and AIDA software.

Acknowledgements

We thank Lars Doelken, Bernd Raedle, and Lisa Marcinowski for help with experimental design and technical support. This work was supported by the Deutsche Forschungsgemeinschaft (SFB684 and SFB1064).

References

1. Pestov DG, Lapik YR, Lau LF (2008) Assays for ribosomal RNA processing and ribosome assembly. Curr Protoc Cell Biol Chapter 22:Unit 22 11
2. Warner JR, Soeiro R (1967) Nascent ribosomes from HeLa cells. Proc Natl Acad Sci U S A 58:1984–1990

3. Melvin WT, Milne HB, Slater AA et al (1978) Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. Eur J Biochem 92:373–379
4. Cleary MD, Meiering CD, Jan E et al (2005) Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat Biotechnol 23:232–237
5. Dolken L, Ruzsics Z, Radle B et al (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA 14: 1959–1972
6. Rabani M, Levin JZ, Fan L et al (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat Biotechnol 29:436–442
7. Schwanhausser B, Busse D, Li N, Dittmar G et al (2011) Global quantification of mammalian gene expression control. Nature 473:337–342
8. Windhager L, Bonfert T, Burger K et al (2012) Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome Res 22:2031–2042
9. Hafner M, Landthaler M, Burger L et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
10. Licatalosi DD, Mele A, Fak JJ et al (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469
11. Burger K, Muhl B, Kellner M et al (2013) 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol 10:1623–1630
12. Kenzelmann M, Maertens S, Hergenhahn M et al (2007) Microarray analysis of newly synthesized RNA in cells and animals. Proc Natl Acad Sci U S A 104:6164–6169
13. Burger K, Muhl B, Rohrmoser M et al (2013) Cyclin-dependent kinase 9 links RNA polymerase II transcription to processing of ribosomal RNA. J Biol Chem 288:21173–21183
14. Dolken L (2013) High resolution gene expression profiling of RNA synthesis, processing, and decay by metabolic labeling of newly transcribed RNA using 4-thiouridine. Methods Mol Biol 1064:91–100
15. Radle B, Rutkowski AJ, Ruzsics Z et al (2013) Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. J Vis Exp (78), doi: 10.3791/50195
16. Chomczynski P, Sacchi N (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nat Protoc 1:581–585

Chapter 11

Metabolic Labeling in the Study of Mammalian Ribosomal RNA Synthesis

Victor Y. Stefanovsky and Tom Moss

Abstract

RNA metabolic labeling is a method of choice in the study of dynamic changes in the rate of gene transcription and RNA processing. It is particularly applicable to transcription of the ribosomal RNA genes and their processing products due to the very high levels of ribosomal RNA synthesis. Metabolic labeling can detect changes in ribosomal RNA transcription that occur within a few minutes as opposed to the still widely used RT-PCR or Northern blot procedures that measure RNA pool sizes and at best are able to detect changes occurring over several hours or several days. Here, we describe a metabolic labeling technique applicable to the measurement of ribosomal RNA synthesis and processing rates, as well as to the determination of RNA Polymerase I transcription elongation rates.

Key words Ribosomal RNA (rRNA), Precursor rRNA, rRNA synthesis rate, rRNA processing rate, RNA polymerase I (RPI or PolI), Transcription elongation, Metabolic labeling, Electrophoresis, Vacuum transfer of RNA, Fluorography, Scintillation counting, Phospho-imaging

1 Introduction

The four ribosomal RNAs (18S, 28S, 5.8S, and 5S) form the catalytic core of the ribosome, and as such, they account for 80% or more of the total RNA in mammalian cells. Their synthesis typically occupies some 30% of the total gene transcription in a proliferating cell and is regulated in a complex manner that is linked to cell growth and proliferation [1]. For many years, ribosomal RNA (rRNA) synthesis was considered as a slowly regulated housekeeping process, but metabolic RNA labeling revealed that in fact it responds within minutes to growth stimuli [2], and is directly regulated by mitogens and growth factors at the levels of transcription elongation and initiation competence [3, 4]. When applied to mammalian cells, methods such as Northern blot or RT-PCR analysis can generally not detect such rapid changes in synthesis rate due to the great abundance of cellular rRNA and the relatively slow processing rates. Quantitative RT-PCR, S1-protection, or Northern monitoring of the extreme 5′ of the 47S

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_11,

rRNA precursor has been widely used to measure rRNA synthesis, since this region of the rRNA is generally processed relatively rapidly, e.g. [4, 5]. However, the rRNA processing rate can vary widely and, more importantly, independently of 47S synthesis [6], making the interpretation of data from these techniques at best questionable and at worst fully misleading. Further, metabolic labeling can provide a direct measure of rRNA processing rates and pathways not available by other means.

Cellular synthesis of rRNA is often analyzed by culturing mammalian cells with a radiolabeled nucleoside such as [^{3}H] or [^{14}C]-uridine, or with [^{32}P] inorganic phosphate, and all have been extensively used for this purpose [2, 3, 7–9]. Given that the rRNA accounts for such a large fraction of total cellular transcription and this transcription is concentrated in a single precursor RNA and a small number of processing products, the newly synthesized rRNAs can be very simply analyzed by gel electrophoresis, e.g., Fig. 1a. Cells are usually labeled for periods from a few minutes to several hours dependent on the parameters of the experiment, and this will define the choice of radiolabel, [^{32}P] inorganic phosphate or [^{3}H]-uridine providing much greater sensitivity than [^{14}C]-uridine due to isotope half-life and disintegration energy. In the example in Fig. 1, labeling of mouse embryonic fibroblasts (MEFs) was performed for 3 h with [^{3}H]-uridine, and EtBr staining used in parallel to observe the steady-state rRNA pools. In this case, MEFs conditional for the gene for the essential factor UBF (UBTF) [10], which can be deleted using 4-hydroxy-tamoxifen (4HT), were labeled before and after gene deletion. In such a 3 h "pulse" labeling, the 47S precursor rRNA and its intermediate and final processing products are all visualized, and all are shown to be products of the rRNA genes since their labeling depends absolutely on the availability of UBF. So-called pulse-chase experiments can be used to follow the rRNA processing pathways and to determine the rates of processing [6, 9, 11], e.g., Fig. 2a. Since processing involves the 2′-O-methylation of the nascent 47S rRNA, it can also be analyzed using [^{3}H]- or [^{14}C]-methylmethionine as the methyl donor [6, 12–14]. Since, during pulse labeling times much shorter than the rRNA processing rates, the label accumulates almost exclusively in the 47S rRNA precursor, such labeling regimes can be used to follow the absolute level of rRNA synthesis, e.g., Fig. 3a. This approach has also been used to directly measure the time taken to complete transcription of a 47S rRNA precursor and hence to determine the average transcript elongation rate achieved by RNA Polymerase I (RPI or PolI) [3], e.g., Fig. 3b. Cells take up the radiolabel very rapidly such that any delay in incorporation into RNA can usually be neglected. However, the uptake rate can be readily determined using standard procedures, e.g. [3] or by measuring the incorporation rate into abundant short RNAs such as the 5S rRNA, e.g. [15].

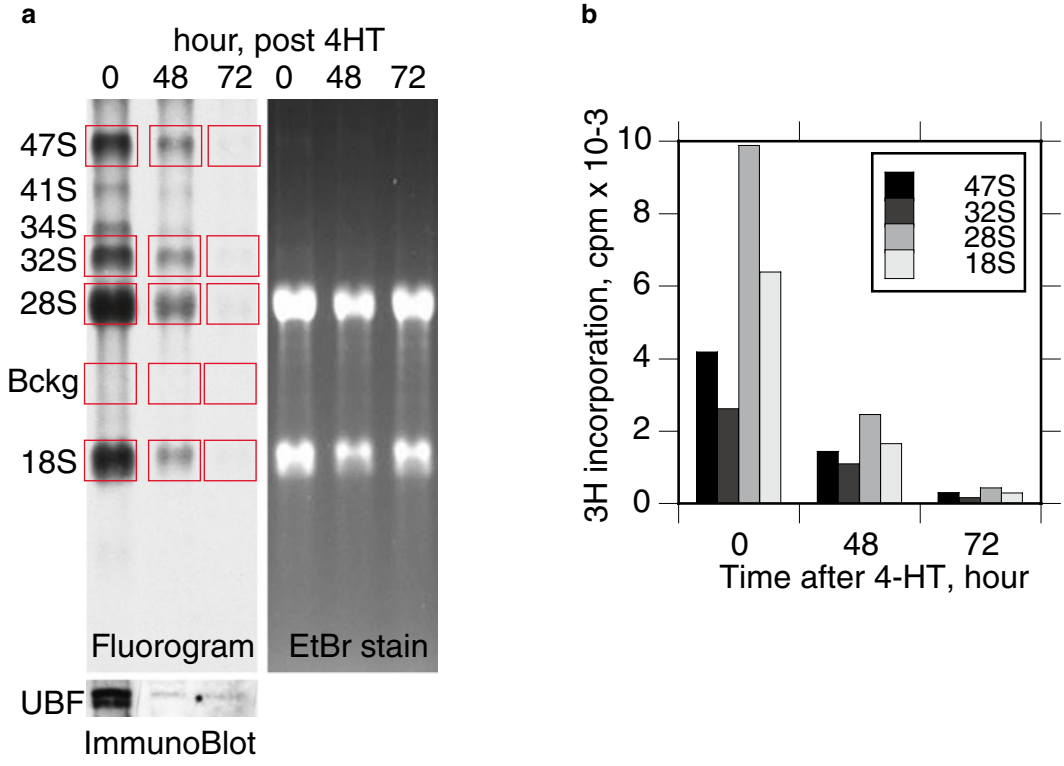

Fig. 1 Example of the pulse labeling of rRNA in mouse embryonic fibroblasts carrying homozygous conditional (Floxed) genes for the RPI factor UBF. (**a**) Electrophoretic analysis of a 3 h [^{3}H]-uridine labeling before and at 48 and 72 h after inducing the deletion of the UBF genes with 4-hydroxy-tamoxifen (4HT). *Left panel* shows the fluorograph and *right panel* the corresponding ethidium bromide stained gel. *Boxes* outlined in *red* indicate the areas cut from the fluorograph for scintillation counting. Immunoblot shows the corresponding levels of UBF. (**b**) Example of the raw counts per min (cpm) quantification data as determined by scintillation counting of the fluorograph in (**a**)

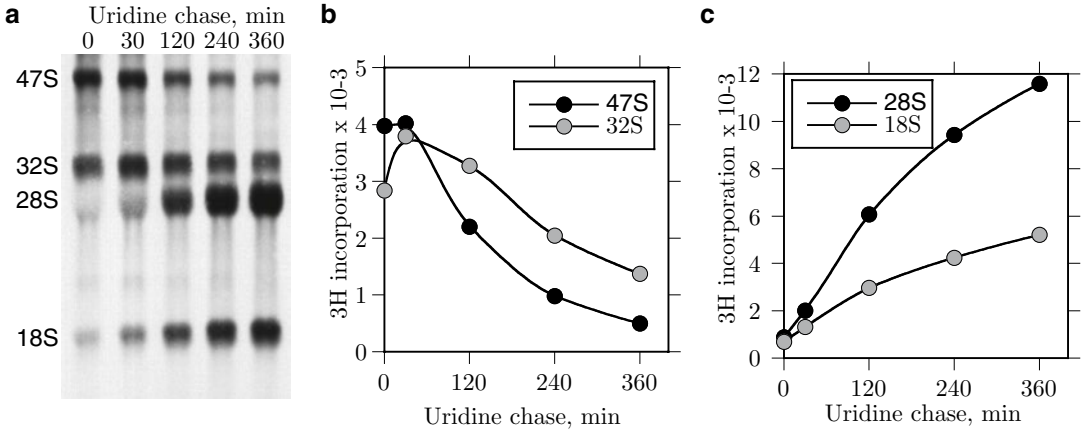

Fig. 2 Example of a pulse-chase experiment. Cells were labeled with [^{3}H]-uridine for 1 h before chasing with unlabeled uridine. (**a**) Fluorograph of RNA samples taken at 0–360 min after start of chase. (**b**) and (**c**) show the raw quantitation data for the 47, 32, 28, and 18S rRNAs. Note the 32S initially accumulates before decaying while the 28 and 18S rRNAs continue to accumulate throughout

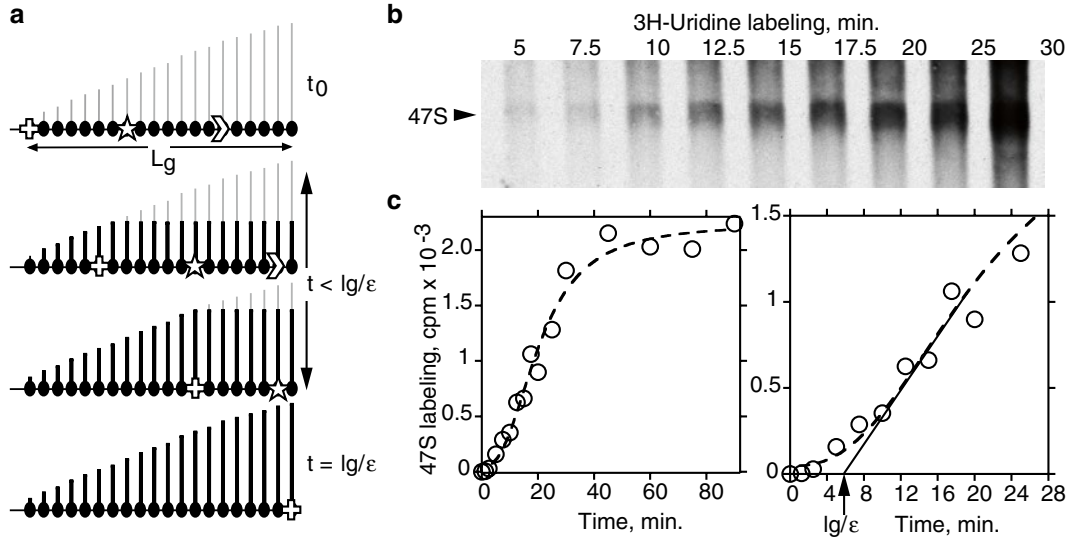

Fig. 3 Example of a series of short labeling pulses to analyze RPI elongation rate. (**a**) Diagrammatic demonstration of label incorporation at times shorter than the time required for RPI to complete a full transcript. Symbols indicate the advance of the transcription complexes. (**b**) A typical fluorograph of 47S rRNA labeling time course. (**c**) *Left*, plot of raw incorporation data from 0 to 90 min of pulse labeling, showing that 47S processing becomes apparent after about 30 min. *Right*, expanded plot of the early time points showing the transition from the nonlinear to linear labeling kinetics, and the extrapolation of the linear domain to determine elongation time. *See* Subheading 3.8 for more details

Here, we describe the procedures to metabolic label and analyze rRNA from mammalian tissues culture, using the NIH3T3 monolayer cells as typical example. The basic procedure uses [^{3}H]-uridine labeling, followed by isolation of total RNA, its fractionation by denaturing formaldehyde agarose gel electrophoresis [16–18], imaging by fluorography, and quantitation by scintillation counting or phosphoimaging. The procedure can be adapted to the use of [^{14}C]-uridine and [^{3}H]- or [^{14}C]-methylmethionine.

2 Materials

Prepare all solutions using single distilled or Milli-Q™ equivalent water and analytical grade reagents. All solutions should be RNase free. Where appropriate, stock solutions and water should be autoclaved for 45 min at 120 °C. Other reagents should be taken fresh from a new bottle and transferred, preferably by pouring not pipetting, into an RNase free, autoclaved, tube. Follow all local waste disposal regulations for radioactivity and toxic compounds.

2.1 Materials for Metabolic Labeling

1. DMEM or other appropriate medium for cell culture. When labeling with [^{3}H]-methylmethionine, use medium without methionine and rinse cells in this medium before beginning the labeling.

2. [5,6-^{3}H|-uridine, [^{14}C|-uridine, or [^{3}H]-methyl-methionine (PerkinElmer, Waltham, Massachusetts, United States).
3. For pulse-chase experiments, prepare a solution of 1 mM uridine or 1 mM methionine (Sigma-Aldrich, St. Louis, Missouri, United States).

2.2 Materials for RNA Extraction

1. Trizol™ Reagent (Life Technologies, Carlsbad, California, United States) or QIAzol Lysis Reagent (QIAGEN, Hilden, Germany). Or alternatively, follow the original guanidinium thiocyanate–phenol–chloroform extraction procedure [19].
2. Chloroform:isoamyl alcohol 24:1 v/v.
3. Isopropanol.
4. 70% v/v Ethanol in water, keep at −20 °C.
5. Formamide (Sigma-Aldrich, St. Louis, Missouri, United States).

2.3 Materials for Formaldehyde Gel Electro-phoresis

1. Horizon 11.14 (11 × 14 cm) horizontal agarose gel electrophoresis chamber (Apogee Designs, Ltd. Baltimore, MD United States) or similar apparatus.
2. Gel comb to form loading slots for up to 40 µL samples, for example, 0.5 cm wide × 0.2 cm thick.
3. RNaseZap® (Ambion Thermo Fisher Scientific) or the cheaper Windex® (SC Johnson, United States).
4. Agarose (e.g., UltraPure™ Agarose, Life Technologies, Carlsbad, California, United States) (*see* **Note 1**).
5. Formaldehyde, 37% w/w (Sigma-Aldrich, St. Louis, Missouri, United States).
6. Ethidium bromide (Sigma-Aldrich, St. Louis, Missouri, United States), 10 mg/mL stock solution.
7. 10× MOPS solution: 0.2 M MOPS (3-*N*-morpholino propanesulfonic acid) (Sigma-Aldrich, St. Louis, Missouri, United States), 0.05 M sodium acetate, 0.01 M Na_2EDTA. Adjust pH to 7.2 with 10 N NaOH at 20 °C.
8. 1× electrophoresis buffer: 1× MOPS.
9. 10× RNA loading buffer without dyes: 50% v/v glycerol, 10 mM EDTA pH 8.0.
10. 1× RNA loading buffer for denaturing formaldehyde gels. Prepare sufficient volume for all samples immediately before use. Each sample requires: 18 µL formamide, 6 µL 37% w/w formaldehyde, 1 µL 1.6 mg/mL ethidium bromide, 3.6 µL 10× MOPS, 4 µL 10× RNA loading buffer without dye (*see* **Note 2**).

2.4 Materials for Fluorescent Imaging of RNA Fractionation

1. 250 or 300 nm UV transilluminator and gel documentation camera suitable for ethidium bromide fluorescence.
2. If ethidium bromide is not included in the RNA loading buffer, a suitable container in which to stain the gel is necessary.

2.5 Materials for the RNA Transfer

1. Biodyne membrane (Pall Corporation, Port Washington, NY, USA).
2. 20× SSC (Standard Saline Citrate): 175 g/l NaCl, 88 g/l sodium citrate.
3. Whatman® 3MM paper.
4. UV cross-linker: Hoefer UVC 500 or similar.

2.6 Materials for Fluorography

1. EN^{3}HANCE™ Spray Surface Autoradiography Enhancer (PerkinElmer, Waltham, Massachusetts, United States). Use only in a chemical hood (*see* **Note 3**).
2. Radioactive ink: Dilute any nontoxic [14C]-radiolabeled molecule, e.g., [14C]-uridine in commercial black writing ink to give 2–3 μCi/mL final concentration (*see* **Note 4**).
3. X-Ray Film Cassette with two sheets of Whatman 3MM paper for protection. The use of EN^{3}HANCE™ Spray will contaminate the cassette, so it is best to dedicate one to this procedure.
4. Biomax XAR film (Carestream Health, Inc., Rochester, NY, United States).
5. Biofreezer for autoradiography exposures, −70 °C or lower.

2.7 Materials for Quantitative Analysis

1. Light-box, a visible light transilluminator.
2. Saran Premium Wrap (SC Johnson, United States) (*see* **Note 5**).
3. Ballpoint pen.
4. Scissors.
5. Glass scintillation vials with caps.
6. Scintillation liquid, e.g., Scintiverse (Thermo Fisher Scientific, Waltham, MA, United States) or similar.
7. Liquid scintillation counter.
8. Computer with ImageJ (National Institutes of Health, United States) and Excel (Microsoft, Redmond, Washington, United States) or similar image analysis and graphing software.

3 Methods

3.1 Pulse Labeling

1. Plating of monolayer cells. Typically, 0.8×10^6 NIH3T3 were plated in 60 mm petris the day before labeling. For other cell lines the cell number may need to be varied according to the cell growth and cell size. Cells should be not more than 60–80% confluent at the time of labeling since most cells will down-regulate ribosomal transcription at higher confluence. At this stage it is important to decide if it is intended to accurately quantify labeling. If this is the case, cells should be plated in triplicate.

2. Prepare a 10× solution of radiolabel in tissue culture medium to give 10 μCi [^{3}H]-uridine per 100 μL medium, keeping in mind the total number of petris to be labeled including triplicates. (*See* **Note 6** for other radiolabels and very short pulse labeling.)
3. Adjust the medium volume in each culture to exactly 2 mL (*see* **Note 7**).
4. Add 100 μL of the 10× radiolabel solution to each culture mixing well but with care.
5. Time of labeling will vary depending on the experiment.

3.2 Pulse-Chase Labeling

1. Follow **steps 1–4** of Subheading 3.1.
2. After labeling for 0.5–1 h, change the medium for fresh medium containing 0.5 mM unlabeled uridine and continue culturing for 1–6 h. (For [^{3}H]-methylmethionine labeling *see* **Note 6**.)

3.3 RNA Extraction

1. Remove culture medium containing the radiolabel. (Store and dispose of contaminated solutions and plastics according to local regulations.)
2. Lyse the cells directly with 1 mL of Trizol™ or equivalent. Homogenize and reduce viscosity by pipetting "up and down." Transfer into 2 mL microcentrifuge tube.
3. Add 200 μL chloroform, close tube, and mix thoroughly by manual shaking, not vortexing, for about 15 s.
4. Centrifuge at 10,000 × *g* for 15 min at 4 °C.
5. Remove 600 μL of the upper aqueous phase, avoiding the interphase, and transfer into 1.5 mL microcentrifuge tube.
6. Add 600 μL of isopropanol and mix well by inversion. Leave 15 min at room temperature.
7. Centrifuge at 10,000 × *g* for 10 min at 4 °C.
8. Remove supernatant as completely as possible (*see* **Note 8**).
9. Add 1 mL of 70 % v/v ethanol and centrifuge at 10,000 × *g* for 5 min at 4 °C.
10. Completely remove supernatant and dissolve the RNA pellet in 40–60 μL of formamide. Leave for 30 min at room temperature to allow RNA to dissolve and finally vortex to ensure solution is homogeneous. If necessary, the solution can be heated to 50 °C for 3 min to help dissolution.
11. To determine the final concentration of the RNA sample, dilute a few microliters of the solution around 40 times in water and determine the optical absorbance at 260 nm in a spectrophotometer using a 1 cm quartz microcuvette or with a NanoDrop® or similar apparatus. $1\,AU_{260\,nm}{}^{1\,cm} = 40$ μg/mL.

3.4 Denaturing Agarose Gel Electrophoresis

1. The electrophoresis chamber should be clean, but there is no need for it to be RNase-free. The gel tray and comb should be clean and free from dried agarose residue. It should be sprayed with RNaseZap® (Ambion Thermo Fisher Scientific), or the cheaper Windex® (SC Johnson, United States), then rinsed well with RNase-free water (*see* **Note 9**).
2. Place the electrophoresis apparatus in a chemical hood.
3. A 1% w/v agarose gel is ideal to fractionate the 47S to 18S ribosomal RNAs. For an 11 × 14 cm tray, weigh 1.2 g of agarose into an Erlenmeyer flask and add 85 mL of water. Boil until the agarose is fully dissolved, then quickly cool it to about 65 °C.

   The following steps should be performed in a chemical hood (*see* **Note 10**).
4. Add 35 mL of a freshly prepared mix of 37% w/w formaldehyde (23 mL) and 10× MOPS buffer (12 mL). Mix rapidly then pour into gel tray. Leave to polymerize for 1 h before submersing in 1× MOPS electrophoresis buffer.
5. Add 36 μL of RNA loading buffer to 1–10 μL of each RNA sample (2–4 μg) and heat for 3 min at 65 °C. Load samples immediately into gel pockets.
6. Electrophorese samples at 2 V/cm for 16–18 h.

3.5 Fluorescent Imaging of Gel and Transfer of RNA to Membrane

1. If ethidium bromide was included in the RNA loading buffer, the RNA fractionation can be imaged immediately following electrophoresis using the UV transilluminator and gel documentation system, e.g., *see* Fig 1a. If not, the gel must first be stained in electrophoresis buffer plus 0.5 μg/mL ethidium bromide. Take care to obtain a series of unsaturated exposures, as these will be used to correct the labeling data for any small variability in RNA loadings.
2. Cut a suitable sized sheet of Biodyne B membrane and soak in 2× SSC.
3. RNA is transferred to the membrane most effectively using a vacuum blotting system. We use a Biometra Vacu-Blot (Biometra GmbH, Göttingen, Germany) and perform transfer using 2× SSC for 1.5 h at a pressure of 200 mbar/cm^2.
4. If necessary, the RNA can be cross-linked to the membrane by UV irradiation at 70 J/cm^2.

For fluorographic imaging continue with Subheading 3.6, for phosphor-imaging skip Subheading 3.6 and go directly to Subheading 3.7.

3.6 Fluorographic Imaging and Quantitation by Scintillation Counting

The treatment of membrane with EN^{3}HANCE™, its subsequent manipulation, and the handling of scintillation liquid should all be performed in a chemical hood (*see* **Note 11**).

1. Label the corners of the membrane with a coded series of radioactive ink dots. Allow to dry thoroughly in a chemical hood on paper towels.
2. Still working in a chemical hood, spray the membrane with EN^{3}HANCE™ from a distance of 20–30 cm, moving evenly from aside. This treatment can be repeated a second time but should not be excessive. Allow membrane to dry in chemical hood for 10–15 min.
3. Place membrane in radiography cassette between Whatman 3MM paper. Place a sheet of Biomax XAR film directly in contact with the membrane. Do not interpose a layer of plastic film as this will negatively affect exposure. Close cassette and expose at −70 °C for 1 to several days depending on the expected signal. Typical fluorographic autoradiographs of pulse labeling experiments are shown in Figs. 1a and 3b and of a pulse-chase experiment in Fig. 2a.
4. After satisfactory fluorographic exposures have been obtained, place an appropriate fluorograph on a light-box, cover it with transparent Saran Wrap and place the membrane on top so that radioactive dots align. Hold membrane firmly in place and with a ballpoint pen mark each RNA band and one or more suitable unlabeled areas for estimates of background levels, *see* Fig. 1a. It is advisable not to use a ruler as it may stick to the membrane.
5. Remove membrane and, using scissors, cut out each RNA band and background area. Place each membrane fragment into a glass scintillation vial and add 10 mL of scintillation liquid. Count each vial in a liquid scintillation counter using an energy window suitable for tritium. To ensure an accurate count, inverse the membrane fragment in each vial and recount. If necessary, repeat this process to avoid counting errors (*see* **Note 12**).
6. To correct for any small variability in RNA loading, the digital ethidium bromide staining images from Subheading 3.5, **step 1** are used. The integrated 28S or 18S rRNA band densities determined from one or several unsaturated images should be equal across the gel if there is no error in gel loading. Small relative variations in 28 or 18S levels can be used to correct the radiolabel counts for each gel track.
7. Enter triplicate counts for each RNA species and track into an Excel or similar worksheet. Correct counts from each track for RNA loading variations and establish mean and standard error of

triplicates, e.g., *see* Fig. 1b. Normalization to one condition allows easy comparison of RNA synthesis and/or processing rates under various cell cultures and labeling regimes, e.g., *see* Fig. 1b.

3.7 Phosphor-Imaging

It is possible to replace the fluorography and quantitation procedures by direct exposure of the transfer membrane to storage phosphor screens, assuming suitable equipment for their scanning is available. The approach requires the use of special tritium-sensitive phosphor screens that lack a protective layer that would otherwise greatly reduce sensitivity of the approach. One disadvantage is that these screens are expensive and easily contaminated. To the best of our knowledge, a direct comparison of the detection sensitivities of fluorography and phosphor-imaging for tritium has not been published, but phophor-imaging has been applied to follow rRNA processing in yeast [15].

3.8 Determination of Transcription Elongation Rates

Direct determination of RPI transcription elongation rates has been successfully achieved in both growth-stimulated and growth-inhibited NIH3T3 cells using a series of labeling pulses of increasing length. These data revealed that the transcription elongation rate of RPI is growth regulated [3]. During radiolabeling, initially existing nascent transcripts will incorporate radiolabel such that only a short 3′-terminal segment of newly completed 47S rRNAs will be labeled, Fig. 3a. But as labeling continues the labeled segment of each completed transcript will increase in length, until by the time taken for a polymerase to traverse the whole 47S rRNA region, label will be incorporated throughout. Subsequently, all transcripts will be fully labeled and radiolabel in the 47S rRNA fraction will accumulate linearly until processing becomes significant. If we assume the cellular uptake of radiolabel is rapid compared with its incorporation into rRNA and processing is negligible during the time taken to synthesize the full-length precursor transcript, then the incorporation I of radiolabel into full length 47S rRNAs is described by;

$$I \propto \int_{t=0}^{t=l_g/\epsilon} n_g \rho_p \varepsilon t dt$$

Thus;

$$I \propto \left[n_g \rho_p \varepsilon t^2 \right]_{t=0}^{t<l_g/\epsilon}$$

Where;

I = incorporation of radiolabel into 47S rRNA,
l_g = gene length in base pairs,
t = time of labeling, minutes,
n_g = number of active genes,
ρ_p = density of polymerases, molecules/bp, and
ε = transcription elongation rate, nucleotides/min.

Simply put, I will vary as t^2 until the first full-length transcript is completed from initiation to termination, and subsequently I will vary linearly with t. By measuring radiolabel incorporation into 47S rRNA at increasing times (Fig. 3b), it is then possible to determine the mean time taken by RPI to synthesize a complete transcript. This elongation time l_g/ε can be read off from the plot as shown in Fig. 3c, and the elongation rate ε calculated from this and from the known length of the gene l_g.

4 Notes

1. Many sources of agarose have been tried and most found to function well.
2. The inclusion of ethidium bromide in the loading buffer is optional, but aids loading of gel and avoids the need for staining after electrophoresis. The ethidium bromide is cross-linked to the RNA during sample heating. Ethidium bromide should not be included in the electrophoresis buffer.
3. Early protocols used 2,5-Diphenyloxazole (PPO) or liquid EN^{3}HANCE™ as scintillant for fluorography [20]. In this case, the gel is not transferred to a membrane but treated directly and then vacuum dried.
4. An autoradiography phosphorescent marker pen can be used in place of the ^{14}C-ink.
5. Other makes of plastic wrap can be used but may be partially permeable to EN^{3}HANCE™.
6. Using the present protocol, after 30 min labeling only the 47S will be detected. To detect ribosomal RNA processing products, a pulse labeling of 2–3 h will be required. When using pulse labeling times of less than 30 min cells should be labeled by the addition of 50 μCi [^{3}H]-uridine in 100 μl medium. For [^{3}H]-methylmethionine labeling cells are washed twice with methionine-free medium and finally 2 mL of methionine-free medium containing 10 μCi of radiolabel is added. Typically cells are labeled for 1 h and either RNAs analyzed directly or after chasing for 1–6 h in fresh medium containing 0.5 mM methionine.
7. If a quantitative measurement of ribosomal RNA synthesis is required, it is preferable not to change the culture medium or to add fresh medium before labeling as this can significantly affect synthesis and hence radiolabel incorporation.
8. Leaving traces of the supernatant can cause migration problems during electrophoresis due to the carryover of guanidinium.
9. RNaseZap® or Windex® solution is strongly alkaline and will affect electrophoresis if not removed thoroughly.

10. Formaldehyde vapor is highly toxic and is especially strongly released from the hot agarose solution.
11. Due to the volatile and toxic solutions, treatment of membranes or gels with EN^{3}HANCE™ or with 2,5-Diphenyloxazole (PPO), their subsequent handling and preparation for scintillation counting should be performed in a chemical hood.
12. The membrane fragment will not be completely transparent and may in some cases affect the efficiency of counting. The potential error cannot be corrected using either external or internal standards, hence it is important to ensure the counts are fully reproducible.

References

1. Moss T, Langlois F, Gagnon-Kugler T, Stefanovsky V (2007) A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. Cell Mol Life Sci 64(1): 29–49
2. Stefanovsky VY, Pelletier G, Hannan R, Gagnon-Kugler T, Rothblum LI, Moss T (2001) An immediate response of ribosomal transcription to growth factor stimulation in mammals is mediated by ERK phosphorylation of UBF. Mol Cell 8(5):1063–1073
3. Stefanovsky V, Langlois F, Gagnon-Kugler T, Rothblum LI, Moss T (2006) Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling. Mol Cell 21(5):629–639. doi:10.1016/j.molcel.2006.01.023
4. Zhao J, Yuan X, Frodin M, Grummt I (2003) ERK-dependent phosphorylation of the transcription initiation factor TIF-IA is required for RNA polymerase I transcription and cell growth. Mol Cell 11(2):405–413
5. Murayama A, Ohmori K, Fujimura A, Minami H, Yasuzawa-Tanaka K, Kuroda T, Oie S, Daitoku H, Okuwaki M, Nagata K, Fukamizu A, Kimura K, Shimizu T, Yanagisawa J (2008) Epigenetic control of rDNA loci in response to intracellular energy status. Cell 133(4):627–639. doi:10.1016/j.cell.2008.03.030, S0092-8674(08)00459-5 [pii]
6. Gagnon-Kugler T, Langlois F, Stefanovsky V, Lessard F, Moss T (2009) Loss of human ribosomal gene CpG methylation enhances cryptic RNA polymerase II transcription and disrupts ribosomal RNA processing. Mol Cell 35(4): 414–425. doi:10.1016/j.molcel.2009.07.008, S1097-2765(09)00503-6 [pii]
7. Scherrer K, Latham H, Darnell JE (1963) Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells. Proc Natl Acad Sci U S A 49:240–248
8. Warner JR, Soeiro R, Birnboim HC, Girard M, Darnell JE (1966) Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA. J Mol Biol 19(2):349–361
9. Loening UE, Jones KW, Birnstiel ML (1969) Properties of the ribosomal RNA precursor in Xenopus laevis; comparison to the precursor in mammals and in plants. J Mol Biol 45(2):353–366, 0022-2836(69)90110-7 [pii]
10. Hamdane N, Stefanovsky VY, Tremblay MG, Nemeth A, Paquet E, Lessard F, Sanij E, Hannan R, Moss T (2014) Conditional inactivation of Upstream Binding Factor reveals its epigenetic functions and the existence of a somatic nucleolar precursor body. PLoS Genet 10(8):e1004505. doi:10.1371/journal.pgen.1004505
11. Penman S, Smith I, Holtzman E (1966) Ribosomal RNA synthesis and processing in a particulate site in the HeLa cell nucleus. Science 154(750):786–789
12. Greenberg H, Penman S (1966) Methylation and processing of ribosomal RNA in HeLa cells. J Mol Biol 21(3):527–535
13. Zimmerman EF, Holler BW (1967) Methylation of 45 s ribosomal RNA precursor in HeLa cells. J Mol Biol 23(2):149–161
14. Maden BE (1986) Identification of the locations of the methyl groups in 18 S ribosomal RNA from Xenopus laevis and man. J Mol Biol 189(4):681–699
15. Kos M, Tollervey D (2010) Yeast pre-rRNA processing and modification occur cotranscriptionally. Mol Cell 37(6):809–820. doi:10.1016/j.molcel.2010.02.024, S1097-2765(10)00206-6 [pii]

16. Schwinghamer MW, Shepherd RJ (1980) Formaldehyde-containing slab gels for analysis of denatured, tritium-labeled RNA. Anal Biochem 103(2):426–434
17. Gurney T Jr (1985) Characterization of mouse 45S ribosomal RNA subspecies suggests that the first processing cleavage occurs 600 +/- 100 nucleotides from the 5′ end and the second 500 +/- 100 nucleotides from the 3′ end of a 13.9 kb precursor. Nucleic Acids Res 13(13):4905–4919
18. Rosen KM, Lamperti ED, Villa-Komaroff L (1990) Optimizing the northern blot procedure. Biotechniques 8(4):398–403
19. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159
20. Laskey RA, Mills AD (1975) Quantitative film detection of 3H and 14C in polyacrylamide gels by fluorography. Eur J Biochem 56(2):335–341. doi:10.1111/j.1432-1033.1975.tb02238.x

Chapter 12

Quantitative Northern Blot Analysis of Mammalian rRNA Processing

Minshi Wang and Dimitri G. Pestov

Abstract

Assembly of eukaryotic ribosomes is an elaborate biosynthetic process that begins in the nucleolus and requires hundreds of cellular factors. Analysis of rRNA processing has been instrumental for studying the mechanisms of ribosome biogenesis and effects of stress conditions on the molecular milieu of the nucleolus. Here, we describe the quantitative analysis of the steady-state levels of rRNA precursors, applicable to studies in mammalian cells and other organisms. We include protocols for gel electrophoresis and northern blotting of rRNA precursors using procedures optimized for the large size of these RNAs. We also describe the ratio analysis of multiple precursors, a technique that facilitates the accurate assessment of changes in the efficiency of individual pre-rRNA processing steps.

Key words Mammalian cells, Ribosome biogenesis, Pre-rRNA, RNA electrophoresis, Northern blot, Hybridization

1 Introduction

Ribosome biogenesis in eukaryotes is a highly complicated process that involves transcription of rRNA precursors (pre-rRNAs), their processing, folding, assembly with ribosomal proteins, and transport of preribosomal particles from the nucleus to the cytoplasm. In mammalian cells, 18S, 5.8S, and 28S rRNAs are initially transcribed in the nucleolus as one contiguous molecule, termed 47S pre-rRNA. Processing of the primary 47S pre-rRNA transcript generates a series of intermediates named according to historically defined conventions, which vary for humans and rodent species [1]. Numerous ribosome assembly factors transiently associate with preribosomal particles during ribosome maturation and participate in pre-rRNA processing [2].

Analysis of pre-rRNA processing intermediates has been essential for understanding functions of ribosomal proteins and ribosome assembly factors. Changes in pre-rRNA processing can also reveal important information about cellular responses to

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_12,

various types of stress and drug treatments that affect the functioning of the nucleolus. In this chapter, we provide a detailed description of the procedures that have been used in our laboratory for characterizing and comparing pre-rRNA processing steps through the quantitative analysis of processing intermediates in mammalian cells. The protocols included here were developed in part to address several technical hurdles encountered in this type of analysis. In particular, agarose gel electrophoresis of RNA has been optimized to improve separation of large pre-rRNA species [3]. A modification of the downward blotting procedure [4] has been adopted to ensure a reliable transfer of pre-rRNAs to a membrane.

Identification of specific pre-rRNAs on northern blots is typically achieved through hybridization with oligonucleotide probes complementary to different parts of the pre-rRNA transcript. Interpretation of changes in pre-rRNA steady-state levels, however, can be challenging because these levels often reflect both the efficiency of several preceding processing steps and the rate of conversion to downstream species. The situation is further complicated by the fact that processing cleavages in pre-rRNA can occur in a varying order. One useful approach, which we term the ratio analysis of multiple precursors (RAMP), is to measure ratios between various pre-rRNAs across the entire pathway and display them in a simple graphical format (Fig. 1). In our experience, examining interrelations between multiple precursor ratios aids in explaining the observed patterns of changes and permits detection of subtle alterations in pre-rRNA processing that can be difficult to discern by directly comparing band intensity on northern blots. In

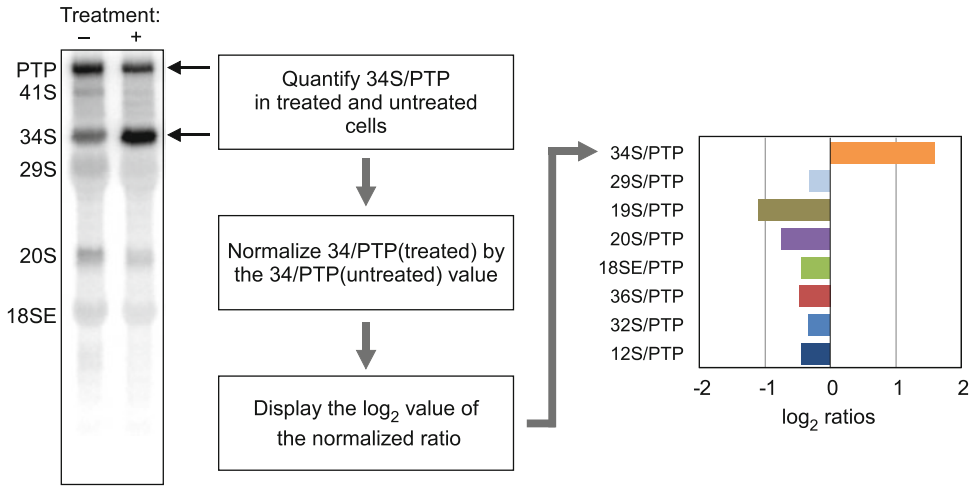

Fig. 1 An outline of steps in the Ratio Analysis of Multiple Precursors (RAMP). In the example shown, hybridization signals are determined in each lane for the 34S pre-rRNA and the comigrating 47S, 46, and 45S pre-rRNAs (primary transcript plus, PTP). The 34S/PTP ratios for treated and untreated cells are calculated. This procedure is then repeated using hybridizations with additional probes to obtain other pre-rRNA ratios. A RAMP profile is built by combining normalized precursor ratios in a single graph

addition, we find that the geometric shape of RAMP profiles exhibits consistency across a range of knockdown efficiencies for a given processing factor [5]. This makes RAMP especially useful for mammalian cell culture experiments, where variations in transfection and gene silencing efficiencies are commonly observed.

In the protocols described later, we refer to "treated" and "reference" cells as the starting biological material for simplicity. In drug treatments, reference cells should be exposed to the drug vehicle. In knockdowns utilizing gene-specific siRNA, reference cells should be transfected with nonsilencing siRNA since a transfection procedure alone can induce changes in pre-rRNA maturation. Thus, the RAMP profiles in siRNA-mediated knockdowns will represent the ratios of precursors in knockdowns normalized to the ratios in cells with nontargeting siRNA. It is recommended that at least three biological replicates be used for both the experimental conditions of interest and their matching no-treatment controls to ensure reproducibility of each profile.

2 Materials

2.1 Preparation of RNA

1. Monophasic reagent for RNA purification, such as RNazol RT or TRI Reagent (Molecular Research Center, Inc).
2. Deionized formamide. Store in 1–2-mL aliquots at –20 °C for up to a year.

2.2 Agarose Gel Electrophoresis of Pre-rRNA

The Hepes-triethanolamine (HT) buffer produces good electrophoretic separation of RNAs in a wide size range. The resolution of pre-rRNA species larger than 5 kb (32S–45S) can be further increased by using the Tricine-triethanolamine (TT) buffer (*see* **Note 1**). All solutions for electrophoresis of RNA must be prepared using molecular biology-grade reagents and deionized water that are free of RNases.

1. 50× HT buffer stock solution: 1.5 M Hepes, 1.5 M triethanolamine. Place a beaker on a balance and pour 56 g triethanolamine (free base) into the beaker. Add 89.4 g Hepes (free acid) and sterile deionized water to 230–240 mL, dissolve reagents completely using a magnetic stirrer and adjust the volume with water to 250 mL. The final pH is 7.5–7.7 and requires no adjustment. Filter through a 0.2 μm high-protein binding filter (nylon or nitrocellulose) to remove dust-borne RNases and other contaminants. The stock solution can be additionally autoclaved at 121 °C for 15 min, although we do not find this necessary when using high-quality reagents. Store at room temperature.
2. Running buffer: 30 mM Hepes, 30 mM triethanolamine. Prepare before use by diluting the 50× HT buffer stock with deionized water.

3. RNA dye: 2.1× HT buffer, 1 mM EDTA, 0.04% bromophenol blue. Prepare from the 50× HT buffer stock, 0.5 M EDTA (pH 8.0), and 0.5% bromophenol blue solutions. Filter through a nylon syringe filter into a disposable plastic tube. Mix well, divide into 1–2-mL aliquots, and store at −20 °C.
4. 37 % w/w formaldehyde.
5. Agarose, molecular biology grade.
6. Gel electrophoresis apparatus (*see* **Note 2**).

2.3 Blotting

1. 20× SSC: 3 M NaCl, 0.3 M sodium citrate. Adjust pH to 7.0 with HCl or NaOH, filter through a 0.2–0.45-μm filter, sterilize by autoclaving. Store at room temperature.
2. 10× SSC: prepare from 20× SSC before use.
3. Methylene blue stain: 0.03% methylene blue in 0.3 M sodium acetate (pH 5.0–5.5). Filter through a 0.2–0.45-μm filter and store at room temperature.
4. Nylon membrane (e.g., Amersham Hybond-N or similar).
5. Thick filter paper (Whatman 3MM or equivalent).
6. Kimberly-Clark WypAll L20 paper towels.
7. Plastic sheet protectors.
8. UV crosslinker.

2.4 Probe Labeling

1. Oligonucleotide probe, 10 μM in TE (*see* **Note 3** for probe design).
2. T4 polynucleotide kinase (10 U/μL) with 10× forward reaction buffer.
3. [γ-^{32}P]ATP, 3000 Ci/mmol, 10 μCi/μL.
4. Labeling stop solution: 10 mM Tris–HCl pH 8, 10 mM EDTA. Store at room temperature.
5. Spin columns containing a gel filtration matrix suitable for the separation of oligonucleotides from unincorporated nucleotides(BioRad Micro Bio-Spin 6 or equivalent).

2.5 Hybridization and Stripping

1. 50× Denhardt's solution: 1 % (w/v) Ficoll (type PM400), 1 % (w/v) polyvinylpyrrolidone, 1 % (w/v) BSA (fraction V). Dissolve all ingredients in deionized water, filter through a 0.45-μm filter. Store in 25–50-mL aliquots at −20 °C. This solution can be frozen and thawed several times.
2. 10 % (w/v) SDS. Filter through a 0.45-μm filter and store at room temperature for up to a year.
3. Hybridization solution: 5× SSC, 5× Denhardt's solution, 0.5% SDS (*see* **Note 4**). Prepare from stock solutions before use and keep at ≥37 °C to prevent precipitate from forming.

4. Hybridization wash solution: 2× SSC, 0.1 % SDS. Prepare from stock solutions before use (*see* **Note 4**).
5. Stripping solution: 0.1× SSC, 0.1 % SDS. Prepare from stock solutions before use.
6. Hybridization equipment, e.g., a hybridization oven.

2.6 Quantification of Pre-rRNA Levels and Data Analysis

1. Phosphorimager with image analysis software (e.g., a Typhoon laser scanning imager with ImageQuant software from GE).
2. Microsoft Excel.

3 Methods

Wear gloves and use RNase- and dust-free reagents and plasticware at all steps to avoid introducing RNase contamination. Formaldehyde is harmful to humans and should be handled in a fume hood. Use appropriate precautions for handling of radioactive materials.

3.1 Preparation of RNA

Total cellular RNA should be prepared using a method appropriate for the cellular material being analyzed. The volumes in the following example are for the RNA isolation from mouse 3T3 cells grown in a 12-well plate. If cells are cultured in other types of plasticware, adjust reagent volumes in proportion to the growth surface area.

1. Plate two sets of cells prior to the experiment, including one set of untreated or reference cells. For example, if using siRNAs to silence expression of a gene of interest, also at the same time set up a transfection with nontargeting siRNA (*see* **Note 5**).
2. Aspirate off cell medium and immediately pipet 300 μL RNazol RT or analogous RNA preparation reagent into each well. Tilt the plate in different directions to ensure rapid and complete lysis of all cells in the wells.
3. Transfer lysates from each well into microcentrifuge tubes and proceed with RNA purification following the protocol supplied with the reagent being used. At the final step, resuspend RNA in 15 μL formamide (*see* **Note 6**). To ensure that RNA is completely dissolved, incubate the pellet after formamide addition for 30 min on ice with occasional gentle mixing, then incubate for 5–10 min at 65–70 °C, mix the contents of the tube again and spin down for 10 s in a microcentrifuge.
4. Measure the RNA concentration using a spectrophotometer or a fluorescence-based assay. Use formamide for the blank samples.

3.2 Agarose Gel Electrophoresis of Pre-rRNA

Load samples from treated and reference cells on a single gel so that they can be transferred to the same membrane at the blotting stage.

1. To prepare a gel, combine in a flask (per 100 mL final volume): 97 mL deionized water, 2 mL 50× HT buffer, and 1 g agarose. Heat the mixture in a microwave oven until the agarose is completely dissolved. Add 3.5 mL of 37% formaldehyde in a fume hood, mix the solution thoroughly by swirling, cover the flask with a piece of aluminum foil and cool to 40–50 °C.
2. Working in a fume hood, pour the agarose mixture into a prepared electrophoresis tray with an inserted comb and cover the tray tightly with plastic wrap while taking care not to touch the molten agarose. Covering the tray reduces evaporative losses of formaldehyde from the gel. Gels should be used on the same day as they are prepared.
3. Pipet RNA samples dissolved in formamide into microtubes or PCR strip tubes (one tube per lane) and keep on ice. Typically, we use 1.5–2.5 μg total RNA per lane for pre-rRNA analysis when using a comb with 4–5-mm teeth.
4. Prepare the 2× formaldehyde loading solution by combining 14 volumes of the RNA dye with 1 volume of 37% formaldehyde. This solution must be freshly prepared as it is not stable upon storage.
5. Add the 2× formaldehyde loading solution prepared before to each RNA sample at a 1:1 volume ratio. Close tubes tightly, mix and spin down the contents of the tubes. Heat the samples at 70 °C for 5 min to denature the RNA and cool to room temperature (*see* **Note** 7).
6. Place the solidified agarose gel in an electrophoresis tank and add a sufficient volume of the running buffer to cover the gel. Load samples into wells without delay to prevent formaldehyde diffusion out of the gel.
7. Run the gel at 5–6 V/cm for the first 5–10 min and then reduce the voltage to 3.5–4 V/cm to lower the heat generation during the electrophoresis. Stop the run when bromophenol blue reaches 1–3 cm from the leading end of the gel.

3.3 Blotting

Gels prepared as described earlier require no further treatment prior to transfer. Gels can be left in the tank after electrophoresis under the running buffer for up to 15 min without harm.

1. Cut out a piece of a nylon membrane to a size 3–5 mm larger than the gel on all sides. Soak the membrane first in deionized water for 1 min and then in 10× SSC for at least 5 min.
2. Cut out six sheets of filter paper to a size 7–10 mm larger than the gel on all sides. Cut out two pieces of filter paper of the same width but long enough to serve as wicks reaching to the container filled with transfer buffer (Fig. 2).

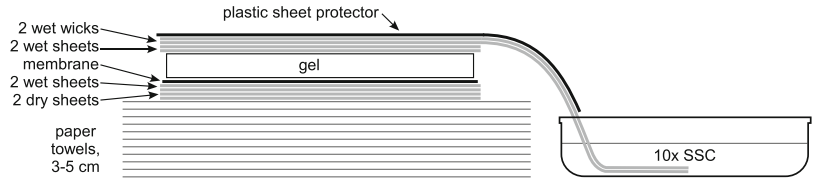

Fig. 2 Downward capillary transfer of RNA from an agarose gel to a nylon membrane. The gel and membrane are sandwiched between sheets of Whatman 3MM filter paper. The transfer solution is supplied from the top by the wicks made from two sheets of 3MM paper. The transfer solution is drawn through the gel into a stack of absorbent paper towels

3. Assemble the transfer stack in the following order (Fig. 2): paper towels, two dry sheets of filter paper, two sheets of filter paper wetted in 10× SSC, membrane, gel. Place narrow pieces of plastic (such as parafilm or sheet protectors) around the gel to prevent liquid from bypassing the gel (*see* **Note 8**). Place two sheets of filter paper wetted in 10× SSC on the top of the gel. Holding the two long pieces of the filter paper together, wet them in 10× SSC and place on the top so that their other end reaches into a suitable plastic container filled with transfer buffer (~1 mL/cm^2 of the gel). Make sure the wicks do not touch any part of the stack underneath the gel. At all steps, care must be taken to avoid air bubbles between gel, membrane, and layers of the filter paper (*see* **Note 9**).
4. Cover the wicks with a plastic sheet protector to prevent them from drying out during transfer. Place an empty tip box on the top to hold the assembly together. It is important not to place anything heavy on the top as compressing the gel matrix will trap large pre-rRNA molecules in the gel. Allow the transfer to proceed for at least 6 h or overnight.
5. Disassemble the transfer stack. Briefly rinse the membrane with deionized water. Stain the membrane with methylene blue to assess RNA integrity and transfer quality. Agitate the blot in the methylene blue stain solution for 1–2 min or until 28S and 18S rRNA bands appear on the blot, then destain with 3–4 changes of deionized water while constantly agitating the membrane. Destaining typically takes less than a minute; when a sufficient contrast between rRNA bands and the membrane is obtained, drain the membrane and blot excess liquid with filter paper from the back of the membrane (*see* **Note 10**).
6. UV crosslink RNA to the membrane using conditions recommended by the membrane or crosslinker manufacturers. This step should be performed while the membrane is still moist.
7. Membranes can be used immediately for hybridizations, kept in 0.1× SSC or dried and stored between sheets of filter paper for up to 2–3 days.

3.4 Probe Labeling

Because pre-rRNAs are abundant cellular RNA species, they can be detected with oligonucleotide probes 5′-labeled using the T4 polynucleotide kinase. In the protocol described here, labeled oligonucleotides are separated from unincorporated [γ-^{32}P]ATP by size-exclusion chromatography using spin columns, yielding ~50 μL of radiolabeled probe, typically sufficient for 3–5 hybridizations.

1. In a microcentrifuge tube, combine 14.4 μL water, 0.6 μL 10 μM oligonucleotide probe (6 pmol), 2 μL 10× T4 polynucleotide kinase reaction buffer, 2 μL [γ-^{32}P]ATP, 1 μL T4 polynucleotide kinase to a total reaction volume of 20 μL. Mix the solution gently, and briefly spin down in a microcentrifuge. Incubate the reaction mixture for 30–60 min at 37 °C.
2. Add 30 μL of the Labeling stop solution. Separate the labeled oligonucleotides from unincorporated label by spin column purification (follow the manufacturer's protocol). Using a handheld Geiger monitor, estimate counts in the eluate and the spin column. A successful labeling of the probe should result in >50% of the radioactivity counts detected in the eluate.

3.5 Hybridization and Stripping

1. Wet the membrane carefully in deionized water. Place the membrane into the hybridization solution (≥0.25 mL/cm^2 of membrane) and incubate for 1 h at 50 °C for prehybridization.
2. If a blot is being probed for the first time, discard the prehybridization solution, add fresh hybridization solution prewarmed to 37–50 °C, and then add the radiolabeled probe. For the membranes that have been previously hybridized and stripped, probe can be added directly to the prehybridization solution. Hybridize at 50 °C overnight.
3. Safely dispose of the hybridization mixture containing the radioactive probe and rinse the membrane once with the hybridization wash solution. Place the membrane in a suitable tray and gently agitate in several changes of the wash solution: twice for 5 min at room temperature, 5 min at 50 °C, and 5 min at room temperature. Rinse the membrane briefly with 2× SSC.
4. Cover the membrane with plastic wrap and expose to a phosphorimager screen. Keeping the membrane moist during the exposure will help with subsequent stripping of the probe.
5. Scan the exposed phosphorimager screen at 100 μm per pixel or better resolution.
6. If rehybridization with a different probe is desired, discard the plastic wrap and strip the membrane: pour boiling stripping solution into a plastic tray, submerse the membrane, and incubate for 10 min while gently rocking the membrane at the room temperature. Discard the stripping solution and repeat stripping one more time (*see* **Note 11**).

3.6 Quantification of Pre-rRNA Levels and Data Analysis

A primary goal of RAMP is to accurately characterize the observed changes in pre-rRNA processing [5]. A set of precursor ratios describing quantitative relations between various processing intermediates is first calculated for a sample of interest and then normalized to the precursor ratios in a reference sample. The resulting values are displayed in a compact color-coded graphical format (Fig. 1).

1. Using software for the analysis of phosphorimager data, determine the total volume of the hybridization signal in the bands corresponding to pre-rRNA species detected with each probe. Note that all hybridization probes internal to the rapidly processed sites A′ and 02 in human pre-rRNA (sites A′ and 6 in the mouse) will also hybridize with the primary transcript 47S and the closely migrating 46S and 45S pre-rRNAs. The combined level of 47S-45S pre-rRNAs, or "primary transcript plus" (PTP, Fig. 1), is a convenient denominator in calculating precursor ratios. Determine PTP values by phosphorimager analysis in all lanes for each hybridization.
2. In Excel, calculate ratios of precursors (e.g., 32S/PTP, 34S/PTP, *see* Fig. 1) for the treated cells and matching reference cells. For each individual ratio calculation, use values obtained from the same lane and in the same hybridization. Log2 transform the ratios.
3. Determine treatment-related change in the ratios by subtracting log-transformed values in reference cells from the corresponding values in treated cells. Calculate the log average and standard deviation for the biological replicas.
4. Plot the normalized log average values according to a predetermined order along the Y axis (Fig. 1). To facilitate profile comparison, this order should be the same for the entire experimental set. Color coding of individual ratios will help to visualize patterns in RAMP profiles.

4 Notes

1. For enhanced separation of very large precursors, use 50× Tricine–triethanolamine (TT) stock solution (1.5 M Tricine, 1.5 M triethanolamine) instead of HT to prepare running buffer, RNA dye, and to make gels. To prepare the 50× TT stock solution, mix 67.3 g Tricine (acid) and 56 g triethanolamine (base) per 250 mL of solution, otherwise following the directions found in the HT recipe. The pH of the TT buffer is 7.8–8.0 and requires no adjustment.
2. Although formaldehyde inactivates RNases to some extent, it is advisable not to run RNA samples in electrophoresis tanks

that are also used for bacterial plasmid preps. Most kits for plasmid preparations contain large quantities of RNases capable of causing RNA degradation even in the presence of formaldehyde and these RNases may be difficult to remove completely from the equipment.

3. We obtain good results with oligonucleotide probes 20–25 nt in length using the hybridization and washing conditions described here. Avoid designing probes with a 5′ terminal C residue because they are poorly labeled by the T4 polynucleotide kinase. In contrast to highly conserved rRNA sequences, transcribed spacers in pre-rRNA diverge rapidly, occasionally leading to sequence variations between different strains of the same species. If a newly designed probe fails to hybridize, redesign the probe using the actual pre-rRNA sequence in the cells under study.
4. To avoid SDS precipitation, do not allow concentrated SDS and SSC to come in contact during preparation of the hybridization and wash solutions. It is convenient to dilute 20× SSC into deionized water first, then add the required volume of 10% SDS and mix thoroughly. If the room temperature is too low, precipitate may gradually form in the wash solution. In that case, warm up the solution to 40–45 °C to dissolve the precipitate and allow the solution to return to ~25 °C.
5. Transcription rates of rRNA in mammalian cells can change significantly in response to growth factors and nutrients in medium. Where possible, compare cells that are at a similar density at the time of lysis. Avoid using cell cultures that are overgrown because inhibition of rRNA transcription may greatly decrease the amount of rRNA precursors in these cells.
6. Dissolving RNA in formamide simplifies handling of samples by stabilizing RNA against the action of RNases [6]. We find that RNA dissolved in formamide can be stored for several years at −20 or −80 °C. If RNA was not dissolved in formamide, adjust the loading dye composition to obtain the following reagent concentrations in samples loaded on a gel: 50% (v/v) formamide, 0.4 M formaldehyde, 1× running buffer, 0.5 mM EDTA, 0.02% bromophenol blue.
7. If RNA pellets were insufficiently dried prior to their reconstitution, samples loaded on a gel may be pushed out of the wells by the residual ethanol mixing with the water. This is probably exacerbated by the small amounts of methanol present in most commercial grades of formaldehyde. We recommend to spin samples prior to heating them but not after. Residual alcohols evaporate from samples during the heating step and condense at the top of the tube.
8. Always check that the wicks and pieces of the filter paper on top of the gel do not touch any part of the paper stack underneath

the gel. Any contact that short-circuits liquid flow may lead to incomplete or uneven transfer of large pre-rRNAs.

9. To prevent air bubbles from being trapped during transfer assembly, roll a pipet over the membrane and over the layers of the filter paper placed on the top of the gel. Another critical step is placing the gel on the membrane. A technique that works in our hands is to pour some 10× SSC along one side of the membrane and immediately place one edge of the gel into the resulting puddle of the buffer, then smoothly lay down the rest of the gel, using the liquid to displace the air from the gap between the gel and the membrane. This must be done quickly before the liquid is drawn into the underlying paper towels.
10. The methylene blue solution can be reused several times. When the staining intensity of rRNA bands begins to decrease, discard the old solution and make a fresh stain. Methylene blue is removed from membranes during prehybridization. A small amount of the stain may remain after the first prehybridization, but this does not affect hybridization quality or efficiency.
11. Exposure to SDS-containing solutions at elevated temperatures are harsh conditions that shorten the useful life of hybridization membranes. Do not exceed the recommended incubation times or stringency of stripping as it might lead to high background in subsequent hybridizations. Probes that hybridize with abundant pre-rRNAs (such as 32S) can be difficult to remove completely. Instead of increasing the duration of stripping, it is advisable to plan a series of hybridizations starting with detection of less abundant pre-rRNA species. A single membrane can be typically stripped and reprobed 4–5 times.

Acknowledgement

This work was supported by NIH grant GM074091 to DGP.

References

1. Henras AK, Plisson-Chastang C, O'Donohue M-F et al (2014) An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip Rev RNA 6:225–242
2. Tafforeau L, Zorbas C, Langhendries J-L et al (2013) The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. Mol Cell 51:539–551
3. Mansour FH, Pestov DG (2013) Separation of long RNA by agarose-formaldehyde gel electrophoresis. Anal Biochem 441:18–20
4. Chomczynski P (1992) One-hour downward alkaline capillary transfer for blotting of DNA and RNA. Anal Biochem 201:134–139
5. Wang M, Anikin L, Pestov DG (2014) Two orthogonal cleavages separate subunit RNAs in mouse ribosome biogenesis. Nucleic Acids Res 42:11180–11191
6. Chomczynski P (1992) Solubilization in formamide protects RNA from degradation. Nucleic Acids Res 20:3791–3792

Part III

Genomics and Proteomics of the Nucleolus

Chapter 13

Complete Sequence Construction of the Highly Repetitive Ribosomal RNA Gene Repeats in Eukaryotes Using Whole Genome Sequence Data

Saumya Agrawal and Austen R.D. Ganley

Abstract

The ribosomal RNA genes (rDNA) encode the major rRNA species of the ribosome, and thus are essential across life. These genes are highly repetitive in most eukaryotes, forming blocks of tandem repeats that form the core of nucleoli. The primary role of the rDNA in encoding rRNA has been long understood, but more recently the rDNA has been implicated in a number of other important biological phenomena, including genome stability, cell cycle, and epigenetic silencing. Noncoding elements, primarily located in the intergenic spacer region, appear to mediate many of these phenomena. Although sequence information is available for the genomes of many organisms, in almost all cases rDNA repeat sequences are lacking, primarily due to problems in assembling these intriguing regions during whole genome assemblies. Here, we present a method to obtain complete rDNA repeat unit sequences from whole genome assemblies. Limitations of next generation sequencing (NGS) data make them unsuitable for assembling complete rDNA unit sequences; therefore, the method we present relies on the use of Sanger whole genome sequence data. Our method makes use of the Arachne assembler, which can assemble highly repetitive regions such as the rDNA in a memory-efficient way. We provide a detailed step-by-step protocol for generating rDNA sequences from whole genome Sanger sequence data using Arachne, for refining complete rDNA unit sequences, and for validating the sequences obtained. In principle, our method will work for any species where the rDNA is organized into tandem repeats. This will help researchers working on species without a complete rDNA sequence, those working on evolutionary aspects of the rDNA, and those interested in conducting phylogenetic footprinting studies with the rDNA.

Key words Ribosomal RNA, rDNA, Whole genome assembly, Whole genome sequencing, Repeats, Sanger, Arachne

1 Introduction

DNA sequencing has allowed the architecture and complexity of many species' genomes to be revealed in the form of genome assemblies. This availability of genome assemblies has facilitated the study of organisms' genomic characteristics by assisting in identification of functional elements [1, 2], understanding of the transcriptional and epigenetic behaviors of the genome [3–9], deciphering of

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_13,

genomic evolutionary history [10–12], and predicting the 3D structure of the genome [13–15], among other goals. However, despite being designated as "complete," most eukaryote genome assemblies contain significant gaps [16]. These gaps usually encompass the repetitive regions of the genome, which are often not seen as a priority to complete as they are thought to be largely nonfunctional. However, the importance of acquiring truly complete genome sequences is increasingly becoming recognized [17, 18].

One of the most prominent gaps in the genome assemblies corresponds to the nucleolar organizer regions (NORs) [19–22]. NORs are the sites of nucleolus formation [23], and the nucleolus is the site of ribosome biogenesis [24, 25], which is essential for cell survival. While the function of the nucleolus as a ribosomal factory has been well established, far less is known about the underlying genomic sequence around which the nucleolus is formed. The primary constituent of NORs is one or more tandem repeat arrays of ribosomal RNA encoding genes (rDNA) [24, 25]. A single rDNA unit consists of a ribosomal RNA (rRNA) coding region, which varies in size from ~4.4 (*Giardia muris*) to ~13.4 kb (*Mus musculus*), and an intergenic spacer (IGS), which varies in size from ~1.4 kb (*Oxytricha fallax*) to 31.9 kb (*Mus musculus*) [26–29] (Fig. 1). rRNA is a major component of ribosomes [30] and forms a large fraction of the total RNA in a cell. However, in addition to its primary role in rRNA production, the rDNA has also been shown to mediate an increasing constellation of "extra-coding" functions, including roles in genome stability [31], cell cycle control [32–34], protein sequestration [35], epigenetic silencing [36], and heat shock response [35]. The majority of these biological processes are mediated by functional elements that lie within the IGS, and these include rRNA transcriptional regulators (promoters [37–39], terminators [40–42] and enhancers [43]),

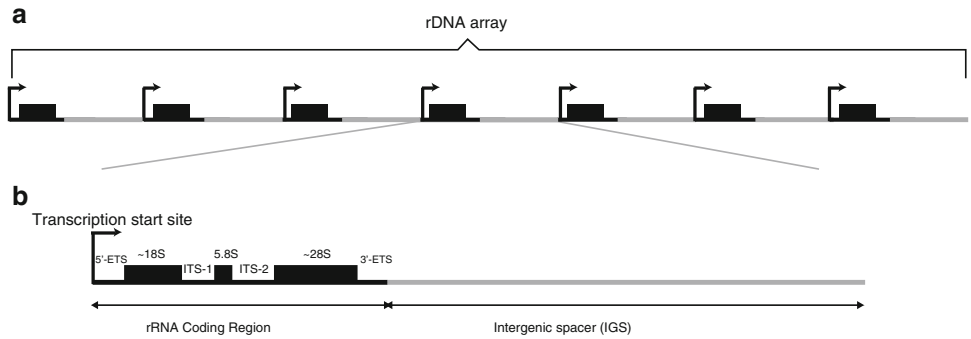

Fig. 1 Eukaryotic ribosomal DNA organization. (**a**) Tandem arrangement of rDNA units. Usually there are more units present in an array than are depicted here. (**b**) Each rDNA unit has an rRNA coding region (*black line*) and an intergenic spacer (IGS; *gray line*). The coding region encodes the ~18S, 5.8S, and ~28S rRNAs (*black boxes*), and these coding units are separated by two internal transcribed spacers (ITS-1 and 2) and flanked by two external transcribed spacers (5′-ETS and 3′-ETS)

origins of replication [44–56], replication fork barrier sites [44, 57–60], protein binding sites [61, 62], and noncoding RNA transcripts [35, 63–67]. Despite this, for most species the only part of the rDNA unit for which sequence information is available is the rRNA coding region. Therefore, determination of the complete rDNA sequence is imperative to characterize the molecular mechanisms that underlie these extra-coding processes. However, the complete rDNA unit sequence is difficult to obtain due to the high copy number and internal repetitive structure of the rDNA. Here, we describe a protocol that can be employed to construct the complete rDNA unit sequence of an organism using publicly available whole genome sequencing (WGS) data.

1.1 Theoretical Basis of the Method

The rDNA copy number of most eukaryotes varies from a few tens to a few thousands of copies [68, 69]. Construction of rDNA unit sequences using whole genome assemblies (WGA) is based on the assumption that rDNA copies are essentially identical in sequence within a genome, something that is largely supported by existing evidence [70–72]. Based on this assumption, the coverage of rDNA reads from WGS data should be many-fold higher than the coverage of reads from the unique regions of the genome. Therefore, if WGS data is subjected to WGA, the rDNA is expected to form high-coverage contigs where the rDNA reads collapse down to a single "consensus" rDNA unit sequence. These rDNA-containing contigs can be identified by searching for similarity to known rDNA sequences. The high level of conservation of the coding regions means that this approach should be successful even with species for which no rDNA sequence is currently available, and the widespread use of rDNA sequences in phylogenetic analyses means that some rDNA sequence information is available for many species.

1.2 Sanger Whole Genome Sequencing Reads Are Required to Overcome rDNA Complexity

In principle, this method should work equally well for both Sanger and NGS sequencing data. However, in practice we have not been able to assemble complete rDNA sequences using NGS (Illumina and 454) sequencing data. Although exploration of the reasons for this is beyond the scope of this article, we believe the problems likely stem from the dual repetitive nature of the rDNA and its unusual sequence characteristics. The rDNA is both present in multiple copies in the genome [68, 69], and it usually contains a number of repeat elements within each unit [27, 73]. The multiplicity of the rDNA can cause problems because most assemblers use a maximum coverage cutoff to differentiate between repeat and unique regions [74]. If the depth of the coverage for a region such as the rDNA is higher than this cutoff, all the reads from the region are tagged as repeats and removed from the assembly. In addition, the rDNA is highly variable in its level of internal repetition, ranging from low (e.g., 2.84% in *Giardia muris*) to high

(e.g., 39.1% in human), with both dispersed and tandem repeats (sub-repeats) found. Dispersed repeats, such as Alu elements in the mammalian rDNA [27, 73], can cause the same read placement ambiguity problems that they cause for the unique regions of the genome. Sub-repeats are not as problematic for assemblers, but their copy number can collapse down, and thus they may be represented in fewer than the true number of repeats, or not even appear as repeats at all. While these issues make the repetitive nature of the rDNA a challenge regardless of sequencing platform, the small read lengths of the common NGS technologies may exacerbate the problems (although see below). Finally, in our experience some regions of the rDNA are highly biased against in NGS libraries. These excluded regions largely, but not completely, correlate with GC-content, thus we speculate that the bias is a consequence of secondary structure. For these reasons, we have found it necessary to use assemblies of Sanger whole genome sequencing data to construct complete rDNA repeat sequences. However, it is possible that long-read sequencing technologies such as PacBio and/or Oxford Nanopore could successfully be used to assemble the rDNA, as the multiplicity of the rDNA would compensate for the higher current error rates of these technologies.

1.3 A Genome Assembler That Can Assemble the rDNA Sequence

By definition, reads from different copies of the same repeat family have similar sequences, making it difficult to differentiate between them. Therefore, as outlined above, assemblers remove the reads from repeat regions to reduce assembly complexity. Most Sanger read assemblers are based on the overlap-layout-consensus algorithm for read assembly [74, 75], and these mark reads as repeats based on the coverage of the contigs. The reads from repeat regions pile up to form high-coverage contigs, and contigs with coverage higher than the cutoff are marked as repeat contigs and usually discarded before proceeding further. Naturally, this process also discards rDNA contigs because of their high coverage levels. Therefore, although the scheme to obtain the rDNA unit sequence using WGA is straightforward in principle, in practice it is more challenging. We compared various assemblers viz. gsAssembler ver. 2.3 (Newbler) [76], the Celera assembler ver. 6.1 [77], Phrap ver. 1.090518 [78], MIRA ver. 3.2.1 [79], and Arachne [80, 81] to assess their efficiency at resolving rDNA repeat sequences. Using assemblies of primate WGS datasets as test cases, we found that only the Arachne assembler is able to assemble the rDNA repeat efficiently. The remaining four failed either because they discard rDNA contigs (gsAssembler and the Celera assembler), or have unreasonably high RAM requirements (Phrap and MIRA). Arachne runs efficiently with achievable RAM requirements (see below), and separates but does not discard the repeat regions [82]. Therefore, we recommend using Arachne with Sanger sequence data to generate complete rDNA unit sequences.

2 Material

1. **Arachne**: Open source software from the Broad Institute (ftp://ftp.broadinstitute.org/pub/crd/nightly/arachne/) that has to be installed on a server. Our protocol was successfully tested using Arachne versions r37405 and r37578, but ver. r37578 is more memory efficient and faster, therefore we recommend using this (ftp://ftp.broadinstitute.org/pub/crd/nightly/arachne/2011/2011-06/) or higher.
2. **CLC Genomics Workbench**: Used for analyzing and visualizing sequencing data (www.clcbio.com/products/clc-genomics-workbench). Other genome visualization applications are also likely to be suitable.
3. **Consed**: Used for visualization, manipulation, and finishing genome assemblies [83]. The installation of Consed involves several command line steps (described in the Consed manual) [84]. Consed accepts contig information in ACE file format. To run Consed, a specific arrangement of the directories is required. First, create a project directory that has subdirectories "edit_dir," and "phd_dir," "chromat_dir." Copy the .ace file in the directory "edit_dir." Run Consed in "edit_dir."
4. **Basic Local Alignment Search Tool** (**BLAST**): The standalone version of this sequence search tool can be obtained from the NCBI website (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/).
5. **Hardware**: To run Arachne WGAs, a 64 bit, multi core server with high RAM is required. We recommend using a server that has at least 75 GB RAM, although the precise computational requirements to finish an assembly differ depending on the number of reads used, and the size and complexity of the genome.

3 Methods

3.1 Acquisition of Whole Genome Sequencing Data

1. Download WGS data (reads) in fasta (sequence information of reads), quality (corresponding quality scores), and traceinfo (ancillary information) files (www.ncbi.nlm.nih.gov/Traces/trace.cgi) using the Perl script "query_tracedb" provided by NCBI (ftp://ftp-private.ncbi.nlm.nih.gov/pub/TraceDB/misc/query_tracedb) (*see* **Note 1**).
   The query string to be used for query_tracedb to download Sanger WGS data for an organism is:

```
SPECIES_CODE='Species name' and TRACE_TYPE_CODE='WGS' and
CENTER_NAME='center code'
```

   The query_tracedb script is sensitive to spaces adjacent to the relation operators (=, !=, >, <), so avoid adding spaces on either

side of "=" in the query. The parameter "SPECIES_CODE" defines the species whose WGS data is downloaded. The parameter TRACE_TYPE_CODE = "WGS" ensures that only whole genome sequencing data are extracted. WGS data can be present from more than one sequencing center, and to download WGS data from only one sequencing center it is necessary to define the "CENTER_NAME". We recommend not using WGS data from different sequencing centers together as input for Arachne, as this will increase the complexity of the assembly (*see* **Note 2**). The details of these parameters can be obtained from Taxonomy browser on the NCBI website (www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root). The query string with correct parameters can be generated using query builder on the Trace archive search website (www.ncbi.nlm.nih.gov/Traces/trace.cgi?view=search).

The steps to download the data using the script are described on the Trace archive searching tips webpage [85]. An example showing the download of WGS data using the query_tracedb script for chimpanzee (*Pan troglodytes*) is given below.

2. Count the number of reads in WGS dataset:

```
query_tracedb "query count SPECIES_CODE='PAN TROGLODYTES' and
CENTER_NAME='BI' and TRACE_TYPE_CODE='WGS'"
```

13,288,075

3. Download trace identifier (TI) number for all the reads. The TI number is a unique id number assigned to every read individually during the NCBI submission. As the query_tracedb script can only download 40,000 reads in one instance, downloading 13,288,075 reads requires 333 files (or "pages") to be downloaded for the complete WGS dataset. Therefore 333 downloads are required to obtain the TI numbers for all the reads (*see* **Note 3**). This can be automated by running the following command in the terminal:

```
export count
for count in {0..332}; do
query_tracedb "query page_size 40000 page_number $count binary
SPECIES_CODE='PAN TROGLODYTES' and CENTER_NAME='BI' and
TRACE_TYPE_CODE='WGS'" >> page_$count.bin
done
```

4. Download the fasta, quality, and traceinfo files using the generated binary files. Again, 333 downloads are required (*see* **Note 4**), which can be automated by running the following command in the terminal:

```
export count
for count in {0..332}; do
(echo -n "retrieve fasta 0b"; cat page_$count.bin) | ./query_tracedb >
data_$count.fasta
```

```
(echo -n "retrieve quality 0b"; cat page_$count.bin) | ./query_tracedb >
data_$count.qual
(echo -n "retrieve xml 0b"; cat page_$count.bin) | ./query_tracedb >
data_$count.xml
done
```

3.2 Preprocessing WGS Data for Arachne Input

The entire downloaded WGS data needs to be preprocessed before use as input for Arachne. The steps for preprocessing the downloaded files are as follows (*see* **Note 5**):

1. **Fasta file**: The header line for each NCBI read entry has three values: the trace identifier (TI) number as the identification code; the trace name; and the trace archive number of the mate pair (*see* **Note 6**).

   Arachne uses the trace name as the identification code, not the TI number, and only accepts the identification code (with no other description) in the fasta header line. Therefore modify the header of each fasta sequence (*see* **Note 7**). After modifying the header lines of the downloaded fasta files, concatenate them into a single fasta file named *reads.fasta*. The header line modification and concatenation can both be achieved, for the *Pan troglodytes* example, using the following command:

```
export count
for count in {0..322}; do
sed -e "s/gnl|ti|[A-Za-z0-9]*\s//g" -e "s/mate:[A-Za-z0-9]*//g"  -e "s/name://g"
data_$count.fasta >> reads.fasta
done
```

2. **Quality file**: Modify the header line of the quality file and concatenate all the modified quality files into a single quality file *reads.qual*. Similar to the fasta file, Arachne needs only the trace name in the header line of quality file. However, the header line for each NCBI read entry has two values: the TI number and the trace name (*see* **Note 8**). The header line modification and concatenation can both be achieved, for the *Pan troglodytes* example, using the following command (the order of the reads in the concatenated quality file should be the same as in the concatenated fasta file):

```
export count
for count in {0..332}; do
sed -e "s/gnl|ti|[A-Za-z0-9]*\s//g" -e "s/name://g" data_$count.qual >> reads.qual
done
```

3. **Traceinfo file**: This file is in xml format and contains the ancillary read information. Arachne only requires a subset of fields from this file to run. These are trace_name, plate_id, well_id, template_id, clip_vector_left, clip_vector_right, trace_end, insert_size, insert_stdev and type (*see* **Note 9**). The fields center_name, seq_lib_id (library_id), and ti are also accepted by Arachne but are not essential. The details of these fields are

given on the Arachne website [86]. Although Arachne accepts the TI number, we recommend removing it to avoid any confusion (*see* **Note 10**). Extract the necessary information from the NCBI traceinfo file and then, as before, concatenate all the traceinfo files into one *reads.xml* file. Again using the *Pan troglodytes* example, the xml files can be appropriately modified and all the modified traceinfo files concatenated into a single quality file *reads.xml* using the following command (the order of the reads in the concatenated quality file should be the same as in the concatenated fasta file):

```
export count
echo "<?xml version=\"1.0\"?>" >> reads.xml
echo "<trace_volume>" >> reads.xml
for count in {0..332}; do
grep -e "<trace>" -e "<trace_name>" -e "<plate_id>" -e "<well_id>" -e
"<template_id>" -e "<clip_vector_left>" -e "<clip_vector_right>" -e "<insert_size>" -e
"<insert_stdev>" -e "<trace_end>" -e "<center_name>" -e "<seq_lib_id>" -e
"</trace>"  data_$count.xml >>  reads.xml
done
echo "</trace_volume>" >> reads.xml
```

3.3 Arachne Input Files and Their Arrangement

In addition to the processed fasta, quality, and traceinfo files, several other auxiliary files are required by Arachne. The files need to be arranged in a very specific directory hierarchy [87]. All the input directories are subdirectories of the folder from which the assembly will be run. This head directory is defined by the variable "PRE." The details of the input files and their arrangement can be obtained from the Arachne website [88]. A brief summary of the "PRE" subdirectories and the files in them follows:

1. **dtds**: contains the XML file *configuration.dtd* that contains rules and macros together with comments to run the assembly [89]. An example of this file can be obtained from the "dtds" directory of the sample projects for Arachne (ftp://ftp.broadinstitute.org/pub/crd/ARACHNE/sample_projects.tar.gz).
2. **e_coli**: contains a file, *contig.fasta* that has the sequence(s) of any potential sources of contamination.
3. **e_coli_transposons**: Similar to the "e_coli" subdirectory, but the *contig.fasta* file has sequences of potential transposon contamination.
4. **vector**: contains a file, *contig.fasta*, that has the sequence of the vector used for sequencing.
5. **data**: the name of this subdirectory can vary and is defined by the variable "DATA" during the assembly run. It contains three subdirectories:

   fasta: contains the file *reads.fasta* with the sequences of the reads in fasta format.

qual: contains the file *reads.qual* with the quality scores of the reads in qual format.

traceinfo: contains the file *reads.xml* with the ancillary information.

The "DATA" subdirectory also contains several files. Detail of the files in the "DATA" subdirectory can be obtained from the Arachne website [88]. The files essential for the assembly are:

genome.size: contains the estimated size of the genome (in bp) to be assembled.

reads_config.xml: has the set of rules that are applied to the input reads. The FindXmlFeatures module in Arachne parses the *reads.xml* file to identify any essential missing features [90]. The rules have to be included in the *reads_config.xml* file to compensate for any missing information. The layout of the file can be obtained from the Arachne website [89].

3.4 Performing the Whole Genome Assembly

1. After arranging the data as required by Arachne in the working directory "PRE," perform the assembly using the script Assemblez. To start the assembly, the command is:

```
/absolute/path/to/Assemblez PRE=/path/to/working/directory DATA=data RUN=run
num_cpus=number_of_cpu num_cpus_pi=10 ACE=True
REMOVE_INTERMEDIATES=False
```

"PRE" defines the absolute path of the directory where the assembly is to be performed. "DATA" defines the name of the subdirectory in the "PRE" directory (PRE/data) where all the input data needed to run the assembly are present. "RUN" defines the name of the directory that will be generated at the start of the assembly (PRE/data/run) and contains the subdirectory "work" that houses all the intermediate files during the assembly. The ACE flag is set to "true" to generate the .ace files for the contigs, as these are required for post processing of the assembled contigs. Further, the flag REMOVE_INTERMEDIATES is set false to retain the intermediate assembly files in case the assembly breaks and needs to be restarted (*see* **Note 11**).

2. The genome assembly can take a few days or even weeks to finish depending on the complexity and size of the genome, and the number of reads used for the assembly. Several files will be generated in the "PRE/data/run" directory, including *assembly.bases.gz* [91] that contains the sequences of all the assembled contigs in fasta format. In addition, a directory "acedir" is generated where the .ace files are stored (*see* **Note 12**). The directory "acedir" also contains a log file *CreateAce.log* that has the names of the contigs in each .ace file.

3.5 Construction of the rDNA Unit Sequence

After generating the WGA, the rDNA repeat sequence must be constructed using the assembled contigs. The steps involved are as follows:

3.5.1 Generate a BLAST Database from the Assembled Contigs

To screen the obtained assembly for rDNA contigs, a database that can be read by BLAST has to be created for the contig file *assembly.bases.gz*. The steps to create such a database are as follows:

1. Decompress the file *assembly.bases.gz* and rename it *assembly.fasta* using the commands:

```
gzip -d assembly.bases.gz
mv assembly.bases assembly.fasta
```

2. Create a BLAST database for the assembled contigs using the command:

```
makeblastdb -input_type fasta -dbtype nucl -in assembly.fasta
```

3.5.2 Define the Threshold Coverage to Differentiate Between rDNA and Non-rDNA Contigs

The coverage of a contig is the average number of reads in the WGS data covering each base of the contig. Multiple copies of the rDNA are present per genome, therefore the average coverage of rDNA-containing contigs should be several fold higher than contigs containing the unique regions of the genome. This property of rDNA-containing contigs can be used to differentiate between true and false-positive rDNA contigs during screening of the assembly database. The steps to calculate the baseline coverage of the unique regions of the genome are as follows:

1. Extract the sequences of five known single-copy genes from the organism using GenBank.
2. BLAST the extracted gene sequences against the database of assembled contigs to identify the contigs containing these five genes.
3. Search the log file *CreateAce.log* to identify the names of the .ace files that contain the identified contigs. Extract the corresponding .ace files from the directory "acedir" and upload them into CLC genomic workbench to calculate the read coverage. The average coverage value of the selected genes is used as the baseline coverage for the genome. The read coverage of false-positive rDNA contigs should be similar to this average coverage value, while the coverage of rDNA-containing contigs should be several fold higher.

3.5.3 Screening the WGA Assembly for rDNA-Containing Contigs and Constructing the rDNA Repeat Sequence

1. The rDNA query sequence for searching the whole genome assembly is selected based on the availability of rDNA sequence from the species in question. If a full or partial rRNA coding region sequence is available then use it as the query sequence. Since the rRNA coding region is highly conserved, a sequence from a related species can also be used. Internal transcribed

spacer (ITS) sequences are available for a large number of species, but their relatively rapid rate of change means they are only suitable for closely-related species.

2. Screen the assembly to identify the rDNA contigs. The basic scheme is illustrated in Fig. 2. Search the assembled contig BLAST database with the rDNA query sequence from above using BLAST. The command for the search is:

```
blastn -query <rdna_contig.fasta>  -task blastn -db assembly.fasta -outfmt
<number_to_define_output_format_type> -num_threads <number_of_threads>
```

Contigs that show matches to the rDNA query sequence are identified based on the BLAST *e*-value.

3. Obtain the fasta sequences (from the assembly file *assembly.fasta*) and .ace files (from the "acedir" directory) of the matching contigs. Calculate the coverage using CLC Genomic workbench, and discard contigs that are <1 kb in length and/or have coverages similar to the unique region of the genome. The remaining contig(s) are potential rDNA-containing contigs. If more than one rDNA contig is identified, select the longest contig for further processing. If the second largest contig is slightly shorter than the largest contig (<1 kb) but has a higher read coverage, use this second contig instead.
4. Open the .ace file for the rDNA contig in Consed. If the ends are of low coverage trim them by right clicking on the first base where low coverage region ends. Select the option "tear contig at this consensus position" and click "Do Tear." This will open two new windows one for low coverage end contig and other has rest of the contig. Discard the low coverage contig and if needed trim the other end of the remaining contig. This trimmed contig is used for downstream analysis.
5. To determine if the rDNA contig can be extended further, BLAST the contig sequence against the assembled contig BLAST database (Fig. 2). If the rDNA contig is incomplete, contigs overlapping with the query sequence will be identified. Discard contigs <1 kb in length and/or with read coverage similar to the unique region of the genome. Also discard contigs that fully overlap the query contig.
6. Repeat this search iteratively using contigs that overlap both ends of the rDNA query sequence (when present) until only contig(s) hits with read coverages similar to that of the unique region of the genome are found. In our experience, usually only one or two iterations are required.
7. Next, merge the identified contigs together to obtain the complete rDNA unit sequence. To achieve this, extract the .ace files for all the rDNA contigs identified. Then combine all the extracted .ace files into one .ace file. This has to be done manually: first, extract the first line of each .ace file and add the value of the fields representing the number of contigs and the

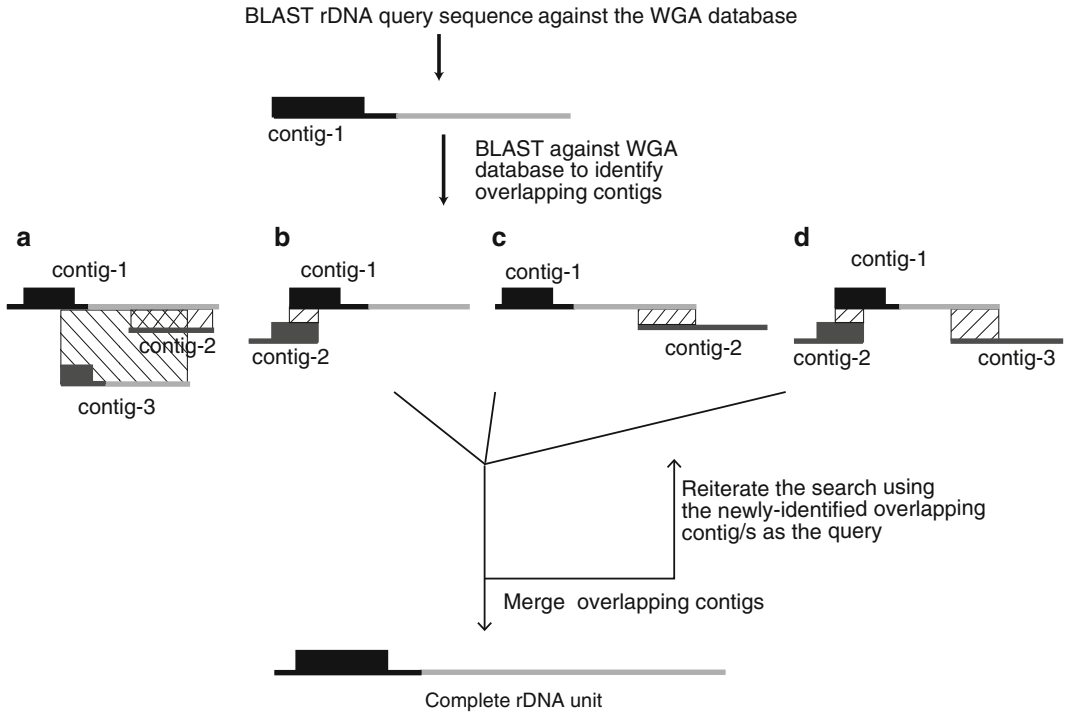

Fig. 2 Flowchart of the strategy to construct the complete rDNA unit sequence. The rDNA query sequence is compared to the contigs obtained from the WGA to identify the primary rDNA-containing contig (contig-1). Contig-1 is then searched against the WGA database to identify additional rDNA-containing contigs. (**a**) If contig-1 has a complete rDNA unit, all other contigs will fully overlap it and nothing further needs to be done. (**b**–**d**) If the rDNA sequence is incomplete, iterations of identifying and merging overlapping contigs are performed until a complete rDNA unit is obtained. See text for more details

number of reads in the file. Make a new header from this information in a new file that will contain information for all the rDNA contigs. For example, if three .ace files have to be concatenated and the values for their first lines are:

```
==> supercontig.ace.1 <==
AS 1 2797
==> supercontig.ace.2 <==
AS 1 138
==> supercontig.ace.3 <==
AS 3 11932
```

Then the first line of the new file *combine.ace* will be:

```
AS 5 14867
```

Next, remove the first line and copy the remaining data from each .ace file into the *combine.ace* file. This can be done using the following commands:

```
sed "/AS 1 2797/d" supercontig.ace.1 >> combine.ace
sed "/AS 1 138/d" supercontig.ace.2 >> combine.ace
sed "/AS 3 11932/d" supercontig.ace.3 >> combine.ace
```

8. To merge the contigs, open the generated *combine.ace* file that contains all the overlapping rDNA contigs in Consed. Click

the "Assembly View" button. This will open the "Assembly View" window followed by a popup "Which Contig to Show in Assembly View" window (if the popup window does not open automatically, go to "What to Show" and click on "In/exclude Contigs"). Increase the depth of coverage cut off value from 50 to an arbitrary value of 1000 to view contigs that have high read coverage, and click "Apply and Restart Assembly View." This will regenerate the "Assembly View" window displaying all the contigs in the .ace file. Then go to "What to Show" and select "Sequence Matches." This will open the "Which Sequence Matches to show in Assembly View" window that has different parameters for the aligner cross_match. Click the "run crossmatch" button to identify the overlapping regions between contigs (which are connected by vertical lines). To merge two contigs, double click the connecting line, opening the "Sequence Matches" window. Select the sequence in the window and click on "Show Alignment." This will open the "Compare Contigs" window. If the second contig is a reverse complement in the "Sequence Matches" window, reverse complement the sequence by clicking the "complement just in this window" button in front of the contig name. Next, click the "Align" button to align the sequences and then click "Join Contigs." This will merge the overlapping reads of both contigs and will generate a new longer merged contig. A new window will open that shows the consensus sequence of the generated contig together with all the reads. Save this new contig in the same .ace file (File → Save assembly). This will replace the contigs that were merged with the new one in the .ace file. Repeat until all overlapping contigs have been merged.

9. By convention, the rDNA sequence is represented with the rRNA coding region transcribed from left to right first, followed by the IGS (Fig. 3a). To demarcate the rRNA coding region and the IGS in the merged rDNA contig, use BLAST with the closest available complete rRNA coding region sequence. Also by convention, the start of the rDNA unit is the transcriptional start site. Therefore, if the information is available, use BLAST to find the homologous transcriptional start site in the new rDNA sequence. Other annotations, such as promoters and terminators, can be added if possible.
10. In some cases, the merged rDNA sequence has more than a single rDNA unit because of duplications at either end (Fig. 3b). Once the rRNA coding region has been identified, such instances are easy to rectify using BLAST to self-compare the merged rDNA sequence, whilst taking care not to remove regions that are actually repetitive within a single rDNA unit.
11. In some cases the rDNA sequence has split the rRNA coding region across the ends of the contig (Fig. 3c). These contigs cannot be extended further, suggesting that the rRNA coding

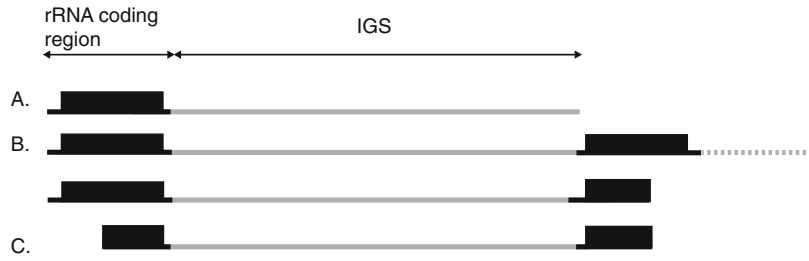

Fig. 3 Potential arrangements of the rRNA coding region and IGS in the merged rDNA contig. The position of the rRNA coding region (*black boxes*) and the IGS (*grey lines*) can vary in the merged rDNA contig. For example, (**a**) complete rRNA coding region followed by the IGS; (**b**) complete rDNA unit followed by a partial rDNA unit. The partial rDNA unit can have a complete rRNA coding region followed by a partial IGS (*dotted grey line*) or just a partial rRNA coding region; (**c**) partial rRNA coding regions on both ends of the contigs

region is complete but segmented. In this case, remove the sequence of one of the partial rRNA coding regions and join it to the other end of the contig sequence such that the 5′-ETS/18S is either at the very start or end of the merged rDNA contig.

12. If the merged rDNA contig has the rRNA coding region in the opposite orientation, make a reverse complement of the sequence to obtain the final rDNA unit sequence.
13. In our experience, this approach is able to generate accurate and complete rDNA repeat sequences. However, to provide further confidence in the sequence obtained, it is possible to verify the rDNA repeat sequences using BAC clones containing rDNA if these are available (for example by screening whole genome BAC libraries). This is particularly desirable if the rDNA sequence contains repeats that are present elsewhere in the genome. There are two levels of verification: verifying the length of the repeat unit to demonstrate that a full-length rDNA sequence has been obtained (*see* **Note 13**); and verifying the sequence with NGS of rDNA-containing BAC clone(s) (*see* **Note 14**).

4 Notes

1. To make this script work behind a network proxy, the following line in the original script:

```
my $res = LWP::UserAgent->new->request($req, sub { print $_[0] });
```

has to be modified to:

```
my $res = LWP::UserAgent->new();
$res->env_proxy();
$res->request($req, sub { print $_[0] });
```

2. WGS data for an organism are sometimes generated as a joint project between two or more sequencing centers. In such cases, the data submitted by one sequencing center may only represent a fraction of the genome, and rDNA-containing contigs will likely be absent from the subsequent WGA. In that case, the assembly has to be performed again using a dataset from another sequencing center. To maximize the chances of using the best dataset, first individually map the data from all the sequencing centers that are involved in a given genome project to the rRNA coding region sequence of the species (or related species) using a Sanger read mapper (e.g., gsMapper). Then use the dataset from the sequencing center with the highest number of rDNA reads.
3. The query_tracedb script command to download the TI number in binary format is:

```
query_tracedb "query page_size 40000 page_number N binary SPECIES_CODE='PAN
TROGLODYTES' AND CENTER_NAME='BI' and TRACE_TYPE_CODE='WGS'"
> page_N.bin
```

where *N* is the page number.

4. The command to download the fasta files is:

```
(echo -n "retrieve fasta 0b"; cat page_N.bin) | ./query_tracedb > data_N.fasta
(echo -n "retrieve quality 0b"; cat page_N.bin) | ./query_tracedb > data_N.qual
(echo -n "retrieve xml 0b"; cat page_N.bin) | ./query_tracedb > data_N.xml
```

where *N* is the page number.

5. It is possible that because of sequencing depth or large genome size, a very large number of WGS reads are present for an organism from a project. Although it is recommended to take all the reads available reads to perform the WGA, it is possible to use just a subset of the WGS reads to speed up the assembly. If computational power is limited or the WGS read number is very high, divide the WGS data into two or three parts and perform WGAs for each of them. Then construct the rDNA sequence using each assembly and compare the rDNA sequences obtained. In theory, all the obtained rDNA sequences should be similar. The reduction of the input data can speed up the process, but it increases the chance that some of the rDNA is misassembled or missed during the assembly.
6. Example of an NCBI entry's fasta file header:
   >gnl|ti|174190180 name:G591P603493FD5.T0 mate: 174194532
   CTAGNNNAAAGATCTGCCTGGGGGATTAGATCTA
   GCTGACAGTCAGGGTGGTGGTGGCC
   ……
7. The following command:

```
sed -e "s/gnl|ti|[A-Za-z0-9]*\s//g" -e "s/mate:[A-Za-z0-9]*//g"
-e "s/name://g" <input_fasta_file>   >>  fasta_file_with_modified_header.fasta
```

will change the header line to:
>G591P603493FD5.T0

C T A G N N N A A A G A T C T G C C T G G G G G A T T A G
ATCTAGCTGACAGTCAGGGTGGTGGTGGCC
……

8. Example of an NCBI entry's quality file header:
>gnl|ti|174190180 name:G591P603493FD5.T0
4 4 4 4 0 0 0 4 6 4 4 6 4 6 8 8 8 8 8 8 8 8 10 8 18 12 9 9 18
18 24 26 25 25 25 26 30 30 30 25 25 23
……
9. The field "type" in the traceinfo file is not part of the Trace archive format and hence its value has to be set in the configuration file *reads_config.xml*. The value of the type field can be "paired_production," "unpaired_production," or "transposon." For Sanger WGS data, the value of the field "type" is usually "paired_production." Further, sometimes the value of the field "insert_stdev" (standard deviation of the insert size in bp) is not defined, in which case 1/10th of the insert size can be used as the standard deviation.
10. The command to modify the XML file to retain only the required information is:

```
grep -e "<trace>" -e "<trace_name>" -e "<plate_id>" -e "<well_id>" -e
"<template_id>" -e "<clip_vector_left>" -e "<clip_vector_right>" -e "<insert_size>" -e
"<insert_stdev>" -e "<trace_end>" -e "<center_name>" -e "<seq_lib_id>" -e
"</trace>"  <input_xml_file>  >>  xml:file_with_filtered_data.xml
```

Furthermore, Arachne requires an XML declaration as the first line of the file, and all the trace information has to be the child of the root element "trace_volume." To fulfill both these requirements, the following two lines should be added at the start of the *reads.xml* file

```
<?xml version="1.0"?>
<trace_volume>
```

and the line below added as the very last line of the *reads.xml* file

```
</trace_volume>
```

11. If the Arachne assembly breaks in the middle of a run, it can be restarted from the point it stopped with the following steps:

Step 1: Perform an Arachne run such that the all modules are run without generating any output using option NOGO = True:

```
/absolute/path/to/Assemblez PRE=/path/to/working/directory DATA=data RUN=run
num_cpus=number_of_cpu num_cpus_pi=10 ACE=True
REMOVE_INTERMEDIATES=False NOGO=True > log.txt
```

The obtained file *log.txt* will have the commands for each step that has to be run during the assembly.

Step 2: The "RUN" directory also has a log file *assemblez.log* that has details of all the commands used during the assembly. Search *assemblez.log* for the last generated output directory:

```
grep "OUTDIR" /path/to/run/assemblez.log | tail -n 1
```

Step 3: Search the module that has generated the last output directory in the *log.txt* file from Step 1:

```
grep "<name of last output directory >" log.txt
```

Step 4: Restart the assembly from the module at which Arachne was broken using the command:

```
/absolute/path/to/Assemblez PRE=/path/to/working/directory DATA=data RUN=run
num_cpus=number_of_cpu num_cpus_pi=10 ACE=True
REMOVE_INTERMEDIATES=False START=<Module_name>
```

where "Module_name" is the module that generated the last output directory.

12. The ACE file format stores the contig information. This file format was initially developed for the visualization and manipulation of contigs using Consed, but was later adopted by other Sanger sequence assemblers. A single .ace file can contain information for more than one contig. The first line of an .ace file shows the number of contigs and total number reads in the file, and subsequent lines have information about the individual contigs. The details of the ACE file format are described in the Consed manual [84].
13. BAC clones, with an average insert size of ~175 kb, are expected to contain at least 4–5 rDNA copies, as the rDNA is up to ~45 kb in length [26, 27]. I-*Ppo*I, a homing enzyme, cuts only at one site in the rDNA repeat (in the 18S rRNA coding region), allowing liberation of single rDNA units from a BAC clone. Therefore, the length of BAC rDNA repeat units can be estimated by gel electrophoresis following I-*Ppo*I digestion. If the rDNA unit length is small (up to ~15 kb), conventional gel electrophoresis using low percent agarose gels can be used to determine the length. However, for larger rDNA units pulsed field gel electrophoretic separations are necessary. We found that Field Inversion Gel Electrophoresis (FIGE) implemented on a CHEF electrophoresis apparatus can accurately determine long rDNA unit lengths. We have found that FIGE gels seem to consistently overestimate the rDNA unit length by ~1 kb. The rDNA unit band can be confirmed by Southern hybridization using an rRNA coding region probe if necessary.
14. A WGA rDNA sequence can be further verified by mapping NGS reads from rDNA-containing BAC clones to the WGA rDNA sequence. This allows detection of relatively small-scale errors in the rDNA sequence. In our experience, some parts of the rRNA coding region are highly biased against in Illumina

sequencing outputs, as discussed earlier. Thus reads from these regions can be absent from the sequencing data for the rDNA BAC clones. When WGA and BAC rDNA sequences are compared, these regions appear as gaps in the BAC rDNA sequences. These gaps should not be interpreted as misassemblies in the WGA rDNA sequences, but as sequencing limitations. Because the rRNA coding regions are well characterized in most eukaryotic lineages, such cases are straightforward to diagnose.

References

1. Yue F, Cheng Y, Breschi A et al (2014) A comparative encyclopedia of DNA elements in the mouse genome. Nature 515:355–364
2. Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
3. Koch CM, Andrews RM, Flicek P et al (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res 17:691–707
4. Pei B, Sisu C, Frankish A et al (2012) The GENCODE pseudogene resource. Genome Biol 13:R51
5. Orom UA, Derrien T, Beringer M et al (2010) Long noncoding RNAs with enhancer-like function in human cells. Cell 143:46–58
6. Djebali S, Davis CA, Merkel A et al (2012) Landscape of transcription in human cells. Nature 489:101–108
7. Nam JW, Bartel DP (2012) Long noncoding RNAs in *C. elegans*. Genome Res 22:2529–2540
8. Brown JB, Boley N, Eisman R et al (2014) Diversity and dynamics of the Drosophila transcriptome. Nature 512:393–399
9. Gerstein MB, Rozowsky J, Yan KK et al (2014) Comparative analysis of the transcriptome across distant species. Nature 512:445–448
10. Scally A, Dutheil JY, Hillier LW et al (2012) Insights into hominid evolution from the gorilla genome sequence. Nature 483:169–175
11. The Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87
12. International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716
13. Naumova N, Imakaev M, Fudenberg G et al (2013) Organization of the mitotic chromosome. Science 342:948–953
14. Ho JW, Jung YL, Liu T et al (2014) Comparative analysis of metazoan chromatin organization. Nature 512:449–452
15. Rao SS, 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
16. Eichler EE, Clark RA, She X (2004) An assessment of the sequence gaps: unfinished business in a finished human genome. Nat Rev Genet 5:345–354
17. Floutsakou I, Agrawal S, Nguyen TT et al (2013) The shared genomic architecture of human nucleolar organizer regions. Genome Res 23:2003–2012
18. Leem SH, Kouprina N, Grimwood J et al (2004) Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. Genome Res 14:239–246
19. Dunham A, Matthews LH, Burton J et al (2004) The DNA sequence and analysis of human chromosome 13. Nature 428:522–528
20. Zody MC, Garber M, Sharpe T et al (2006) Analysis of the DNA sequence and duplication history of human chromosome 15. Nature 440:671–675
21. Hattori M, Fujiyama A, Taylor TD et al (2000) The DNA sequence of human chromosome 21. Nature 405:311–319
22. Heilig R, Eckenberg R, Petit JL et al (2003) The DNA sequence and analysis of human chromosome 14. Nature 421:601–607
23. McClintock B (1934) The relation of a particular chromosomal element to the development of the nucleoli in *Zea mays*. Z Zellforsch Mikrosk Anat 21:294–326
24. Ritossa FM, Spiegelman S (1965) Localization of DNA complementary to ribosomal RNA in the nucleolus organizer region of *Drosophila melanogaster*. Proc Natl Acad Sci U S A 53:737–745

25. Phillips RL, Kleese R, Wang SS (1971) The nucleolus organizer region of maize (*Zea mays* L.): chromosomal site of DNA complementary to ribosomal RNA. Chromosoma 36:79–88
26. Thiry M, Lafontaine DL (2005) Birth of a nucleolus: the evolution of nucleolar compartments. Trends Cell Biol 15:194–199
27. Grozdanov P, Georgiev O, Karagyozov L (2003) Complete sequence of the 45-kb mouse ribosomal DNA repeat: analysis of the intergenic spacer. Genomics 82:637–643
28. van Keulen H, Gutell RR, Campbell SR et al (1992) The nucleotide sequence of the entire ribosomal DNA operon and the structure of the large subunit rRNA of *Giardia muris*. J Mol Evol 35:318–328
29. Spear BB (1980) Isolation and mapping of the rRNA genes in the macronucleus of *Oxytricha fallax*. Chromosoma 77:193–202
30. Birnstiel ML, Chipchase MI, Hyde BB (1963) The nucleolus, a source of ribosomes. Biochim Biophys Acta 76:454–462
31. Kobayashi T, Ganley AR (2005) Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science 309:1581–1584
32. Donati G, Montanaro L, Derenzini M (2012) Ribosome biogenesis and control of cell proliferation: p53 is not alone. Cancer Res 72:1602–1607
33. Derenzini M, Montanaro L, Chilla A et al (2005) Key role of the achievement of an appropriate ribosomal RNA complement for G1-S phase transition in H4-II-E-C3 rat hepatoma cells. J Cell Physiol 202:483–491
34. Deisenroth C, Zhang Y (2010) Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. Oncogene 29:4253–4260
35. Audas TE, Jacob MD, Lee S (2012) Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. Mol Cell 45:147–157
36. Zhang LF, Huynh KD, Lee JT (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell 129:693–706
37. Clos J, Normann A, Ohrlein A et al (1986) The core promoter of mouse rDNA consists of two functionally distinct domains. Nucleic Acids Res 14:7581–7595
38. Doelling JH, Gaudino RJ, Pikaard CS (1993) Functional analysis of Arabidopsis thaliana rRNA gene and spacer promoters *in vivo* and by transient expression. Proc Natl Acad Sci U S A 90:7528–7532
39. Haltiner MM, Smale ST, Tjian R (1986) Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. Mol Cell Biol 6:227–235
40. Pfleiderer C, Smid A, Bartsch I et al (1990) An undecamer DNA sequence directs termination of human ribosomal gene transcription. Nucleic Acids Res 18:4727–4736
41. Grummt I, Maier U, Ohrlein A et al (1985) Transcription of mouse rDNA terminates downstream of the 3′ end of 28S RNA and involves interaction of factors with repeated sequences in the 3′ spacer. Cell 43:801–810
42. Nemeth A, Perez-Fernandez J, Merkl P et al (2013) RNA polymerase I termination: where is the end? Biochim Biophys Acta 1829:306–317
43. Pape LK, Windle JJ, Mougey E et al (1989) The Xenopus ribosomal DNA 60-and 81-base-pair repeats are position-dependent enhancers that function at the establishment of the preinitiation complex: analysis *in vivo* and in an enhancer-responsive *in vitro* system. Mol Cell Biol 9:5093–5104
44. Little RD, Platt TH, Schildkraut CL (1993) Initiation and termination of DNA replication in human rRNA genes. Mol Cell Biol 13:6600–6613
45. Yoon Y, Sanchez JA, Brun C et al (1995) Mapping of replication initiation sites in human ribosomal DNA by nascent-strand abundance analysis. Mol Cell Biol 15:2482–2489
46. Gencheva M, Anachkova B, Russev G (1996) Mapping the sites of initiation of DNA replication in rat and human rRNA genes. J Biol Chem 271:2608–2614
47. Yu GL, Blackburn EH (1990) Amplification of tandemly repeated origin control sequences confers a replication advantage on rDNA replicons in *Tetrahymena thermophila*. Mol Cell Biol 10:2070–2080
48. Daniel DC, Johnson EM (1989) Selective initiation of replication at origin sequences of the rDNA molecule of *Physarum polycephalum* using synchronous plasmodial extracts. Nucleic Acids Res 17:8343–8362
49. Van't Hof J, Hernandez P, Bjerknes CA et al (1987) Location of the replication origin in the 9-kb repeat size class of rDNA in pea (*Pisum sativum*). Plant Mol Biol 9:87–95
50. Botchan PM, Dayton AI (1982) A specific replication origin in the chromosomal rDNA of *Lytechinus variegatus*. Nature 299:453–456
51. Gogel E, Langst G, Grummt I et al (1996) Mapping of replication initiation sites in the mouse ribosomal gene cluster. Chromosoma 104:511–518

52. Muller M, Lucchini R, Sogo JM (2000) Replication of yeast rDNA initiates downstream of transcriptionally active genes. Mol Cell 5:767–777
53. Coffman FD, He M, Diaz M-L et al (2006) Multiple initiation sites within the human ribosomal RNA gene. Cell Cycle 5:1223–1233
54. Coffman FD, Georgoff I, Fresa KL et al (1993) *In vitro* replication of plasmids containing human ribosomal gene sequences: origin localization and dependence on an aprotinin-binding cytosolic protein. Exp Cell Res 209:123–132
55. Akamatsu Y, Kobayashi T (2015) The human RNA polymerase I transcription terminator complex acts as a replication fork barrier that coordinates the progress of replication with rRNA transcription activity. Mol Cell Biol 35:1871–1881
56. Dimitrova DS (2011) DNA replication initiation patterns and spatial dynamics of the human ribosomal RNA gene loci. J Cell Sci 124:2743–2752
57. Brewer BJ, Fangman WL (1988) A replication fork barrier at the 3′ end of yeast ribosomal RNA genes. Cell 56:637–643
58. Lopez-estrano C, Schvartzman JB, Krimer DB et al (1998) Co-localization of polar replication fork barriers and rRNA transcription terminators in mouse rDNA. J Mol Biol 277:249–256
59. Lopez-Estrano C, Schvartzman JB, Krimer DB et al (1999) Characterization of the pea rDNA replication fork barrier: putative cis-acting and trans-acting factors. Plant Mol Biol 40:99–110
60. Wiesendanger B, Lucchini R, Koller T et al (1994) Replication fork barriers in the Xenopus rDNA. Nucleic Acids Res 22:5038–5046
61. Grandori C, Gomez-Roman N, Felton-Edkins ZA et al (2005) c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. Nat Cell Biol 7:311–318
62. Kern SE, Kinzler KW, Bruskin A et al (1991) Identification of p53 as a sequence-specific DNA-binding protein. Science 252:1708–1711
63. Jacob MD, Audas TE, Mullineux ST et al (2012) Where no RNA polymerase has gone before: novel functional transcripts derived from the ribosomal intergenic spacer. Nucleus 3:315–319
64. Bierhoff H, Schmitz K, Maass F et al (2010) Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. Cold Spring Harb Symp Quant Biol 75:357–364
65. Mayer C, Neubert M, Grummt I (2008) The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. EMBO Rep 9:774–780
66. Mayer C, Schmitz KM, Li J et al (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. Mol Cell 22:351–361
67. Saka K, Ide S, Ganley AR et al (2013) Cellular senescence in yeast is regulated by rDNA noncoding transcription. Curr Biol 23:1794–1798
68. Prokopowich CD, Gregory TR, Crease TJ (2003) The correlation between rDNA copy number and genome size in eukaryotes. Genome 46:48–50
69. Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49:727–764
70. Stage DE, Eickbush TH (2007) Sequence variation within the rRNA gene loci of 12 Drosophila species. Genome Res 17:1888–1897
71. Ganley AR, Kobayashi T (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res 17:184–191
72. James SA, O'Kelly MJ, Carter DM et al (2009) Repetitive sequence variation and dynamics in the ribosomal DNA array of *Saccharomyces cerevisiae* as revealed by whole-genome resequencing. Genome Res 19:626–635
73. Gonzalez IL, Sylvester JE (1995) Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer. Genomics 27:320–328
74. Li Z, Chen Y, Mu D et al (2012) Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. Brief Funct Genomics 11:25–37
75. Flicek P, Birney E (2009) Sense from sequence reads: methods for alignment and assembly. Nat Methods 6:S6–S12
76. Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380
77. Myers EW, Sutton GG, Delcher AL et al (2000) A whole-genome assembly of Drosophila. Science 287:2196–2204
78. Green P. Documentation for phrap and cross_match. http://www.phrap.org/phredphrap/phrap.html. Accessed 21 August 2015
79. Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. In: German Conference on Bioinformatics, pp 45–56

80. Jaffe DB, Butler J, Gnerre S et al (2003) Whole-genome sequence assembly for mammalian genomes: Arachne 2. Genome Res 13:91–96
81. Batzoglou S, Jaffe DB, Stanley K et al (2002) ARACHNE: a whole-genome shotgun assembler. Genome Res 12:177–189
82. Burton J (2008) Repeat. https://www.broadinstitute.org/crd/wiki/index.php/Repeat. Accessed 26 August 2015
83. Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8:195–202
84. Gordon D (2015) Consed 29.0 Documentation. http://www.phrap.org/consed/distributions/README.29.0.txt. Accessed 26 August 2015
85. Searching Tips. NCBI. http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?view=search_tips. Accessed 26 August 2015
86. Heiman D, Burton J (2007) XML ancillary files. https://www.broadinstitute.org/crd/wiki/index.php/XML_ancillary_files. Accessed 26 August 2015
87. Burton J, Gnerre S (2006) Directory tree. http://www.broadinstitute.org/crd/wiki/index.php/Directory_tree. Accessed 26 August 2015
88. Heiman D, Burton J, Grabherr MG (2006) Input. http://www.broadinstitute.org/crd/wiki/index.php/Input. Accessed 26 August 2015
89. Heiman D, Burton J (2007) Reads config.xml. http://www.broadinstitute.org/crd/wiki/index.php/Reads_config.xml. Accessed 26 August 2015
90. Burton J (2008) FindXmlFeatures. http://www.broadinstitute.org/crd/wiki/index.php/FindXmlFeatures. Accessed 26 August 2015
91. Burton J, Gnerre S, Grabherr MG (2006) Output. https://www.broadinstitute.org/crd/wiki/index.php/Output. Accessed 26 August 2015

Chapter 14

Analysis of rRNA Gene Methylation in *Arabidopsis thaliana* by CHEF-Conventional 2D Gel Electrophoresis

Gireesha Mohannath and Craig S. Pikaard

Abstract

Contour-clamped homogenous electric field (CHEF) gel electrophoresis, a variant of Pulsed-field gel electrophoresis (PFGE), is a powerful technique for resolving large fragments of DNA (10 kb–9 Mb). CHEF has many applications including the physical mapping of chromosomes, artificial chromosomes, and sub-chromosomal DNA fragments, etc. Here, we describe the use of CHEF and two-dimensional gel electrophoresis to analyze rRNA gene methylation patterns within the two ~4 million base pair nucleolus organizer regions (NORs) of *Arabidopsis thaliana*. The method involves CHEF gel electrophoresis of agarose-embedded DNA following restriction endonuclease digestion to cut the NORs into large but resolvable segments, followed by digestion with methylation-sensitive restriction endonucleases and conventional (or CHEF) gel electrophoresis, in a second dimension. Resulting products are then detected by Southern blotting or PCR analyses capable of discriminating rRNA gene subtypes.

Key words Arabidopsis, 45S rRNA gene, rDNA, Nucleolus organizer region (NOR), DNA methylation, Protoplast, Pulsed-field gel electrophoresis (PFGE), Contour-clamped homogenous electric field (CHEF), Two-dimensional gel analysis, Southern blotting

1 Introduction

Pulsed-field gel electrophoresis (PFGE) is an invaluable technique to resolve large fragments of DNA (10 kb–9 Mb) that cannot be resolved by conventional gel electrophoresis [1, 2]. Contour-clamped homogenous electric field (CHEF) gel electrophoresis is a variation of PFGE that enables excellent resolution of long DNA molecules [3]. The CHEF procedure involves the application of an electric field that alternates between two orientations at a predetermined angle [3, 4]. For the analyses of chromosomes and sub-chromosomal fragments, whole cells are first embedded in an agarose gel matrix and digested with proteases and one or more restriction endonucleases that cut the chromosomes within the agarose matrix, whose dimensions match the wells of an agarose gel [5, 6]. The agarose "plugs" containing the cut DNA are then inserted into the wells and electrophoresis is conducted.

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_14,

CHEF gel electrophoresis conducted either in one or two dimensions, combined with the Southern blotting technique [7–10], has been useful for multiple applications including, Restriction Fragment Length Polymorphism (RFLP) analysis, physical mapping of gene order, sizing and isolation of intact yeast chromosomes or Bacterial Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs), analysis of genomic polymorphism in bacteria called epidemiological typing, etc. [1, 4, 5, 11–13].

45S ribosomal RNA (rRNA) genes in eukaryotes are tandemly repeated at chromosomal loci known as nucleolus organizer regions (NORs) [14–16]. In *Arabidopsis thaliana*, rRNA genes are organized into two NORs: *NOR2* and *NOR4* on chromosomes 2 and 4, respectively, with each NOR consisting of 350–400 gene copies [11, 17]. About half of the rRNA genes are silenced during development by repressive chromatin modifications, including DNA hypermethylation [18–21]. Recent evidence has indicated that rRNA gene silencing in Arabidopsis takes place at a sub-chromosomal level, affecting *NOR2*. A prediction is that DNA fragments derived from *NOR2* are expected to be hypermethylated relative to *NOR4*. This hypothesis can be tested by analyzing rRNA gene methylation within large DNA fragments derived from the two NORs. We describe here a method to perform such studies using a two-dimensional gel electrophoresis approach. The method involves CHEF gel electrophoresis of agarose-embedded DNA following restriction endonuclease digestion with an enzyme that cuts infrequently within one or both NORs, followed by CHEF gel electrophoresis to resolve the fragments in a first dimension. The products resolved in the first dimension are then subjected to in-gel restriction digestion using a methylation-sensitive endonuclease, followed by a conventional (or CHEF) gel electrophoresis in a second dimension. Southern blotting, probing, and detection of resulting products provide insight into the methylation status of large DNA fragments derived from the different NORs.

The method described here can also be adapted to analyze NOR methylation patterns in other species.

2 Materials

All solutions should be made using ultra pure water with 18.2 MΩ·cm (25 °C) resistivity (e.g., Milli-Q water, type 1) and analytical/molecular biology grade reagents. Filter-sterilize the solutions using a 0.22 or 0.45 μm vacuum filtration unit and store them at room temperature (unless indicated otherwise). Follow recommended safety procedures for handling and disposing of chemicals.

2.1 Seed Sterilization and Plant Growth

1. Seed Sterilization Solution: 76% (v/v) ethanol, 20% (v/v) commercial bleach (*see* **Note 1**), 4% (v/v) water.
2. Murashige & Skoog (MS) plates (0.5× strength): For 1 L of MS media, dissolve 2.2 g of MS salts (with vitamins) in water and add sucrose to a final concentration of 1% (w/v) (10 g/L). Adjust pH of the solution to 5.7 with potassium hydroxide (KOH). Add 5 g of phyto agar, autoclave for 20 min and then pour MS media into sterile petri dishes (100 mm × 20 mm) under sterile conditions in a laminar flow hood (*see* **Note 2**).

2.2 Materials Needed to Generate Protoplasts Embedded in Agarose Plugs

1. Protoplasting Solution (prepare fresh prior to use): 0.4 M mannitol, 0.33% (w/v) cellulase (10 kU/g), 0.17% (w/v) macerozyme R-10 (*see* **Note 3**), 3 mM MES Buffer, pH 7.5 (adjust pH with KOH), 7 mM $CaCl_2$, 5 mM glutathione.
2. Filtering devices with the following pore sizes: 1.13 mm, 230 μm, 38 μm (*see* **Note 4**).
3. Protoplast Resuspension Solution: 0.34 M mannitol, 7 mM $CaCl_2$.
4. 50 mM TBE: 50 mM Tris–HCl, 50 mM Borate, 1 mM EDTA.
5. Reusable agarose plug molds (Bio-Rad Cat#1703622).
6. NDS solution: 100 mM EDTA (*see* **Note 5**), 1% (w/v) *N*-laurylsarcosine, 1 mM Tris. For 1 L, dissolve EDTA and Tris in 750 mL of water. To help dissolve EDTA, add solid sodium hydroxide to reach a pH between 8.0 and 9.0. In a separate vessel, dissolve the *N*-laurylsarcosine in 100 mL of water (incubation in a 65 °C water bath for few minutes might be necessary for complete solubilization). Add the sarcosine solution to the EDTA solution, adjust pH to 8.5 with sodium hydroxide or hydrochloric acid and adjust the final volume with water.

2.3 Components for Restriction Digestion of Agarose-Embedded DNA and CHEF Gel Analysis

1. $T_{10}E_{10}$ buffer: 10 mM Tris, 10 mM EDTA, pH 8.0.
2. Stop Solution: 20 mM Tris, 50 mM EDTA, pH 8.0.
3. Certified megabase agarose (Bio-Rad) and low-melting temperature agarose.
4. CHEF gel apparatus.

2.4 Southern Blotting Components

1. Ethidium Bromide Solution: Prepare ethidium bromide solution (1.5 μg/mL) in 50 mM TBE.
2. Depurination Solution: 0.25 M HCl.
3. Denaturation Solution: 3 M NaCl, 0.4 M NaOH.
4. Transfer Buffer: 3 M NaCl, 8 mM NaOH.
5. *GeneScreen Plus* hybridization transfer membrane (PerkinElmer).
6. Neutralization Buffer: 50 mM sodium phosphate, pH 6.5.

7. UV Crosslinker apparatus (e.g., GS UV crosslinker, Bio-Rad)
8. Southern transfer apparatus.
9. Reaction components to generate radioactive random-primed probe: Gel-purified 45S rRNA gene DNA template, random hexamers, New England Biolabs (NEB) Buffer 2, dATP, dTTP, dGTP, [α-^{32}P]-labeled dCTP, Klenow fragment (3′ → 5′ exo-).
10. Church's Hybridization Buffer: 1% (w/v) BSA, 1 mM EDTA, 0.5 M sodium phosphate, pH 7.2, 7% (w/v) SDS.
11. Non-stringent Wash Buffer: 0.2× SSC, 0.1% (w/v) SDS (20× SSC buffer contains 3 M NaCl, 0.3 M sodium citrate, pH 7.0).
12. Stringent Wash Buffer: 0.1× SSC, 0.1% (w/v) SDS.
13. Phosphorimaging screen and cassette.

3 Methods

Protoplasts can be generated from plants grown on soil or on MS plates. However, the quality of DNA appears to be better using cells of plants grown on MS plates. Therefore, the method below involves growing plants on MS media for protoplast generation. To develop this method, we have incorporated previously described procedures with some modifications [5, 22–26].

3.1 Seed Sterilization

1. Add a desired amount of Arabidopsis seeds (~25 μL/plate) to a 1.7 mL microcentrifuge tube and add 1 mL of Seed Sterilization Solution. Mix and incubate the tube at room temperature for 15 min using a rotating mixer (*see* **Note 6**).
2. Spin down the seeds in a microcentrifuge at 13,000 × *g* for 1 min and discard the solution. Wash the seeds twice with 1 mL of 100% (v/v) ethanol.
3. Add 1 mL of 100% (v/v) ethanol to the seeds and spread them using a pipettor on an empty sterile petri dish containing a piece of sterile filter paper, in a laminar flow hood. Allow the seeds to dry.
4. Spread the seeds onto freshly prepared MS plates under sterile conditions (*see* **Note 7**).
5. Seal the plates with Micropore paper tape to prevent desiccation while allowing aeration, and place them at 4 °C for 3 days before transferring them to a growth chamber set for short day growth conditions (8 h light, 24 °C: 16 h dark, 23 °C, 60% relative humidity). Place the plates on a shelf covered with an aluminum foil, surface-sterilized with 70% (v/v) ethanol. Within the growth chamber, the plates are kept about 40 cm away from light source consisting of fluorescent and incandescent bulbs. Let the plants grow for 2 weeks (*see* **Note 8**).

3.2 Generation of Protoplasts

1. After 2 weeks take the MS plates to a hood and gently harvest the plants either by cutting them with sterilized surgical scissors or pulling them out of the media with forceps, without collecting any chunks of solid media (*see* **Note 9**).
2. Combine plant material harvested from two plates on the lid of a sterile petri dish and lightly mince the plants into small pieces using sterile surgical scissors or a razor blade.
3. After mincing, place the plant material in a sterile petri dish (100 mm × 20 mm) and add enough 0.5 M mannitol to cover the plant material. Allow the cells to plasmolyze for 30–60 min.
4. After plasmolysis, remove the mannitol using a pipette (or by gently pouring into a waste container) and add 15 mL of freshly made Protoplasting Solution. Gently mix the plant material with the solution, seal the petri dish with a parafilm tape and cover with aluminum foil. Place the plates on a rotary shaker with gentle agitation (30 rpm) and allow cell wall digestion to proceed overnight (*see* **Note 10**).

   (*From this stage onwards be very gentle with the protoplasts. It is critical to minimize physical trauma to the protoplasts, therefore avoid pipetting until the very last step*).
5. Separate liberated protoplasts from the rest of the digested plant material by gently pouring the digestion mixture, successively, through three different filtering devices with pore sizes of 1.13 mm, 230 μm, and 38 μm (*see* **Note 11**). When using the filtering device with the pore size of 38 μm, it will help to pre-wet the filter from below by placing it onto a drop of water placed on a sterile surface. This overcomes surface tension and removes trapped air bubbles so that the protoplasts in solution can flow through.
6. Gently transfer the final filtrate to a clean 50 mL conical tube placed on ice and centrifuge at 30 × *g* for 10 min in a swinging bucket rotor at room temperature. Remove the supernatant (usually very green) using a sterile transfer pipette without touching the protoplast pellet. Add 13 mL of Protoplast Resuspension Solution and resuspend protoplasts by gently inverting the tube several times. After the resuspension of the protoplasts is complete, gently transfer the solution to a 15 mL conical tube placed on ice.
7. Centrifuge the protoplasts again at 30 × *g* for 10 min, remove the supernatant, and resuspend the protoplasts as before but in 10 mL of Protoplast Resuspension Solution and store them on ice (*see* **Note 12**).
8. Take 10 μL of resuspended protoplasts and dilute them in 90 μL of Protoplast Resuspension Solution and count the protoplasts using a hemocytometer. Adjust the volume to give 2×10^7 protoplasts/mL, a concentration sufficient for analyzing high copy 45S rRNA genes (*see* **Note 13**).

3.3 Preparation of Agarose-Embedded DNA

1. We use reusable plug molds (Bio-Rad) to make agarose plugs. Keep clean, assembled plug molds ready on ice. After resuspending protoplasts to the required concentration, prepare an equal volume of 2% (w/v) low-melting agarose solution in 50 mM TBE and heat in a boiling water bath until the agarose melts (*see* **Note 14**).
2. Equilibrate the temperatures of both the protoplasts and the 2% (w/v) agarose solution by incubating them in a 37 °C water bath for 5–10 min. Then mix the agarose and the protoplast solutions together by gentle but thorough pipetting and immediately pipette the mixture into the plug molds kept on ice (*see* **Note 15**). Allow about 15–20 min for the agarose plugs to solidify.
3. Once the agarose plugs are solidified, gently disassemble the plug molds. Holding the plug-retaining side of the plug mold upside down, gently displace the protoplast-containing agarose plugs (using a small spatula) into a 50 mL conical tube containing 40 mL of NDS with 1 mg/mL protease (source: *Streptomyces griseus*, Sigma, Cat#5147). Seal the tubes with parafilm tape and incubate them overnight in a 50 °C water bath (*see* **Note 16**).
4. Replace the NDS/protease with fresh solution and let it incubate overnight again at 50 °C (*see* **Note 17**).
5. Replace the NDS/protease with fresh NDS but *without* protease and incubate at 50 °C overnight again.
6. The following day replace the NDS with fresh NDS and store at 4 °C (*see* **Note 18**).

3.4 Restriction Digestion of Agarose-Embedded DNA for First Dimension CHEF Gel Electrophoresis

To obtain large rDNA fragments, choose a restriction enzyme which is an infrequent cutter of the 45S rRNA genes and provides a distinct banding pattern. For example, only some rRNA gene subtypes in *A. thaliana* ecotype Col-0 have a *Hind*III site [5].

1. Take two pieces of an agarose plug (one for first dimension analysis and the other for second dimension analysis) slightly smaller than the well of a CHEF gel and incubate in a 15 mL conical tube containing 10 mL of $T_{10}E_{10}$ buffer with 2 mM PMSF for 1 h or more at 4 °C (*see* **Note 19**).
2. Replace the $T_{10}E_{10}$ buffer with fresh 10 mL solution *without* PMSF and equilibrate for 30 min at room temperature. Repeat this step three more times.
3. Place the sample in a flat-bottom 2 mL centrifuge tube containing 1 mL of 1× restriction enzyme buffer and equilibrate for 1 h at room temperature (*see* **Note 20**). Repeat this step one more time by replacing the buffer.
4. Remove the buffer and then add about 200–250 μL (enough to cover the agarose plug) of 1× restriction enzyme buffer

containing 50 units of restriction enzyme and incubate at the appropriate temperature (e.g., 37 °C for *Hind*III digestion) for few hours to overnight (*see* **Note 21**).

5. At the end of the reaction, remove the reaction solution and incubate the plug in 500 μL of Stop Solution for 5–10 min. Then proceed to CHEF gel electrophoresis (*see* **Note 22**).

3.5 CHEF Gel Electrophoresis (First Dimension)

1. About 45 min before the restriction digestion ends, cast a 1% (w/v) agarose gel using Certified Megabase Agarose (Bio-Rad) (or agarose of equivalent quality) in 50 mM TBE. A 1% gel is ideal to resolve 10 kb–1.5 Mb DNA. Use the comb type/size depending on your choice of gel type. You can run either a wide gel or a long gel (Fig. 1). The CHEF gel running conditions given below are for a wide gel. After melting the agarose in TBE, it helps to slightly seal the edges and corners (but not the round holes!) of the gel casting tray with the melted agarose solution before pouring the gel. Once cooled sufficiently, pour the melted agarose solution on to the gel casting tray without introducing air bubbles. Let it cool and solidify for about 40 min. During this time prepare 1% (w/v) low-melting temperature agarose solution (about 4–5 mL) and incubate at 65 °C until it is needed (*see* **Note 23**).
2. Once the gel is dried, carefully remove the comb. Take the digested plugs from the reaction tube and gently insert them into the wells using a small spatula. Two sets of plug samples

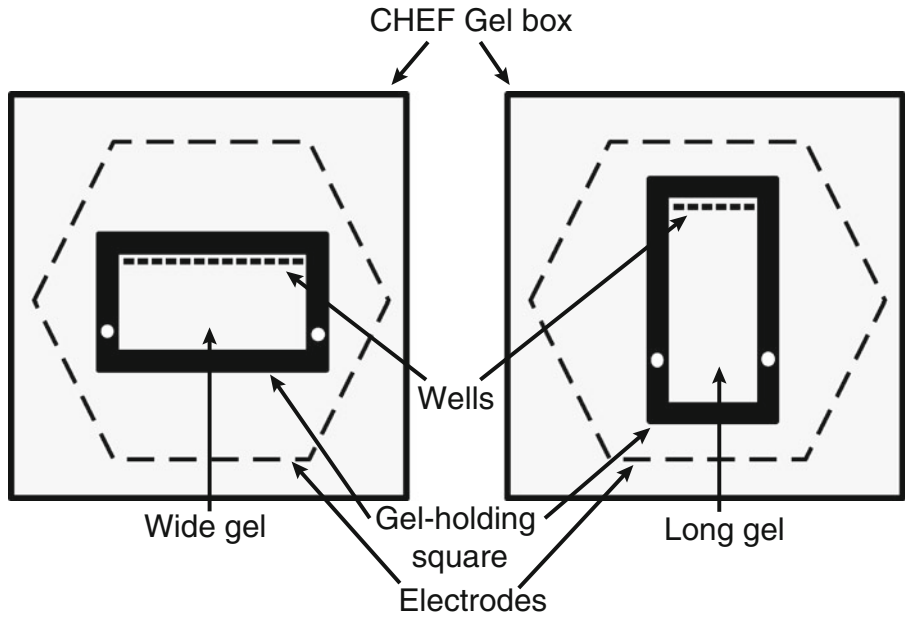

Fig. 1 CHEF gel orientations: Wide gel (on the *left side*) allows more space for samples with less resolving distance wherein long gel (on the *right side*) provides less sample space with longer resolving distance

should be separated on the gel to allow cutting the gel into two halves with each half having one set of samples. In one of the wells, insert a piece of agarose plug containing CHEF gel markers (e.g., Yeast Chromosome PFG marker from NEB which has size range of 225 kb–1900 kb; product # N0345S). Gently press the plugs against the front wall of the wells to remove air bubbles. Seal the wells with the molten 1 % (w/v) low-melting temperature agarose solution without introducing air bubbles and let it solidify for about 10 min (*see* **Note 24**).

3. After the gel is cast, lift the casting tray holding the gel and place it inside the gel-holding square positioned in a CHEF gel apparatus (Fig. 1) containing 2 L of cold 50 mM TBE. We run CHEF gels in a cold room at 4 °C without further cooling (*see* **Note 25**).
4. Load any additional DNA markers as needed (e.g., Kb-plus DNA marker) and run the CHEF gel with the following parameters (*see* **Note 26**). The parameters mentioned in this protocol are for running gels using a CHEF DRII pulsed-field gel electrophoresis system from Bio-Rad. Some modifications in the parameters may be required if you use other systems (e.g., CHEF DRIII or CHEF Mapper system).

   Stage I

   Initial switch time: 60 s

   Final switch time: 60 s

   Run time: 13 h

   Voltage 200 V

   Stage II

   Initial switch time: 90 s

   Final switch time: 90 s

   Run time: 7 h

   Voltage: 200 V

3.6 Southern Analysis of CHEF Gels (First Dimension)

Procedures described for Southern transfer of DNA from CHEF gels onto a membrane are similar to those described for conventional gels except for the requirement of pretreatment of the gel with UV irradiation (*see* **Note 27**).

1. After gel running is complete, carefully transfer the gel to a container and cut into half. Use one half for one-dimensional analysis and save the other half for two-dimensional analysis (*see* **Note 28**).
2. Place one half of the gel to be used for one-dimensional analysis in a plastic box containing fresh Ethidium Bromide Solution and stain for about 40 min by gently agitating the gel on a rotary shaker (*see* **Note 29**).

3. Take a picture of the ethidium bromide-stained gel after UV exposure and proceed to pretreatment (*see* **Note 30**).
4. Irradiate the gel with 60 mJ of 254 nm UV light. We use a Bio-Rad GS Gene Linker UV chamber on setting "nic".
5. Treat the gel with the Depurination Solution for *no longer* than 6–8 min.
6. Wash the gel twice for 30 min each with the Denaturation Solution.
7. Equilibrate the gel with the Transfer Buffer for 15 min.
8. Equilibrate a GeneScreen Plus (PerkinElmer) membrane (cut to the shape of the gel) with the Transfer Buffer at least for 5 min.
9. Transfer DNA from the agarose gel to the membrane for about 24 h by assembling the Southern transfer apparatus (*see* **Note 31**).
10. After the DNA transfer, mark the DNA side of the membrane with a pencil and neutralize the membrane twice, 5 min each, using the Neutralization Solution.
11. Briefly air-dry the membrane and UV cross-link with 150 mJ of 254 nm UV light. We use the C3 setting on a Bio-Rad GS UV Crosslinker for this purpose.
12. Place the membrane in a hybridization bottle containing about 15–25 mL of Church's Hybridization Buffer. Prehybridize the membrane for 1 h or longer at 60 °C (*see* **Note 32**).
13. While the prehybridization of the membrane is occurring, prepare the probe. First, mix the following components and incubate them at 95 °C for 5 min to denature the template DNA.

1 µL (~50 ng)	gel-purified 45S rRNA gene DNA template (*see* **Note 33**)
1 µL	random hexamers (250 ng/µL)
8 µL	water

Add the following components, mix, and incubate for 45–60 min at 37 °C in a radioactive area

2 µL	NEB Buffer 2
3 µL	dNTP solution (1 mM each of dATP, dTTP, and dGTP)
4 µL	α-^{32}P-dCTP (40 µCi)
1 µL	Klenow fragment (3′ → 5′ exo-) (*see* **Note 34**)
20 µL total reaction volume	

14. Bring the volume up to 100 μL with water and run the sample through a PERFORMA spin column (Edge Bio) to remove the unincorporated radioactive nucleotide. Monitor the flow through with a Geiger counter to make sure that it is "hot".
15. Boil the microcentrifuge tube with the probe for 5 min and then add the denatured probe to the bottle containing the pre-hybridized membrane and the hybridization buffer (*see* **Note 35**). Allow probe hybridization to occur overnight at 60 °C.
16. Rinse the membrane twice with Non-stringent Wash Buffer at room temperature and then wash with the same buffer for 10 min at 65 °C.
17. Wash twice with the Stringent Wash Buffer for 15 min each at 65 °C (*see* **Note 36**).
18. Briefly dry the radioactive membrane, wrap in plastic wrap, and expose to a phosphorimager screen for 6 h to overnight. Process the screen using a phosphorimaging scanner to obtain an image of the gel (*see* **Note 37**).

An image of the membrane obtained for agarose-embedded DNA of ecotype (strain) Col-0, subjected to *Hind*III digestion and CHEF gel electrophoresis, is given in Fig. 2. After one-dimensional CHEF gel electrophoresis, gel elutions corresponding to each band on the gel image can be obtained from gel lanes carrying the same sample but not subjected to Southern blot analysis (*see* **Note 38**). Similar images of one-dimensional analysis for other ecotypes of Arabidopsis have been published earlier [5, 11, 12].

3.7 In-Gel Restriction Digestion of Resolved CHEF Gel Samples (for Second Dimension Gel Electrophoresis)

For the second dimension electrophoretic step you can perform either CHEF gel or conventional gel electrophoresis depending on the size of DNA fragments to be resolved. We describe below the use of a conventional gel for the second dimension. For methylation analysis, choose a methylation-sensitive endonuclease (e.g., *Hpa*II).

1. Use the half of the CHEF gel stored at 4 °C for second dimension gel electrophoresis. Slice an individual gel lane containing the sample lengthwise, into two equal halves. Use one half of the lane for restriction digestion with a methylation-sensitive enzyme (e.g., *Hpa*II) and keep the other half for a control (no enzyme) (*see* **Note 39**).
2. Place the gel slice in a plastic container with 100 mL of 1× restriction enzyme buffer and incubate at room temperature for 2 h.
3. Repeat **step 2** and then place the gel slice in a sealable bag containing 10 mL of 1× restriction enzyme buffer and 1000 units of restriction enzyme (or no enzyme for the control sample).
4. Seal the bag, avoiding bubbles, and incubate overnight at the appropriate temperature (e.g., 37 °C for *Hpa*II) (*see* **Note 40**).

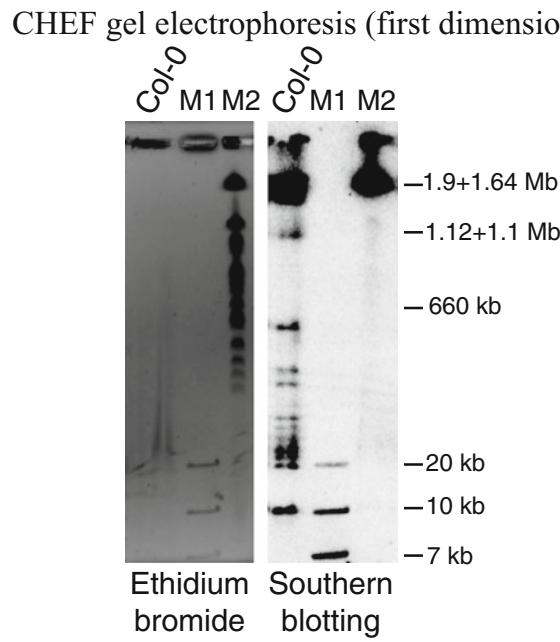

Fig. 2 CHEF gel electrophoresis of *Hind*III-digested agarose-embedded DNA of *Arabidopsis thaliana* ecotype Col-0 (first dimension). On the *left* side is an ethidium bromide-stained gel image. On the *right* side is the image obtained after southern blotting and probing with rRNA gene radioactive probe. M1 = GeneRuler 1 kb Plus DNA ladder (Thermo Scientific #SM1331); M2 = Yeast Chromosome PFG marker (NEB: N0345S). The topmost band in the M2 lane corresponds to yeast (*Saccharomyces cerevisiae*) chromosome 12, which contains ribosomal genes [28] and cross-reacts with the probe. Also, some fragments in M1 cross-react with the probe

For efficient digestion, hang the bag to a wall of a plastic container using glue tape and then place the container in an incubator maintaining the required temperature (Fig. 3).

5. The following morning add 1000 units of restriction enzyme again (except for the control sample) and incubate for an additional 6 h (*see* **Note 41**).
6. After digestion, remove the gel slice from the bag and place it in a tray with 200 mL of 50 mM TBE and incubate at room temperature with gentle agitation for 1 h.
7. Repeat **step 6**. Just prior to the end of the incubation, prepare 1 % (w/v) melted agarose solution in 50 mM TBE. After the wash, lay the gel strip at the top of a gel casting tray perpendicular to the direction of electrophoresis and pour the gel as shown in Fig. 4. Optionally, a small gel comb can be held (with the help of a clamping system) next to the gel slice to create few lanes for loading markers (Fig. 4) (*see* **Note 42**).

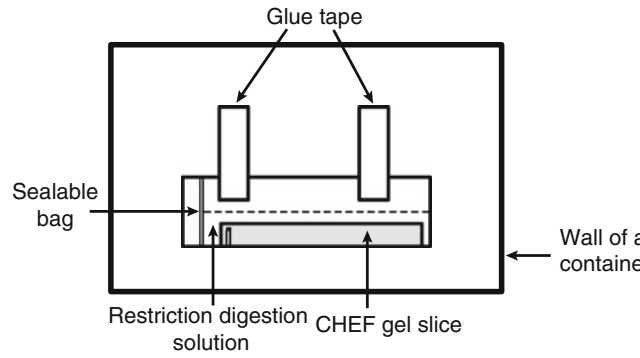

Fig. 3 Setting up a CHEF gel slice for in-gel restriction digestion

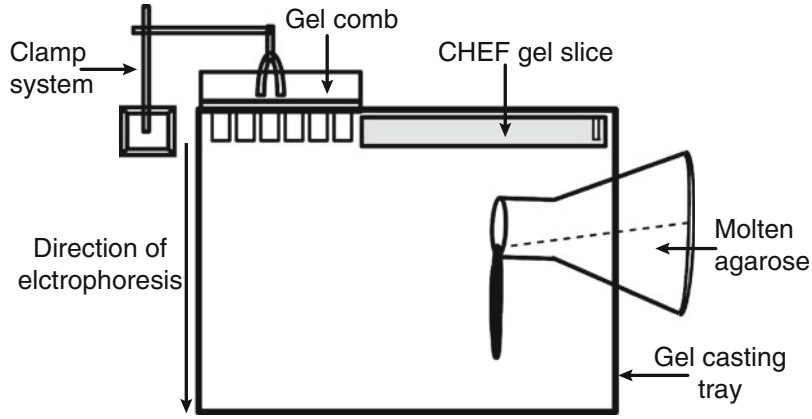

Fig. 4 Casting a new agarose gel with a CHEF gel slice at the top, for the second dimension gel electrophoresis

8. After the gel is hardened, load the appropriate markers and perform standard electrophoresis in 1× TBE at 160 V for 3–4 h (*see* **Note 43**).
9. Perform Southern blotting, probing, and detection as described above (*see* **Note 44**).

A gel image obtained after *Hind*III digestion and CHEF gel electrophoresis in the first dimension and *Hpa*II digestion and conventional gel electrophoresis in the second dimension is shown in Fig. 5. See also [11]. Differential *Hpa*II sensitivities of fragments resolved in the first dimension is apparent in the second dimension (Fig. 5). Note that after two-dimensional gel electrophoresis, DNA fragments can be eluted from different positions of a gel for PCR or other analyses, rather than Southern blot analysis.

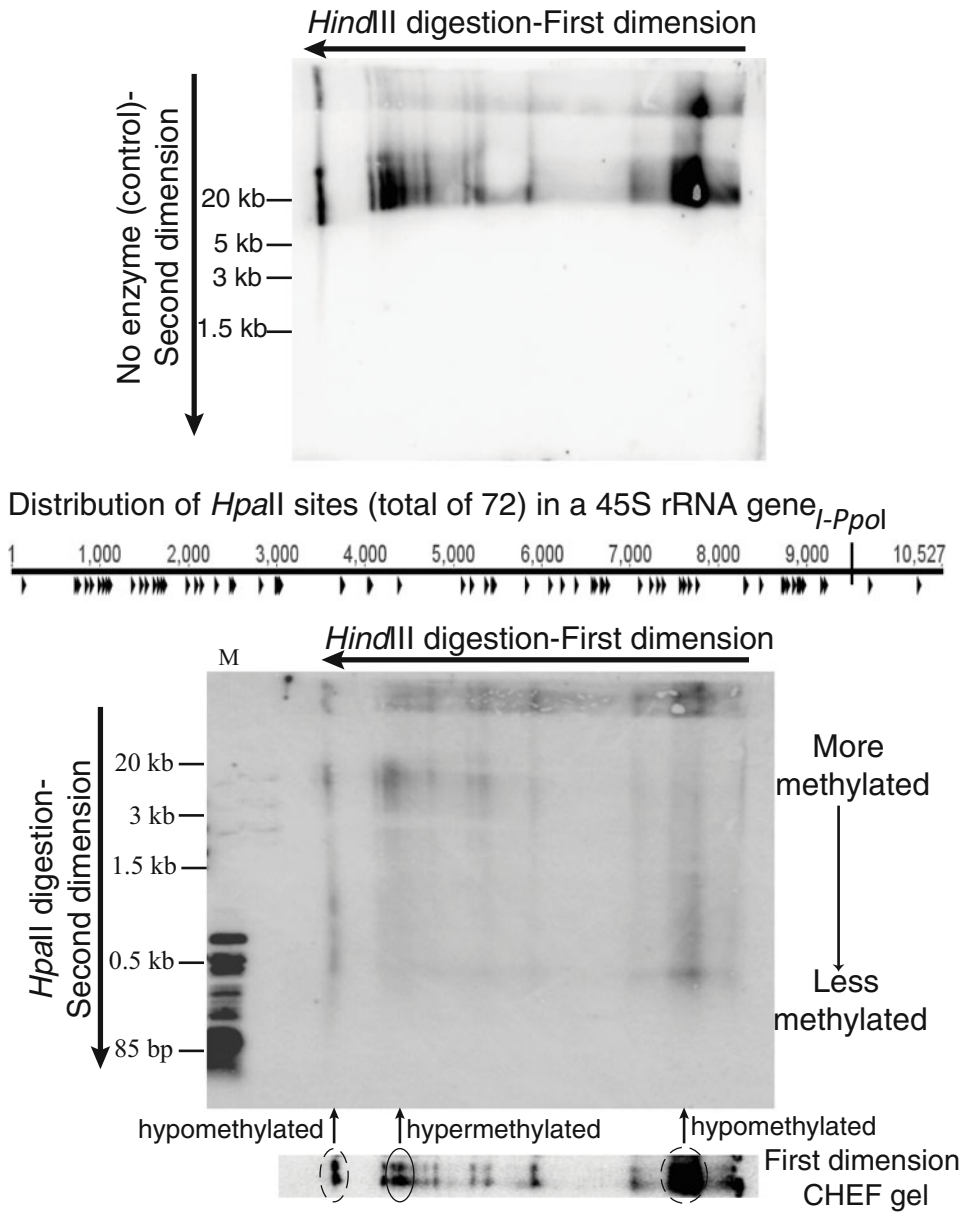

Fig. 5 Two-dimensional gel electrophoresis of agarose-embedded DNA of *Arabidopsis thaliana* ecotype Col-0. CHEF gel electrophoresis of *Hind*III-digested products in the first dimension and conventional gel electrophoresis of non-digested (image at the *top*) or *Hpa*II (methylation-sensitive endonuclease)-digested products (image at the *bottom*) in the second dimension. The DNA fragments marked with *dotted circles* are hypomethylated compared to the DNA fragments marked with *undotted circles*. M = *Hpa*II-digested products of four Arabidopsis rRNA gene-containing BAC clones [27]

4 Notes

1. Bleach made by the Clorox company works fine but bleach from some of the companies can be harsh on seeds and could severely affect Arabidopsis seed germination rate. Ensure the safety of your Seed Sterilization Solution before performing any experiment.
2. Although MS plates can be prepared in advance and stored at 4 °C, we recommend preparing them fresh before use. A drop of 1 M KOH solution is enough to bring the pH of MS solution to 5.7. While pouring and further handling MS plates take care not to contaminate the media! For small plates (100 mm × 20 mm) pour ~25 mL of media and for large plates (150 mm × 25 mm), pour ~120 mL of media. Partially open the lids and allow the plates to dry for about 45 min in a laminar flow hood.
3. Cellulase and macerozyme from different companies may not work equally well. We obtain these enzymes from Biophoretics.
4. Our filter units are constructed by gluing stainless steel mesh disks (cut from sheets supplied by the manufacturer) to the bottoms of 50 mL plastic conical tubes with their bottom cones cut off.
5. Some protocols recommend 500 mM EDTA in NDS solution but we find that 100 mM EDTA is sufficient.
6. Do not plan to analyze more than four genotypes at a time because isolation of protoplasts and their embedding into agarose plugs is a tedious procedure. Plants from two MS plates (150 mm × 25 mm) are enough to generate ten agarose plugs at a concentration of 2×10^7 protoplasts/mL.
7. Take all possible precautions and care to avoid contamination of MS plates. Use only sterile equipment/solutions to spread seeds. There are several ways to place seeds onto MS plates. One of the methods is: place sterilized seeds on a sterile piece of paper (autoclaved). Fold the paper and gently tap it to drop seeds on to MS plates. Alternatively, you can also resuspend the sterilized seeds in a sterile 0.08% (w/v) agarose solution and drop the seeds onto MS plates using a pipettor and a tip.
8. Alternatively, you can store sterilized Arabidopsis seeds in a sterile container at 4 °C for 3 days and then place them on freshly made MS plates. In this case, MS plates will be immediately placed (after sealing with a Micropore tape) in a growth chamber. Do not grow plants for more than 14 days. Plan your experiments accordingly.
9. Discard contaminated plates and use only healthy plants for protoplast generation. It might be a good idea to grow an extra plate of plants (if there is no limitation of seeds) to compensate for the loss of plants, if any, due to contamination.

10. Digestion for longer time (>18 h) may not be desirable and therefore plan the timings for this procedure accordingly.
11. To avoid physical trauma to the protoplasts, slowly pour the protoplast-containing solution along the walls of a filtering device using a clean funnel. The filtrate should be collected into a clean 50 mL beaker placed on ice. While collecting the filtrate, make sure that the filter is as close to the bottom of the beaker as possible. The filters will get clogged frequently by plant debris and therefore, it will be convenient to have multiple clean filtering devices of each pore size to avoid washing and reusing of the filtering devices while performing the protocol.
12. Filtered protoplast solutions will be light green in color and the protoplasts will form a distinct dark green pellet if they are healthy and intact.
13. Most likely, the resuspended protoplast concentration will be less than 2×10^7/mL. If this is the case, then centrifuge the protoplasts again at $30 \times g$ and remove supernatant to the required volume. For example, if the protoplast concentration in the 10 mL solution is 4×10^6/mL, then remove 8 mL of the supernatant after pelleting the protoplasts. This leads to a five-fold increase in the protoplast concentration to 2×10^7/mL.
14. It is important to maintain the concentration of the agarose solution at 2% (w/v). Lower concentration will increase the fragility of the agarose plugs and higher concentration will decrease the porosity of the agarose plugs, which in turn might affect all the downstream steps such as protease digestion, restriction digestion, and release of large DNA fragments from the agarose plugs into the CHEF gel. Alternatively, you can buy 2% agarose solution from Bio-Rad (CleanCut Agarose #1703594). To make agarose plugs, you may use disposable plug molds (Bio-Rad, Cat#1703713) in place of reusable molds.
15. Because the protoplast and the agarose solutions will be viscous, use 1 mL wide orifice tips to mix them. If you don't have the wide orifice tips, cut about 2 mm off of the tips of 1 mL pipette tips, with scissors or a razor blade, before use.
16. While displacing the agarose plugs into NDS solution stay as close to the tube as possible to minimize the impact of dropping. About ten agarose plugs can be placed in 40 mL of NDS per 50 mL conical tube. Do not use dry bead baths in place of water baths. At least in our hands, bead baths did not work well. In a 50 °C bead bath, the agarose plugs melted due to higher temperatures at the bottom of the bead bath.

    We use protease from *Streptomyces griseus* (Sigma Cat#5147) which is cheaper than Proteinase K. NDS with protease sufficient for two successive incubations can be prepared and stored at 4 °C.
17. While discarding NDS solution, be gentle and use a spatula to prevent agarose plugs from falling out of the tube. Alternatively,

you can use screened plug caps from Bio-Rad (Cat#170-3711), or make your own, to discard the solution safely without losing the agarose plugs. While adding fresh NDS solution, take care to ensure the safety of the agarose plugs.

During incubation, gently invert the tubes once or twice to avoid tight stacking of the agarose plugs, which could hamper the protease digestion of the embedded protoplasts.

18. After protease digestion, agarose plugs should have lost all green color. However, some plugs may still have a greenish tinge, which is fine. DNA should now be free from all histones and other proteins due to the protease digestion. A high concentration of EDTA (100 mM) will prevent contaminating DNase activity, if any. Agarose plugs can be stored at 4 °C for several months. Because agarose plugs are fragile, repeatedly opening and closing the storage tube might physically damage the plugs. To avoid this, and for convenience of handling, aliquot plugs into flat 6 well culture plates for storage.

19. The well size of a CHEF gel varies depending on your choice of comb. Measure the size of a well from the comb you intend to use and cut the agarose plug accordingly to fit the well.

    The first equilibration period in $T_{10}E_{10}$ buffer can be extended to overnight, if needed. It is a bit hard to see the plug clearly in the solution because it is transparent, therefore take care that you don't lose the plug while discarding the buffer during $T_{10}E_{10}$ washes.

20. We buy most restriction enzymes from NEB. All NEB restriction enzyme buffer compositions are available at the NEB website.

21. The amount of enzyme and incubation time needed may have to be empirically defined.

    Some restriction enzymes require incubation at higher temperatures. Because the DNA is embedded in low-melting temperature agarose, the incubation temperatures should not exceed 55 °C. For restriction enzymes that require higher temperature (e.g., *BstE*II requires incubation at 60 °C) NEB sells high fidelity versions of the same enzyme that sometimes require lower incubation temperature (*BstE*II-HF requires incubation at 37 °C).

22. If you intend to run the CHEF gel a day later, remove the Stop Solution and store the digested agarose plug at 4 °C. If you store the plug along with the solution, shorter DNA molecules may diffuse from the plug into the solution.

23. Don't add ethidium bromide while casting the gel. We use a Bio-Rad wide-long gel casting stand (Cat#170-3704). For 1.5–3.0 Mb separations, use 0.8% gel.

24. Pieces of agarose plugs often stick to the walls of microcentrifuge tubes. Using a spatula to draw the sample out is not recom-

mended as this will likely damage the plug. Rather, gently invert the tube along with the buffer few times until the plug is suspended in the buffer. Then open the cap of the tube to pour the buffer along with the plug onto a lid of a clean petri dish (use the inner surface). Take care not to damage plugs while inserting into the wells. CHEF gel markers of different sizes are available from NEB and Bio-Rad.

25. Under 4 °C cold room conditions, the mean gel running temperature is ~14 °C. If you don't have cold room facilities to run CHEF gels at 4 °C, use a chiller system with the cooling module in conjunction with CHEF gel apparatus to maintain the gel running temperature of 14 °C.
26. While running a CHEF gel, the TBE buffer is constantly circulated using a pump. Make sure that the speed setting is not too high as this may result in damage and/or displacement of the agarose gel. We use a speed setting of 35 with the Variable Speed Pump from Bio-Rad.

    If you need to vary the CHEF gel running parameters, first try varying the gel running times and if results are not still satisfactory, try varying switch times.
27. For efficient blot transfer of DNA fragments larger than 20 kb, the CHEF gel is treated with UV irradiation, which results in cleavage of DNA molecules. Depurination of the gel with acid treatment also results in DNA cleavage. Although either of these gel pretreatments is enough for DNA cleavage, we use both. If you intend to use only one of these two, then go for UV irradiation.
28. Remember that restriction digestion has been performed on two identical sets of agarose plugs (one set for one-dimensional analysis and the other set for two-dimensional analysis). Therefore, it is recommended to use one whole agarose plug (cut into required pieces) for both one-dimensional and two-dimensional analysis for consistent results.

    It is a good idea to first complete one-dimensional analysis before starting two-dimensional analysis. Half of the gel required for two-dimensional analysis can be saved for few days at 4 °C in a clean container with a lid. Cover the gel with a couple of clean paper towels, soaked with 50 mM TBE, to prevent desiccation.
29. Ethidium bromide staining of DNA is necessary for DNA nicking by UV light and therefore, a fresh staining solution is ideal.
30. To see if the ethidium bromide staining is sufficient or not, check the DNA marker staining. Digestion products from the agarose plugs will likely form a smear.
31. We use a downward transfer apparatus (Schleicher and Schuell) for DNA transfer to blots. You may use upward transfer, if preferred.

To make sure that DNA transfer is complete, stain the gel after transfer and take a picture after UV exposure, using the same exposure settings used prior to DNA transfer. Check the DNA markers on the gel to determine the efficiency of DNA transfer. If need be, the duration of the DNA transfer step can be extended up to 48 h.

32. Alternatively, the blot membrane can be wrapped in a plastic wrap and stored between −20 and 4 °C for later use.
33. We digest a 45S rDNA-containing BAC clone (e.g., F1J10) [27] with I-*P*poI endonuclease (which cuts once per gene) and gel-purify the 10.5 kb rDNA fragment (corresponding to a single copy of 45S rRNA gene) to be used as DNA template for making probes. Alternatively, you can use gel-purified PCR amplicons of 45S rDNA.
34. Klenow fragment ($3' \rightarrow 5'$ exo-) has DNA polymerase activity but not $3' \rightarrow 5'$ exonuclease activity.
35. It is important to remove some hybridization buffer just before adding the probe. The amount of hybridization buffer in the bottle should be just enough to soak the membrane. Too much hybridization buffer will dilute the probe and reduce hybridization efficiency.
36. The amount of wash buffer should be just enough to immerse the membrane. Excess wash buffer and longer wash times may result in weaker signals during detection.
37. Monitor the membrane with a Geiger counter before and after each wash step to check the levels of radiation. The membrane should remain fairly "hot" even after all the washes but not nearly as "hot" as after hybridization.
38. For eluting the bands from a CHEF gel, first measure the distance of each band from the well on the gel image obtained after Southern blot analysis. According to these measurements cut the bands from a gel lane containing the same sample, but not subjected to Southern blot analysis. If you plan to use a kit to extract DNA from gel slices, choose a kit designed to extract longer than 10 kb DNA fragments (e.g., QIAEX II from Qiagen, Cat#20021).
39. Once you standardize this protocol in your hands, you can straightaway proceed to the second dimension analysis after the first dimension analysis without storing the CHEF gel at 4 °C. In this case, however, you will not be able to see the results of the first dimension analysis before proceeding to the second dimension analysis.
40. You can use zip lock bags for incubating sliced gel lanes. We use bags with a dimension of 6 in × 2 in. Using bigger bags will increase the amount of buffer and restriction enzyme required for efficient digestion.

41. You can vary the amount of restriction enzyme and incubation time as needed. To further increase the efficiency of digestion, you can also incubate the bag (secured to a container) (Fig. 3) in a 37 °C shaker set for gentle shaking (not more than 25 rpm!). Overshaking may result in the loss of some small DNA fragments from the gel slice into the buffer solution.
42. The width of the gel casting tray should allow for a few lanes for loading markers. The length of the gel can be varied to obtain a desirable resolution. We use a large gel casting tray (10 in × 9 in) and a wall comb (a flat comb with no wells) to make gels of desirable size. If you make a gel bigger than you need you can cut the unwanted portion using a razor blade either before or after gel electrophoresis.

    Instead of using a comb held by a clamping system, you can also gently cut out wells in an area next to the CHEF gel slice to load markers.
43. The duration of electrophoresis will vary depending on the gel electrophoresis system and the dimension of the gel. Tracking the loading dye front relative to markers will help you decide when to stop the electrophoresis.
44. To detect bands of interest, long exposures to a phosphorimager screen may be needed (up to 5 days). If sensitivity is an issue, you can increase the concentration of protoplasts in the agarose plugs. Additionally, you can also try generating higher specific activity radioactive probes.

References

1. Cox EC, Vocke CD, Walter S, Gregg KY, Bain ES (1990) Electrophoretic karyotype for Dictyostelium discoideum. Proc Natl Acad Sci U S A 87(21):8247–8251
2. Schwartz DC, Cantor CR (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37(1):67–75
3. Chu G, Vollrath D, Davis RW (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science 234(4783):1582–1585
4. O'Brien FG, Udo EE, Grubb WB (2006) Contour-clamped homogeneous electric field electrophoresis of Staphylococcus aureus. Nat Protoc 1(6):3028–3033. doi:10.1038/nprot.2006.382
5. Copenhaver GP, Doelling JH, Gens S, Pikaard CS (1995) Use of RFLPs larger than 100 kbp to map the position and internal organization of the nucleolus organizer region on chromosome 2 in Arabidopsis thaliana. Plant J 7(2):273–286
6. Ryan AJ, Squires S, Strutt HL, Johnson RT (1991) Camptothecin cytotoxicity in mammalian cells is associated with the induction of persistent double strand breaks in replicating DNA. Nucleic Acids Res 19(12):3295–3300
7. Bautsch W (1992) Bacterial genome mapping by two-dimensional pulsed-field gel electrophoresis (2D-PFGE). In: Burmeister M, Ulanovsky L (eds) Methods in molecular biology: pulsed-field gel electrophoresis, vol 12. Humana Press Inc., Totowa, NJ, pp 185–201

8. Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98(3):503–517
9. Southern EM (1992) Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. Biotechnology 24:122–139
10. Warburton PE, Wevrick R, Mahtani MM, Willard HF (1992) Pulsed-field and two dimensional gel electrophoresis of long arrays of tandemly repeated DNA. In: Burmeister M, Ulanovsky L (eds) Methods in molecular biology: pulsed-field gel electrophoresis, vol 12. Humana Press Inc., Totowa, NJ, pp 299–317
11. Copenhaver GP, Pikaard CS (1996) Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in Arabidopsis thaliana, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. Plant J 9(2):273–282
12. Copenhaver GP, Pikaard CS (1996) RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of Arabidopsis thaliana adjoin the telomeres on chromosomes 2 and 4. Plant J 9(2):259–272
13. Goering RV (1993) Molecular epidemiology of nosocomial infection: analysis of chromosomal restriction fragment patterns by pulsed-field gel electrophoresis. Infect Control Hosp Epidemiol 14(10):595–600
14. Phillips RL, Kleese RA, Wang SS (1971) The nucleolus organizer region of maize (Zea mays L.): chromosomal site of DNA complementary to ribosomal RNA. Chromosoma 36(1):79–88. doi:10.1007/BF00326423
15. Raska I, Shaw PJ, Cmarko D (2006) Structure and function of the nucleolus in the spotlight. Curr Opin Cell Biol 18(3):325–334. doi:10.1016/j.ceb.2006.04.008
16. Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49:727–764. doi:10.1146/annurev.bi.49.070180.003455
17. Albini SM (1994) A karyotype of the Arabidopsis thaliana genome derived from synaptonemal complex analysis at prophase I of meiosis. Plant J 5(5):665–672. doi:10.1111/j.1365-313X.1994.00665.x
18. Earley KW, Pontvianne F, Wierzbicki AT, Blevins T, Tucker S, Costa-Nunes P, Pontes O, Pikaard CS (2010) Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. Genes Dev 24(11):1119–1132. doi:10.1101/gad.1914110
19. Lawrence RJ, Pikaard CS (2004) Chromatin turn ons and turn offs of ribosomal RNA genes. Cell Cycle 3(7):880–883
20. Pontvianne F, Abou-Ellail M, Douet J, Comella P, Matia I, Chandrasekhara C, Debures A, Blevins T, Cooke R, Medina FJ, Tourmente S, Pikaard CS, Saez-Vasquez J (2010) Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in Arabidopsis thaliana. PLoS Genet 6(11):e1001225. doi:10.1371/journal.pgen.1001225
21. Pontvianne F, Blevins T, Chandrasekhara C, Mozgova I, Hassel C, Pontes OM, Tucker S, Mokros P, Muchova V, Fajkus J, Pikaard CS (2013) Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. Genes Dev 27(14):1545–1550. doi:10.1101/gad.221648.113
22. Doelling JH, Gaudino RJ, Pikaard CS (1993) Functional analysis of Arabidopsis thaliana rRNA gene and spacer promoters in vivo and by transient expression. Proc Natl Acad Sci U S A 90(16):7528–7532
23. Doelling JH, Pikaard CS (1993) Transient expression in Arabidopsis thaliana protoplasts derived from rapidly established cell suspension cultures. Plant Cell Rep 12(5):241–244. doi:10.1007/bf00237127
24. Bio-Rad L (1992) CHEF DRIII pulsed field electrophoresis systems: instructions manual and applications guide. Bio-Rad Laboratories, Hercules, CA
25. Southern E (2006) Southern blotting. Nat Protoc 1(2):518–525. doi:10.1038/nprot.2006.73
26. Van Daelen R, Zabel P (1992) Pulsed-field gel electrophoresis. In: Beckman JS, Osborn TC (eds) Plant genomes. Kluwer Academic Publishers, Dordrecht, pp 71–100
27. Mozo T, Fischer S, Shizuya H, Altmann T (1998) Construction and characterization of the IGF Arabidopsis BAC library. Mol Gen Genet 258(5):562–570
28. Machin F, Torres-Rosell J, Jarmuz A, Aragon L (2005) Spindle-independent condensation-mediated segregation of yeast ribosomal DNA in late anaphase. J Cell Biol 168(2):209–219. doi:10.1083/jcb.200408087

Chapter 15

Fluorescence-Activated Nucleolus Sorting in Arabidopsis

Frédéric Pontvianne, Myriam Boyer-Clavel, and Julio Sáez-Vásquez

Abstract

Nucleolar isolation allows exhaustive characterization of the nucleolar content. Centrifugation-based protocols are not adapted to isolation of nucleoli directly from a plant tissue because of copurification of cellular debris. We describe here a method that allows the purification of nucleoli using fluorescent-activated cell sorting from *Arabidopsis thaliana* leaves. This approach requires the expression of a specific nucleolar protein such as fibrillarin fused to green fluorescent protein *in planta*.

Key words Nucleolus, *Arabidopsis thaliana*, Fibrillarin, FACS

1 Introduction

Fluorescent-activated cell sorting (FACS) allows the isolation of specific cells from a heterogeneous mixture of biological samples using its particular fluorescence characteristics. Therefore, a specific type of cell can be labeled with a fluorescent marker either because it expresses, such a marker, or because it possesses an epitope that specifically reacts with an antibody coupled with a fluorescent marker [1]. Cell volumes usually range from 50 to 2000 μm^3, but most cell sorters are able to isolate other biological materials like bacteria or unicellular algae that can be as small as 1 μm^3 [2]. Eukaryotic cells contain several organelles or cell compartments that are embedded in the cytoplasm. Mitochondria and chloroplasts, which are bacteria-derived organelles, have been analyzed by FACS in the past [3, 4]. Chlorophyll fluorescence from chloroplasts is currently used to isolate protoplast from cell suspensions by FACS [5, 6]. However, because other tissues or cell debris autofluorescence, an additional fluorescent marker can be used [7].

Plant nuclei can also be isolated using a DNA marker like DAPI or via the expression of a nuclear protein fused to a fluorescent protein [6, 8]. In fact, virtually any compartment or very large aggregate from eukaryotic cells can be isolated using FACS as long as they meet the following requirements:

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_15,

1. Sufficient size.
2. Capacity to remain intact from the rest of the cell content after the separation protocol.
3. Specific fluorescent characteristics.

For example, FACS allows the sorting of chromosomes and provides an excellent tool for genomic studies, especially to assemble very large genomes [9].

The nucleolus, which represents by far the largest nuclear compartment, is a direct consequence of ribosome biogenesis, but is also a plurifunctional body implicated in other mechanisms including stress sensing, cell cycle progression, viral replication, and RNP biogenesis [10–12]. Usually, protocols to purify nucleoli are based on the isolation of nuclei, then nucleoli, by successive centrifugation steps on sucrose gradients and variable salt concentrations. Although this type of purification allows a significant enrichment of nucleoli, traces of nucleoplasm, and nuclear debris remain [13]. In plants, this approach is even more complicated because of the presence of numerous chloroplasts in certain cell-types. One way to decrease the debris is to use plant protoplasts. In 2005, the first nucleolar proteome of Arabidopsis was performed on protoplast cells [14]. We describe here a new strategy that allows nucleolar isolation directly from an entire organism. This approach, named FANoS for fluorescent-activated nucleolus sorting uses a fluorescent nucleolar marker to sort nucleoli by FACS [15, 16].

2 Materials

Stock solution can be prepared in advance, autoclaved and stored at room temperature for several months.

1. 1 M Tris–HCl pH 7.5: weigh 60.5 g of Tris EDTA-free powder, add water to 450 mL, mix and adjust the pH with concentrated HCl prior to make up the volume to 500 mL.
2. 0.5 M NaEDTA pH 8: weigh 93.1 g of Na EDTA powder, add water to 450 mL, mix and adjust the pH with concentrated NaOH prior to make up the volume to 500 mL.
3. 5 M NaCl: weigh 146.1 g of NaCl powder, add water to 500 mL and mix.
4. 1 M $MgCl_2$: weigh 47.6 g of $MgCl_2$ powder, add water up to 500 mL and mix.
5. 1 M MOPS (3-(*N*-morpholino)propanesulfonic acid) pH 7.5: weigh 104.5 g of MOPS powder, add water up to 450 mL and adjust the pH with concentrated NaOH prior to make up the volume to 500 mL.
6. 1 M Sodium Citrate: weigh 129 g of sodium citrate ($Na_3C_6H_5O_7$), add water up to 500 mL and mix.

7. 10× Phosphate buffer saline (PBS): 1370 mM sodium chloride (NaCl), 27 mM potassium chloride (KCl), 100 mM disodium hydrogen phosphate (Na_2HPO_4), and 18 mM potassium dihydrogen phosphate (KH_2PO_4).
8. Wash Solution: 10 mM Tris–HCl pH 7.5, 10 mM Na EDTA pH 8, 100 mM NaCl. Has to be freshly prepared.
9. FACS Solution (or Galbraith's buffer): 45 mM $MgCl_2$, 20 mM MOPS, 30 mM sodium citrate, 0.1% Triton X-100 adjusted to pH 7 with NaOH. Can be prepared in advance and stored at −20 °C for several months.
10. Bioruptor® Standard (diagenode, Belgium) sonication device with a 1.5 mL tube adapter.
11. PARTEC CellTrics (SysmexPartec GmbH, Germany).
12. DAPI: 1 mg/mL of stock solution that is diluted 1000× for nuclear DAPI staining in the homogenate.
13. A flow cytometer/cell sorter: FACS is carried out by flow cytometry with a cell sorter. The protocol described here uses an FACSAria I, (BD Biosciences, USA), but any cell sorter able to sort small particles and with sufficient laser power can be used. YFP or GFP is excited with a 488 nm blue laser of 100 mW and signal emission is captured with a 505 long pass filter and a 530/30 (515–545 nm) filter with the photomultiplier tube (PMT) set at 495. DAPI is excited with a 20 mW 355 nm UV laser at a PMT of 544. Typically, settings of the PMT voltages are 360 for side scatter (SSC) and 357 for forward scatter (FSC) light, respectively. Nozzle size is 70 μm, frequency is 60 kHz and pressure 37 PSI. The sorting procedure is carried out in sterile conditions and the sheath fluid is 1× PBS. A refrigerated chamber keeps the sample at 4 °C and is constantly agitated at 300 rpm. The Aria software is BD FACSDiva version 6.1.2.
14. RNase A is kept as a 10 mg/mL stock and is used to at a concentration of 10 μg/mL.
15. Proteinase K is kept as a 10 mg/mL stock and is used at a concentration of 10 μg/mL.

3 Methods

First, nucleoli have to be labeled via a fluorescent marker that will delimit the nucleolar position directly in the cell of interest *in planta*. A nucleolus-specific protein is required here. Fibrillarin (or NOP1 in *Saccharomyces cerevisiae*) is a nucleolar protein implicated in ribosomal RNA maturation [17, 18]. It has already been widely used in Eukaryotes to label the nucleolus and is therefore a good candidate. The fibrillarin 2 (FIB2) encoding gene from *Arabidopsis thaliana* (At4G25630) was successfully used to sort nucleoli by FACS (Fig. 1a) [15]. In that case, we expressed AtFIB2 CDS under

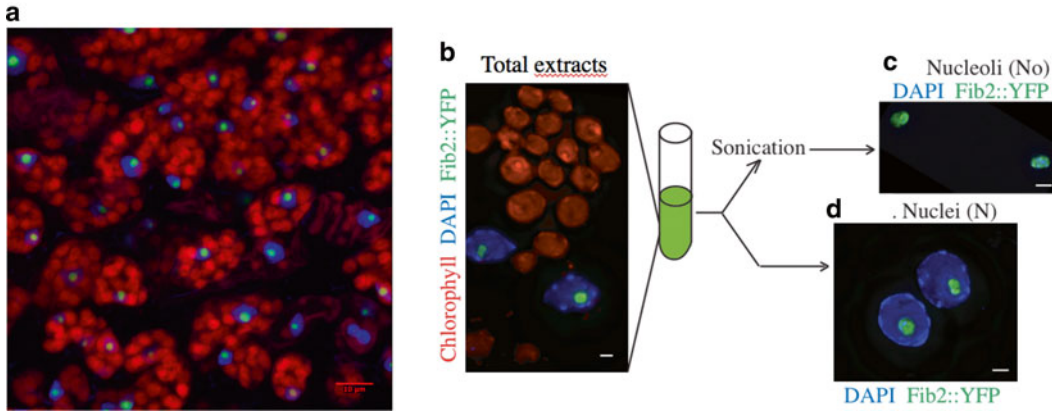

Fig. 1 Fluorescent-activated nucleolus sorting (FANoS) protocol. (**a**) Fixed plant leaf of a plant stably expressing the nucleolar marker FIB2:YFP (*green*) subjected to fluorescence microscopy. Nuclei are stained with DAPI (*blue*) and chloroplasts fluoresce red. (**b**) Leaf cell filtered homogenate subjected to fluorescence microscopy where chloroplasts fluoresce *red*, DAPI-stained DNA is *blue*, and nucleolar FIB2:YFP is *green*. (**c**) Purified nucleoli obtained by FANoS. (**d**) Purified nuclei obtained by fluorescence-activated nuclei sorting (FANS)

the nearly ubiquitous promoter 35S with the Yellow Fluorescent Protein at its C-terminal part. Note that cell-specific nucleolar isolation can be carried out here by replacing the 35S promoter with a cell-specific promoter. *A. thaliana* plants stably expressing this nucleolar marker were then used to prepare samples.

The size of the nucleoli varies depending on the cell-type. For example, some root cell-types possess a large nucleolus that can represent up to 25% of the nuclear volume, while it usually represents 5–10% of this volume in leaf cells. Fully elongated leaves from 3- to 4-week-old plants are well adapted to FANoS because their cells contain large nuclei. Note that the ploidy level in the leaf varies from 2C to 32C depending on the position of the cell in the leaf. A ploidy gradient is indeed observed from the petiole, where cells display a lower ploidy level, to the leaf tip, where this level is higher. Ploidy does not affect the number of nucleoli, but the size of the nucleolus is proportional to the level of ploidy of the cell. Therefore, 16C or 32C cell-types contain larger nucleoli compared to 2C or 4C cell-types.

Here we describe the protocol to isolate nucleoli from at least 3-week-old leaf tissues. Plants which do or do not express the nucleolar marker (YFP:FIB2 plants) have to be prepared at this stage. Wild-type plants are required later as a negative fluorescent control to perform the FACS settings. All buffers are freshly prepared and kept on ice until use. It is also important to keep the samples on ice between each step of the following procedure.

3.1 Fixation

To avoid any loss of the nucleolus content during the nucleolar isolation procedure, the biological samples are fixed with a

formaldehyde solution (*see* **Note 1**). Around 1 g of fully elongated leaves from 4-week-old plants are carefully placed in a 15 mL falcon tube and then fixed 20 min in 10 mL of 4% formaldehyde in Wash Solution. Incubation is carried out on a rotator wheel. The samples are then washed twice 10 min in 10 mL of ice-cold Wash Solution.

3.2 Homogenate Preparation

To prepare the homogenate that will be used for sorting, leaves are minced with a razor blade (*see* **Note 2**). Following the fixation step, 0.5 g of leaves are quickly dried on a paper towel and placed on a plastic weigh boat with 1.5 mL of FACS solution. A new carbon steel razor blade is used to mince the sample around 45 s. The obtained homogenate is filtered through a 30 μm mesh diameter nylon membrane (PARTEC CellTrics) on ice in a 2 mL Eppendorf tube. 0.5 mL of FACS solution is added on the plastic weigh boat to collect extra samples and added to the filter. Typically, 2 min of gravity filtration is sufficient, although a quick manual spin will increase sample recovery.

3.3 Sonication

At this stage, the homogenate can be used directly to sort nuclei as described in the following paragraph. However, to sort nucleoli, a sonication step is required to open the nuclear membrane, shear the DNA, and release the nucleoli from the nuclei (*see* **Note 3**). 700 μL aliquots of filtered homogenate are prepared in 1.5 mL tubes and subjected to sonication. This protocol was optimized for a sonicator bath like the Bioruptor (Diagenode), although other types of sonicator can be used here. For optimal nucleolar isolation, samples are sonicated three times for 5 min (30 s ON/30 s OFF) at medium power. Note that the samples have to be maintained at 4 °C during the procedure. If the sonicator is not equipped with a water-cooling system, place the sample 2 min on ice between each of the three runs and cool down the water bath with ice.

3.4 Sorting

Sorting is carried out with a FACS. The procedure described here is adapted for a FACSAria (BD Biosciences—Special Order Research Product) but could be realized with other types of cell sorter as soon as they meet the technical requirement that include 1–2 μm^3 size particle sorting. The specification and the setting of the machine are described in Subheading 2. Because the filtered homogenate (sonicated or not) contains various autofluorescent particles, it is important to set the area of sorting using a control that does not contain the fluorescent marker (*see* **Notes 4** and **5**).

3.4.1 Nuclear Sorting

Nuclear sorting can be performed using the fluorescent signal of YFP:FIB2. The control sample (without a nucleolar marker) is used to determine the area where no YFP signal is detected from the particle cloud (Fig. 2a, b). Although the nucleolus only represents 5–15% of nuclear volume, the fluorescent signal is sufficient

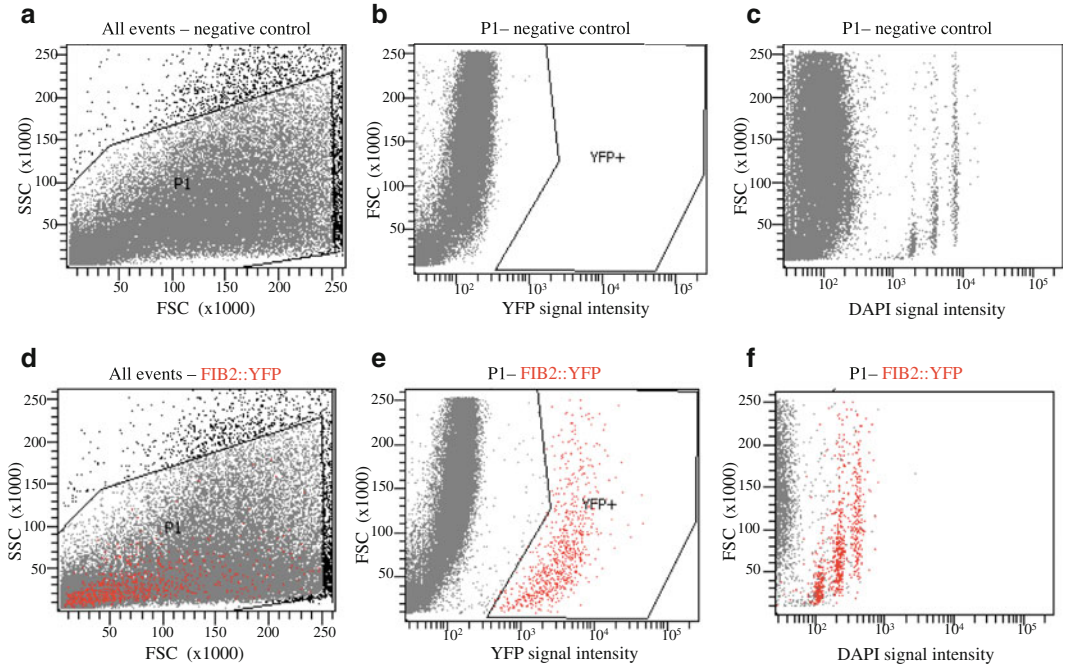

Fig. 2 Fluorescence-activated nucleus sorting (FANS). Analyzed event pattern with the negative control (**a**–**c**) or with the YFP:FIB2 homogenate (**d**, **e**). (**a**, **d**) Biparametric FSC/SSC dot plot with all events. The P1 window represents the events analyzed with a biparametric FSC/YFP (**b**, **e**) or FSC/DAPI (**c**, **f**) dot plot

to isolate most of the nuclei (Fig. 2). DAPI-stained nuclei from leaf cells are usually divided into 3–4 groups that correspond to different ploidy levels (2C, 4C, 8C, and 16C), as shown in the control plant stained with DAPI (Fig. 2c). In the extract of plants carrying the nucleolar marker, more than 95% of the DAPI-stained nuclei also display a significant YFP signal (Fig. 2f). This approach has the advantage of only sorting nuclei with YFP:FIB2 signals and is especially recommended when the nucleolar marker is not ubiquitously expressed.

3.4.2 *Nucleolar Sorting*

Nucleolar sorting is performed using the YFP:FIB2 signal (*see* **Notes 6** and 7). As for nuclear sorting, the negative control sample (without a nucleolar marker) is used to determine the area where no YFP signal is detected from the particle cloud (Fig. 3a, b). In the sample with the YFP:FIB2 marker, a clear shift is observed in the P1 area that corresponds to nucleoli (Fig. 3f) (*see* **Note 8**). To demonstrate that only positive YFP:FIB2 signals are collected by the FACS, we can analyze our run again, the FACS-isolated nucleoli from Fig. 3f. In that case, more than 95% of signals correspond to isolated nucleoli (Fig. 3g–i). Note that the 5% particles present in the reanalyzed sample and that form a small cloud next to the graph's origin are also present when milliQ water is subjected to sorting.

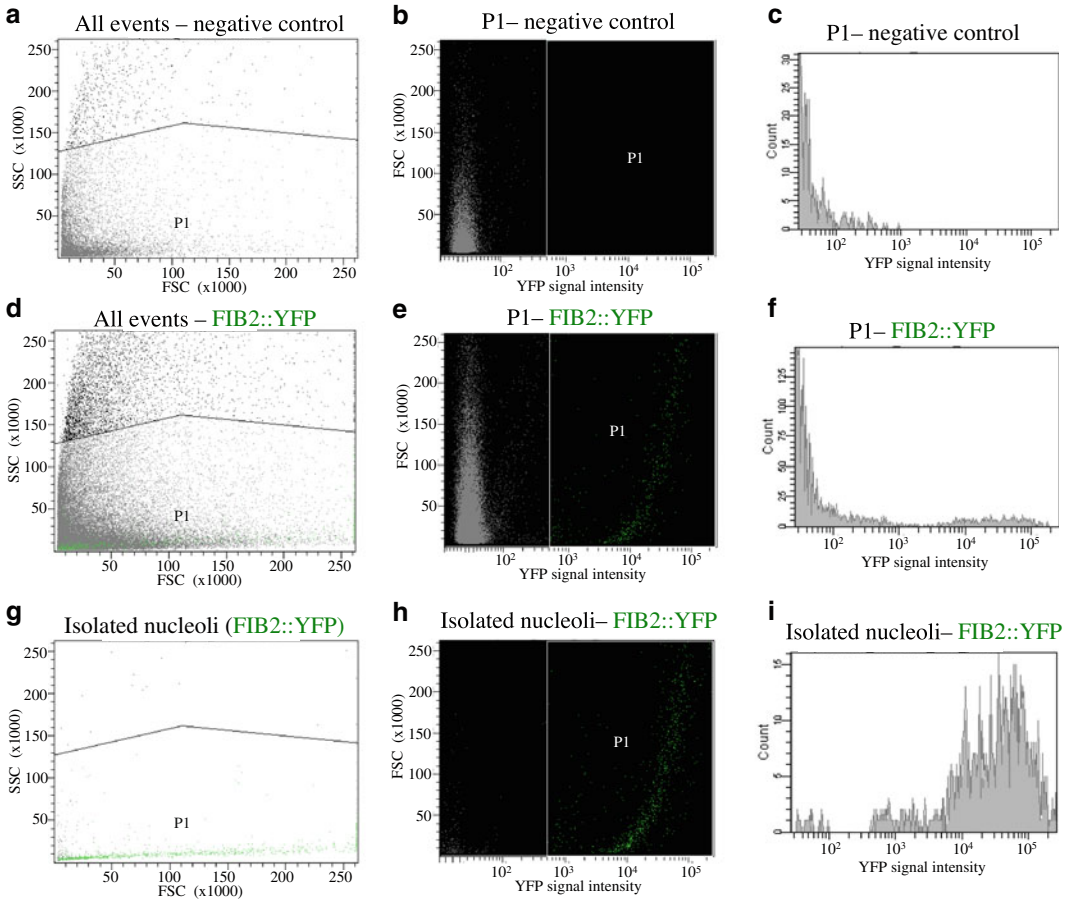

Fig. 3 Fluorescence-activated nucleolus sorting (FANoS). Analyzed event pattern with the negative control (**a**–**c**), with the YFP:FIB2 homogenate (**d**, **e**), and with FANoS-isolated nucleoli (**g**, **i**). (**a**, **d**, **g**) Biparametric SSC/FSC dot plot with all events. The P1 window represents the events analyzed with a biparametric FSC/YFP dot plot (**b**, **e**, **h**). (**c**, **f**, **i**) Diagram showing the distribution of the events depending on their YFP intensities

3.5 Nucleoli Concentration

When sorted by FACS, nucleoli are isolated in a drop that corresponds to a volume of 3.88 nL. Therefore, 500,000 nucleoli correspond to nearly 2 mL. To concentrate the nucleoli in a smaller volume, samples are centrifuged at 2000×*g* for 20 min at 4 °C. After centrifugation, up to 1.8 mL of the supernatant is carefully withdrawn to only keep the nucleoli in about 200 μL.

3.6 Nucleolar Content Analyses

The DNA, RNA, or protein content of FACS-isolated nuclei or nucleoli can now be analyzed:

– For DNA analyses, samples are placed 10 min at 95 °C to liberate the nucleoli content, digested with RNAse A 30 min at 37 °C, then with proteinase K for 30 min at 45 °C and finally boiled 10 min at 99 °C to inactivate the proteinase K. At this stage, DNA can be used directly for PCR or quantitative-PCR

analyses. For high throughput DNA sequencing, the sample should first be cleaned and concentrated using a commercial kit like ChIP DNA Clean and Concentrator (Zymo Research).

- For RNA analyses, samples can be directly subjected to Trizol (Life Technology) extraction. However, note that all buffers used to prepare the filtered homogenate have to be RNAse free, and RNAse inhibitor should always be added to the sample.
- For protein analyses, plant protease inhibitors have to be added to the FACS solution immediately after the sonication step. Because nucleoli are mostly composed of protein, no RNase or DNase treatments are necessary prior to mass-spectrometry analyses. For Western blot analyses, Laemmli buffer can be directly added to the sample.

4 Notes

1. Due to the high density of the nucleolus, this fixation step is not obligatory to obtain nucleolar isolation by FACS, but it certainly stabilizes its content.
2. Optimal homogenate preparation is obtained by mincing no more than the equivalent of 0.5 g of leaves.
3. The filtered homogenate is divided into two in order to obtain nuclei and nucleoli from the same sample and avoid bias during manipulation.
4. To facilitate the setting of the sorting window, it is better to first analyze the sample containing the fluorescent marker, then the negative control.
5. For optimum sorting, it is preferable to adjust the particle flow through the capillary around 10,000 events per second. This setting can be changed by adjusting the flux (number of mL/s) or by diluting the sample.
6. The nucleolar fraction corresponds to approximately 1% of the sample.
7. The size of the nucleoli detected also varies because they tend to aggregate with each other once released from the nucleus.
8. With the settings described in Subheading 3, intensity of the fluorescence is usually between 10^4 and 10^5.

Acknowledgments

This work was supported by the CNRS, the UPVD foundation BQR 2014 FANoS and the ANR NucleoReg ANR-15-CE12-0013-01 to FP, and ANR SubCellif to JSV. The cell sorter is an instrument of the Montpellier Rio Imaging platform. The authors thank Richard Cooke for English corrections.

References

1. Galbraith DW, Anderson MT, Herzenberg LA (1999) Flow cytometric analysis and FACS sorting of cells based on GFP accumulation. Methods Cell Biol 58:315–341
2. Troussellier M, Courties C, Vaquer A (1993) Recent applications of flow cytometry in aquatic microbial ecology. Biol Cell 78(1-2):111–121
3. Galbraith DW, Harkins KR, Jefferson RA (1988) Flow cytometric characterization of the chlorophyll contents and size distributions of plant protoplasts. Cytometry 9(1):75–83
4. Cottet-Rousselle C, Ronot X, Leverve X, Mayol JF (2011) Cytometric assessment of mitochondria using fluorescent probes. Cytometry A 79(6):405–425
5. Bargmann BO, Birnbaum KD. (2010) Fluorescence activated cell sorting of plant protoplasts. J Vis Exp (36)
6. Galbraith DW, Janda J, Lambert GM (2011) Multiparametric analysis, sorting, and transcriptional profiling of plant protoplasts and nuclei according to cell type. Methods Mol Biol 699:407–429
7. You MK, Lim SH, Kim MJ, Jeong YS, Lee MG, Ha SH (2014) Improvement of the fluorescence intensity during a flow cytometric analysis for rice protoplasts by localization of a green fluorescent protein into chloroplasts. Int J Mol Sci 16(1):788–804
8. Zhang C, Barthelson RA, Lambert GM, Galbraith DW (2008) Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. Plant Physiol 147(1):30–40
9. Vrána J, Simková H, Kubaláková M, Číhalíková J, Doležel J (2012) Flow cytometric chromosome sorting in plants: the next generation. Methods 57(3):331–337
10. Boisvert FM, van Koningsbruggen S, Navascués J, Lamond AI (2007) The multifunctional nucleolus. Nat Rev Mol Cell Biol 8:574–585
11. Pederson T (2010) "Compact" nuclear domains: reconsidering the nucleolus. Nucleus 1:444–445
12. Boulon S, Westman BJ, Hutten S, Boisvert FM, Lamond AI (2010) The nucleolus under stress. Mol Cell 40(2):216–227
13. Hacot S, Coute Y, Belin S, Albaret MA, Mertani HC, Sanchez JC, Rosa-Calatrava M, Diaz JJ (2010) Isolation of nucleoli. Curr Protoc Cell Biol. Chapter 3: Unit 3.36
14. Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JW, Shaw PJ (2005) Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol Biol Cell 16(1):260–269
15. Pontvianne F, Blevins T, Chandrasekhara C, Mozgová I, Hassel C, Pontes OM, Tucker S, Mokros P, Muchová V, Fajkus J, Pikaard CS (2013) Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. Genes Dev 27(14):1545–1550
16. Durut N, Abou-Ellail M, Pontvianne F, Das S, Kojima H, Ukai S, de Bures A, Comella P, Nidelet S, Rialle S, Merret R, Echeverria M, Bouvet P, Nakamura K, Sáez-Vásquez J (2014) A duplicated NUCLEOLIN gene with antagonistic activity is required for chromatin organization of silent 45S rDNA in Arabidopsis. Plant Cell 26(3):1330–1344
17. Barneche F, Steinmetz F, Echeverría M (2000) Fibrillarin genes encode both a conserved nucleolar protein and a novel small nucleolar RNA involved in ribosomal RNA methylation in *Arabidopsis thaliana*. J Biol Chem 275(35):27212–27220
18. Sáez-Vasquez J, Caparros-Ruiz D, Barneche F, Echeverría M (2004) A plant snoRNP complex containing snoRNAs, fibrillarin, and nucleolin-like proteins is competent for both rRNA gene binding and pre-rRNA processing *in vitro*. Mol Cell Biol 24(16):7284–7297

Chapter 16

Purification of RNA Polymerase I-Associated Chromatin from Yeast Cells

Astrid Bruckmann, Jan Linnemann, and Jorge Perez-Fernandez

Abstract

The native template of all eukaryotic nuclear RNA polymerases is chromatin. To understand how transcription occurs in vivo, it is important to define the chromatin environment of transcribing RNA Pols. Here, we describe a method used to characterize the distribution and the protein environment of RNA Pol I on ribosomal DNA during transcription in the yeast *S. cerevisiae*. The method is based on conventional chromatin immunoprecipitation and we propose quality control analyses at different steps of the procedure. Finally, the obtained samples are a useful source for downstream analyses by semiquantitative mass spectrometry or quantitative PCR.

Key words rRNA, rDNA, RNA Polymerase I, Transcription, Affinity purification, Yeast, *Saccharomyces cerevisiae*, Chromatin purification, Crosslink, ChIP, Proteome-ChIP

1 Introduction

The nucleolus is a nuclear subdomain in eukaryotic cells where the ribosome synthesis takes place. Ribosome biogenesis starts with the transcription of the ribosomal DNA (rDNA) by the RNA polymerases (RNA Pol) I and the co-transcriptional association of biogenesis factors to the nascent rRNA. Although the composition of different ribosomal precursor particles is well defined, little is known about the intricate protein–protein interactions occurring during Pol I mediated transcription.

Biochemical purification of protein complexes coupled to mass spectrometry is the state-of-the-art to identify interacting partners. Nevertheless, the characterization of DNA-protein interactions shows several methodological problems during the affinity purification procedure. Major complications include the insolubility of the DNA–protein macromolecular complexes, the liability of the DNA-protein interactions, and the dual presence of proteins in DNA-bound and -unbound fractions.

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_16,

Several methods have been described for the unbiased characterization of chromatin from affinity purification of DNA–protein complexes. These affinity purifications rely either on the use of an oligonucleotide to specifically capture a given DNA locus [1] or the use of a DNA-binding protein which recognizes sequences previously inserted at the locus of interest [2]. In this regard, we have previously reported a method to purify chromatin domains released from chromosomes of *Saccharomyces cerevisiae* [3]. The chromatin purified by this method preserves many characteristic features of the chromatin at the endogenous *locus* [4]. Although these methods are suitable to define the molecular composition of chromatin at defined endogenous chromosomal *loci*, they do not discriminate between different functional states of the purified chromatin.

The development of methods based on chromatin immunoprecipitation (ChIP) may circumvent the described drawbacks. On one side, the use of in vivo formaldehyde cross-linking fixes even large multisubunit protein complexes to their respective positions along the DNA and probably preserves transient interactions. Thus, these methods allow harsh purification conditions to solubilize chromatin. On the other side, in ChIP-derived methods the purification of a tagged bait protein of choice, which associates specifically with a particular functional state of a chromosomal locus, is used to copurify distinct chromatin populations.

We describe a method based on affinity purification of cross-linked material to obtain rDNA fractions enriched in RNA Pol I. The method involves a gentle cross-linking with formaldehyde, the solubilization of DNA–protein complexes by sonication, a mild washing step to discard proteins not bound to the DNA and the affinity purification of protein–DNA complexes. By introducing slight variations in the protocol, the purified chromatin can be used to characterize the distribution of Pol I along the rDNA or to analyze the proteome associated to Pol I with semiquantitative mass spectrometry approaches. Although the affinity purification described here makes use of a bait protein tagged with a protein A epitope [5], the protocol can certainly be adapted to be used with differently tagged proteins. In this regard, we used this method to describe the interaction of proteins with DNA-associated Pol I and Pol II, and found specific interactions of groups of proteins with either Pol I or Pol II [6]. Moreover, the associated proteome in the Pol I purification strongly overlaps the proteome copurified with the native 35S rDNA domain by other methods [4].

2 Materials

2.1 Coupling of Rabbit IgG Antibodies to Epoxy-Activated Magnetic Beads

1. BcMag™ Epoxy-Activated Magnetic Beads (Bioclone Inc.).
2. Rabbit IgG antibodies.
3. 50% v/v acetone.
4. 0.1 M sodium phosphate buffer pH 7.4.
5. 3 M ammonium sulfate in 0.1 M sodium phosphate buffer pH 7.4.
6. Phosphate Buffered Saline (PBS): 0.02 M sodium phosphate, 0.15 M NaCl, pH 7.4.
7. PBS pH 7.4 with 0.5% w/v Triton X-100.
8. 100 mM glycine-HCl pH 2.5.
9. 10 mM Tris–HCl pH 8.8.
10. 0.1 M triethylamine solution (should be prepared fresh).
11. 10% w/v sodium azide.
12. BcMag Separator (Bioclone Inc.).

2.2 Cell Cultures and Cross-Linking

1. Y2423. Yeast strain encoding the second largest subunit of Pol I (Rpa135) with C-terminal Protein A tag. Genotype: *his3-1*, *leu2-0*, *met15-0*, *ura3-0 RPA135*::*RPA135-TEV-PROTA*::*KANMX6* [6].
2. YP medium (1% w/v yeast extract, 2% w/v Peptone) containing either 2% w/v glucose (YPD) or 2% w/v galactose (YPG) (*see* **Notes 1** and **2**).
3. Spectrophotometer to follow cell growth by measuring the absorbance of cell cultures at 600 nm.
4. 37% w/v formaldehyde (stabilized in 10–15% v/v methanol, *see* **Note 3**).
5. Tris/glycine solution (10 mM Tris–HCl pH 8.0, 2.5 M glycine).
6. PBS.

2.3 Whole Cell Extracts Preparation

1. Buffer PQ: 50 mM HEPES-KOH pH 7.5, 200 mM NaCl, 0.5 mM EDTA, 1% v/v Triton, 0.1% v/v sodium deoxycholate, 0.1% v/v SDS. Prepare fresh and keep at 4 °C until the end of the experiment (*see* **Note 4**).
2. Buffer DQ1: 50 mM HEPES-KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% v/v Triton, 0.1% v/v sodium deoxycholate, 0.1% v/v SDS. Prepare fresh and keep at 4 °C until the end of the experiment (*see* **Note 4**)
3. Protease inhibitor cocktail (*pic*) is a 100× concentrated ethanol solution containing 200 mM benzamidine and 100 mM PMSF (*see* **Note 5**).
4. Glass beads (diameter 0.75–1 mm) stored at 4 °C.

5. Vibrax-VXR (IKA), orbital shaker.
6. PARAFILM® M
7. Branson Sonifier.

2.4 Affinity Chromatography

1. Magnetic beads coupled to IgG (*see* Subheading 2.1 and **Note 6**).
2. BcMag Separator (Bioclone Inc.).
3. Rotation wheel.
4. Buffer AC: 0.1 M ammonium acetate and 0.1 mM $MgCl_2$ (*see* **Note 7**).
5. 0.5 M ammonium hydroxide.
6. Buffer DQ2: 10 mM Tris–HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% v/v NP-40, 0.5% v/v sodium deoxycholate (*see* **Notes 4** and **8**).
7. TE buffer: 50 mM Tris–HCl pH 8.0, 10 mM EDTA pH 8.0 (*see* **Notes 4** and **8**).
8. Buffer DQE: 50 mM Tris–HCl pH 8.0, 10 mM EDTA, 1% SDS (*see* **Notes 4** and **8**).
9. Thermomixer.

2.5 Quality Control

1. 6× Laemmli loading buffer. 12% w/v SDS, 0.06% bromophenol blue, 47% v/v glycerol, 12% v/v 0.5 M Tris–HCl pH 6.8, 9.3% w/v DTT.
2. RNase A.
3. Glycogen.
4. Proteinase K.
5. Phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v).
6. 100% v/v ethanol.
7. 70% v/v ethanol.
8. 10 mM Tris–HCl pH 8
9. anti-PAP antibody (Sigma #P1291) for western blot analysis.

3 Methods

3.1 Coupling of Rabbit IgGs to Epoxy-Activated Magnetic Beads

1. 300 mg of epoxy-activated beads are suspended in 10 mL of 50% acetone (300 mg corresponds to ~5.1×10^{10} beads) in a 50 mL polypropylene conical tube under vigorous shaking on a Vortex mixer (*see* **Note 6**).
2. Capture beads at the side of the tube using the BcMag Separator and remove the supernatant.
3. Wash beads with 20 mL of 0.1 M sodium phosphate buffer. Capture the beads and remove the supernatant as described in **step 2**. Repeat the wash step twice more.

4. Suspend the beads in 16 mL of 0.1 M sodium phosphate buffer and incubate with gentle rotation for 5 min at room temperature (*see* **Note 9**).
5. Dissolve 100 mg of rabbit IgGs in 7 mL of water (final concentration approx. 14 mg/mL).
6. Clarify the IgG suspension by centrifugation at 13,000×*g* for 10 min at 4 °C. Transfer 3.5 mL of the supernatant (50 mg of rabbit IgG) in a new 50 mL conical tube.
7. Dilute the IgG solution with 9.85 mL of 0.1 M sodium phosphate buffer and add dropwise 6.65 mL of 3 M ammonium sulfate in 0.1 M sodium phosphate pH 7.4 under gentle mixing (*see* **Note 10**).
8. Centrifuge the IgG solution at 820×*g* for 3 min at 4 °C and the supernatant is added to the magnetic bead suspension (*see* **step 4**).
9. Incubate the sample overnight, at least 18 h, at 30 °C with gentle rotation.
10. Capture the beads and remove the supernatant as described in **step 2**.
11. Wash the beads with 20 mL of 100 mM glycine-HCl to remove IgGs, which have not been covalently attached to the matrix (*see* **step 2** and **Note 11**).
12. Wash the beads once more with 20 mL of 10 mM Tris–HCl pH 8.8 and remove the supernatant as described in **step 2**.
13. Add 20 mL of 0.1 M triethylamine solution and incubate for 5–10 min under gentle rotation in order to inactivate residual reactive epoxy groups. The supernatant is removed as described in **step 2**.
14. Wash the beads with 20 mL of PBS pH 7.4 for 5 min with gentle rotation and remove the supernatant as described in **step 2**. Repeat the washing step three times more.
15. Wash the beads with 20 mL of PBS pH 7.4, 0.5% w/v Triton X-100 for 5 min under gentle rotation. Remove the supernatant as described in **step 2** and repeat the wash step but with 15 min incubation. Remove the supernatant as described in **step 2**.
16. Suspend the beads in a final volume of 16 mL PBS pH 7.4, 0.02% w/v sodium azide and store as 1 mL aliquots at 4 °C until use.

3.2 Cell Culture, Cross-Linking, and Sonication

1. Inoculate 500 mL culture of appropriate medium (YPD, YPG, *see* **Notes 1** and **2**) between 0.1 and 0.2 AU/mL (*see* **Notes 12** and **13**).
2. When the culture reaches an optical density between 0.5 and 0.8 AU/mL, add 6.7 mL of 37% formaldehyde (final concentration of 0.5% *see* **Notes 14** and **15**) and incubate the culture for 10 min at 30 °C (*see* **Note 16**).
3. Add 27 mL Tris/glycine (*see* **Note 16**) to quench the cross-linking reaction.

4. Harvest cells by centrifugation at 4000 × *g* for 5 min at room temperature.
5. Place the tubes containing the cell pellets on ice. Wash the cell pellets twice with 10 mL ice-cold PBS. Transfer the cell pellets when required to a 50 mL conical tube. After each washing step, cells are harvested by centrifugation at 2500 × *g* for 5 min at 4 °C and supernatants are discarded.
6. Wash once the cell pellets with 10 mL PQ-lysis buffer. Harvest the cells as described in **step 5** and discard the supernatant (pellets can be stored at −80 °C).
7. Suspend the cell pellets in 1.5 mL PQ-lysis buffer containing *pic*, per 1 g cells.
8. The cell suspension is aliquoted in 2 mL tubes (*see* **Note 17**) containing glass beads (0.8 mL cell suspension and 1.4 g glass beads 0.75–1 mm).
9. Lyse the cells by vortexing in an orbital shaker at 4 °C four times 10 min, 2 min cooling in between (*see* **Note 18**).
10. Pierce the bottom of the tube containing the cell extract using a hot syringe. Insert the tube immediately on a 15 mL conical tube and secure them with parafilm.
11. Transfer the lysate to the 15 mL tube by centrifugation at 2500 × *g* for 5 min at 4 °C.
12. Save 25 µL of supernatant (S1) for quality control analysis by Western blot (Fig. 1). The rest of the supernatant is discarded.
13. Wash the pellet, containing the chromatin fraction, with 1 mL of PQ-lysis buffer containing *pic*. Centrifuge the sample at 2500 × *g* for 10 min at 4 °C and discard the supernatant (*see* **Note 19**).
14. Suspend the pellet in a final volume of 1 mL of PQ-lysis buffer containing *pic*.
15. Shear the chromatin by sonication in a Branson sonifier: Microtip limit 3, 90% Duty cycle to constant, five times ten pulses with 30 sec cooling in between.
16. Centrifuge samples at 2500 × *g* for 5 min at 4 °C. Transfer the supernatant to a 1.5 mL tube (*see* **Note 17**) and save 25 µL (S2) for quality control analysis by western blot (Fig. 1).
17. Centrifuge samples at 15,000 × *g* for 10 min at 4 °C.
18. Save three times 25 µL of the supernatant (W) as the input control for the quantitation of DNA and for quality control analysis by western blot and agarose gel electrophoresis (Fig. 1). The rest is used as the whole cell lysate (WCL) for the affinity purification.

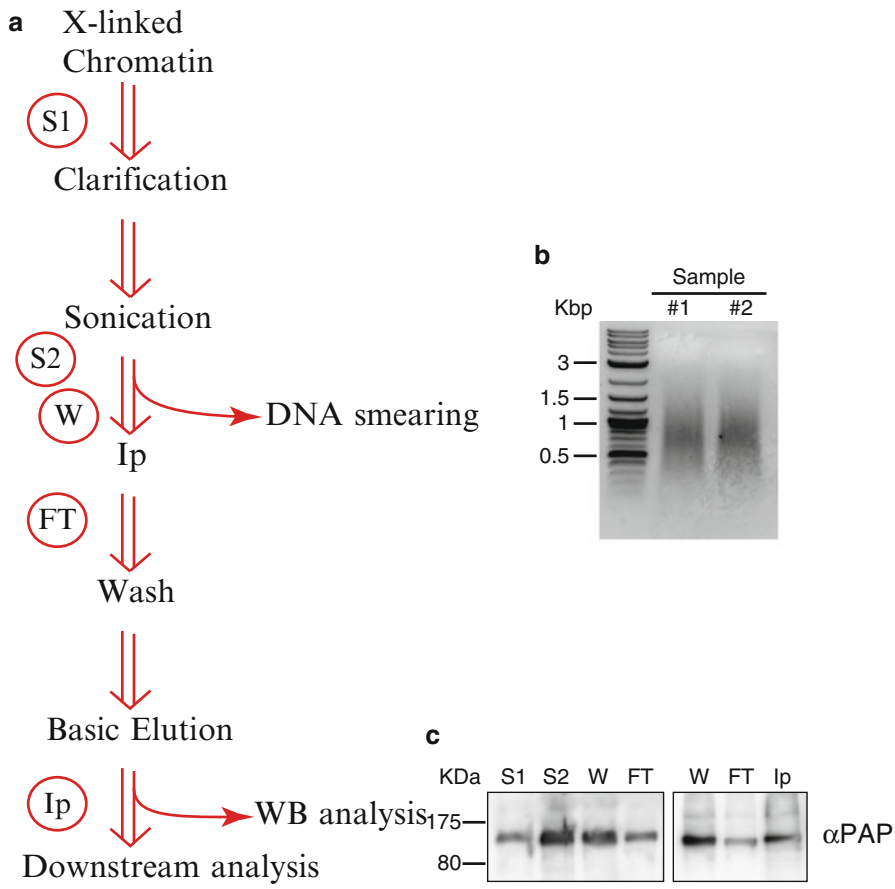

Fig. 1 Scheme for the workflow of the affinity purification method (**a**). The indicated fractions at the *left side* are analyzed by agarose gel (**b**) and western blot (**c**). The smearing of the DNA is a critical step when different samples are compared. Obtained fragments after sonication are analyzed on 1 % agarose-TBE gel. The analysis of DNA fractionation shows a distribution centered around 700 bp (**b**). In order to test the efficiency of the affinity purification, 10 % of the elution fraction and 2.5 % of the other fractions are analyzed by western blot using anti-PAP antibody (**c**). In this analysis we show that most of the Rpa135-ProtA is enriched in the chromatin fraction (S2 and W in comparison with S1). Rpa135-ProtA is depleted from the extract after the affinity purification (compare W and FT). The efficiency of the affinity purification can be calculated by quantifying the signals in W and Ip

3.3 Affinity Purification of C-Terminal ProteinA Tagged Rpa135

1. Wash 200 µL magnetic beads (*see* Subheading 3.1) once with 1 mL ultrapure water. Suspend the beads and capture them using the BcMag separator. Supernatant is discarded (*see* **Note 20**).
2. Wash the beads three times with 1 mL PQ-lysis buffer and remove supernatants after each washing step as described in **step 1**.
3. Add the WCL sample to the beads (*see* **Note 21**).
4. Incubate the beads for 3 h at 4 °C under gentle rotation (*see* **Note 22**).

5. Beads are captured as indicated in **step 1**. Save 25 µL of supernatant (FT, flow through) for quality control analysis by western blot (Fig. 1), discard the rest.
6. Wash the beads four times with PQ-lysis buffer at 4 °C and discard supernatants as indicated in **step 1** (*see* **Note 23**).
7. Wash the beads two times with buffer AC at 4 °C (*see* **Note 24**).
8. Elute by adding 0.5 mL 500 mM ammonium hydroxide (freshly prepared)and incubate for 20 min at room temperature under gentle rotation. Repeat elution and combine the two eluates (*see* **Note 25**).
9. Lyophilize overnight (*see* **Note 26**). The sample can be used for mass spectrometry analysis. Save a small amount for the quality control analysis (10–20%).

3.4 Quality Control Analysis of the Affinity-Purified Rpa135-ProtA

1. Add 5 µL of Laemmli loading buffer to the 25 µL of the saved fractions. Boil samples for 5 min at 95 °C and let them cool down.
2. Spin samples 10 s in a centrifuge.
3. Analyze 15 µL on an SDS-PAGE, transfer to a nitrocellulose membrane and perform western blot analysis (Fig. 1).

3.5 Quality Control Analysis of DNA Fragmentation

1. Spin tubes for 10 s (*see* **Note 26**).
2. Add 10 µg of RNase A and incubate for 1 h at 37 °C.
3. Add 20 µg of glycogen and 50 µg of proteinase K in final 200 µL reactions (*see* **Note 27**) and incubate at 56 °C for 3 h.
4. Add 200 µL of phenol:chloroform:isoamyl alcohol, vortex for 10 s and centrifuge at room temperature for 5 min at 15,000 × *g*.
5. Recover 175 µL of the aqueous phase and transfer to new tubes. Add 18 µL of 3 M sodium acetate pH 5.2, mix and add 460 µL of 100% ethanol. Incubate the samples for at least 2 h at –20 °C.
6. Centrifuge the samples at 15,000 × *g* for 30 min at 4 °C. Discard the supernatant, add 600 µL of 70% ethanol and centrifuge again at 15,000 × *g* for 5 min at 4 °C. Discard the supernatant and air dry the pellets.
7. Solubilize the DNA pellets from the input and the elution samples with 200 µL of 10 mM Tris–HCl pH 8.0. Use 20 µL of 10 mM Tris–HCl pH 8.0 and analyze by electrophoresis in an agarose gel (Fig. 1).

4 Notes

1. YPD or YPG medium usage will depend on the use of strains containing conditionally expressed mutants using GAL promoter.

2. Alternatively, when plasmid maintenance is mandatory, YNB medium supplemented to satisfy auxotrophic requirements could be used. Carbon source selection will vary as indicated in **Note 1**.
3. Stabilization of formalin with 10–15% methanol avoids the formation of paraformaldehyde, which will increase the cross-linking length, thus the number of unspecific contacts.
4. Stock solutions could be prepared. 0.5 M HEPES-KCl pH 7.5, 5 M NaCl, 0.5 M EDTA pH 8.0, 10% v/v Triton, 2% w/v sodium deoxycholate, 10% w/v SDS.
5. *pic* is stored at −20 °C. Due to the short half-life of PMSF in water solutions, *pic* will be added when required to the individual tubes and not to the general buffer.
6. Coupling of Magnetic beads to IgG has been previously described by our group [3].
7. Stock solutions could be prepared. 5 M ammonium acetate pH 7.4 and 1 M $MgCl_2$.
8. Stock solutions could be prepared. 1 M Tris–HCl pH 8.0, 1 M LiCl, 10% v/v NP-40.
9. Prolonged incubation times prior to the addition of the rabbit IgGs should be avoided because the epoxide group is not stable under these conditions.
10. High local concentration of ammonium sulfate due to a fast addition of the salt could result in the precipitation of rabbit IgGs.
11. The solution should be rapidly removed as described in **step 2** to avoid denaturation of IgG polypeptides.
12. AU refers to Absorbance Units at 600 nm.
13. Volume culture depends on the downstream techniques which will be used. We recommend 500 mL for large scale purifications (required for the mass spectrometry analysis of associated proteins) and 50 mL for small-scale purifications (quantitative analysis of bound DNA, ChIP).
14. The indicated cell concentration may vary between yeast strains. In genetic backgrounds S288 and W3O3 these conditions would reflect logarithmic growth phase. However, if SC media is used we will strongly recommend to keep the concentration below 0.6 UA/mL.
15. The final concentration of formaldehyde depends on the downstream techniques which will be used. We recommend 0.5% w/v formaldehyde for downstream mass spectrometry analysis of associated proteins and 1% w/v formaldehyde for the quantitative analysis of bound DNA.
16. For 50 mL cultures, 1.4 mL of formaldehyde and 2.7 mL of Tris/glycine would be added. Time of cross-linking should be experimentally tested. For RNA Pol I, we recommend

10–15 min. Longer incubations could mask the epitope or increase the number of unspecific associations.

17. It is important to use tubes of high quality to prevent interference with downstream analyses. We use Eppendorf original (#0030120086).
18. Alternatively cells could be processed in a Fast Prep instrument (screw caps tubes needed) three times 20″ at 6.0, 1′ cooling in between in ice water.
19. The supernatant is free of genomic DNA and contains proteins not cross-linked to DNA, which could include our bait. Thus, the use of this fraction will considerably reduce the yield of purified chromatin.
20. 50 µL of slurry IgG beads is enough for small-scale analyses.
21. The ratio of WCL and beads is experimentally determined for the quantitative analysis of bound DNA. The capacity of the beads could be saturated with small amounts of sample because of the high size of DNA–protein complexes. In this case, a large increase of input sample would not cause a proportional increase of the purified protein and the ratio of "affinity-purified DNA vs input DNA" will decrease. We propose to fix the volume of beads to 50 µL and try different concentrations of WCL. We have observed that the ratio beads/WCL strongly depends on the bait protein.
22. Overnight incubation will not harm the sample. For slow growing cell strains, culture could be attempted in day one and cell pellet stored at −80 °C as it is mentioned in **step 9** in Subheading 3.1. In that case, we recommend 3 h incubation in order to shorten the procedure.
23. Due to its higher stringency, washes are performed with the DQ-lysis buffer when the sample is used for the quantitative analysis of bound DNA.
24. Washes will be done in DQ2 buffer when the sample is used for the quantitative analysis of bound DNA.
25. Wash the beads with 1 mL TE buffer when the sample is used for the quantitative analysis of bound DNA. This step will aid the elution. Discard the supernatant as in Subheading 3.3, **step 1**. Add 75 µL of DQE buffer and incubate 20 min in a thermomixer at 37 °C followed by centrifugation at 15,000 × *g*. Collect the supernatant as in Subheading 3.3, **step 1**. Save 25 µL for quality control.
26. Incubate the sample for the quantitation of input DNA (Subheading 3.2 **step 18**) and the 50 µL from the elution sample (in **Note 25**) at 65 °C overnight. Afterwards proceed with the DNA precipitation (*see* **steps 1–7** in Subheading 3.5).
27. Add 20 µg of proteinase K to the elution samples.

Acknowledgments

We thank all members of the "Institute für Biochemie III" for their support and for helpful discussions. The help of Prof. Dr. Rainer Deutzmann and Eduard Hochmuth in establishing downstream mass spectrometric analysis is gratefully acknowledged. We thank Prof. Dr. Herbert Tschochner, Dr. Philipp Milkereit, and Dr. Joachim Griesenbeck for their support and critical reading of the manuscript.

References

1. Déjardin J, Kingston RE (2009) Purification of proteins associated with specific genomic loci. Cell 136:175–186
2. Byrum SD, Raman A, Taverna SD et al (2012) ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus. Cell Rep 2:198–205
3. Hamperl S, Brown CR, Perez-Fernandez J et al (2014) Purification of specific chromatin domains from single-copy gene loci in Saccharomyces cerevisiae. Methods Mol Biol 1094:329–341
4. Hamperl S, Brown CR, Garea AV et al (2014) Compositional and structural analysis of selected chromosomal domains from Saccharomyces cerevisiae. Nucleic Acids Res 42:e2
5. Puig O, Caspary F, Rigaut G et al (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 24:218–229
6. Hierlmeier T, Merl J, Sauert M et al (2013) Rrp5p, Noc1p and Noc2p form a protein module which is part of early large ribosomal subunit precursors in S. cerevisiae. Nucleic Acids Res 41:1191–1210

Chapter 17

Bioinformatic Analysis of ChIP-seq Data on the Repetitive Ribosomal RNA Gene

Uwe Schwartz and Gernot Längst

Abstract

Ribosomal RNA genes are highly repetitive and therefore not annotated in genome assemblies. We did recently analyze the epigenetic and architectural regulation of murine ribosomal genes by the Transcription Termination Factor I and made use of genome-wide histone modification ChIP-seq data. This method paper describes how repetitive genomic regions can be integrated into custom genomic assemblies and be used with genome-wide profiling data.

Key words ChIP-seq, Bioinformatics, Ribosomal RNA gene, Repetitive genomic elements

1 Introduction

Epigenetic control of rRNA genes plays an essential role in cell growth and genomic stability. The epigenetic information resides in self-propagating molecular signatures like histone modifications, DNA methylation, chromatin-associated RNAs, and nucleosome positions. These marks provide a memory of previously experienced environmental signals, such as environmental stimuli, developmental cues, or internal events, without changing the genetic information. Specific chromatin modifying activities are setting and changing these epigenetic modifications resulting in a metastable marking of genomic regions. The epigenetic marks, the chromatin structure of genes, and the nuclear architecture can be assessed genome-wide by various techniques, including ChIP-seq, MNase-seq, and 3C assays, elucidating the current activity state and marking regulatory regions of genes. These kind of epigenetic profiling and bioinformatics analysis are routine methods for single copy genes but fail for repetitive genomic elements.

The essential ribosomal RNA genes (rDNA), located in the nucleolus and transcribed by the RNA Polymerase I, belong to such repetitive genomic regions. The analysis of rDNA chromatin

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_17,

structure by genomic approaches has been problematic, as rDNA is not included in current genome assemblies and sequencing reads corresponding to rDNA are generally discarded during analysis. Mammalian cells contain several hundred rRNA genes that are clustered in tandem repeats on five acrocentric chromosomes at Nucleolar Organizer Regions (NORs) in a telomere to centromere orientation. With the nucleolus being the most prominent nuclear body and the largest cellular transcription factory, the nucleolus represents a paradigm for studying both the organization and regulation of gene expression. Studies centered on the rRNA genes revealed that epigenetic processes may lock rRNA gene copies in specific active, repressed, and poised states that coexist in growing and growth-arrested cells [1].

In a recent study, we analyzed the chromatin structure and architecture of the murine ribosomal RNA gene, showing that the Transcription Termination Factor I (TTFI) is playing an essential role in orchestrating and regulating the epigenetic status and activity of these genes [2]. We established the bioinformatic pipeline to analyze genome-wide datasets with the repetitive rRNA genes. Active enhancer elements are characterized by specific histone marks, like H3K27ac and H3K4me1, which are involved in long-range chromatin interactions [3], Notably, our integrative genomic analysis revealed characteristic enrichment of histone marks at the TTFI binding sites of the rRNA gene terminator region, which can be observed in human as well [4]. These results suggest that the gene terminator has all hallmarks of a functional enhancer, such as distal location, regulatory histone marks, and the potential to exert gene activation by direct, protein-mediated DNA loops. The terminator contains a homotypic cluster of TTFI binding sites enabling high-affinity binding of TTFI to chromatin, initiating the formation of a DNA loop between the gene terminator and the promoter, followed by TTFI dependent gene activation and coordinating gene transcription with transcription termination [2].

To explore epigenetic profiling of the murine rDNA, we previously constructed a build of the murine genome containing a single copy of rDNA to which we aligned short sequence reads from ChIP-seq studies [2]. Using this approach, we were able to define the location of regulatory elements of the murine rDNA, taking into account that we monitor all the different epigenetic activity states of the transcription unit at the same time. This method could, in principle, be applied to any repetitive DNA element for which the sequence is available, but the genomic annotation is missing.

2 Material

2.1 Local Hardware, Software, and Sequence Data

1. PC workstation with UNIX operating system.
2. FastQC.
3. Bowtie 2.
4. SAMtools.
5. HOMER.
6. IGV.
7. Mouse reference genome sequence (https://support.illumina.com/sequencing/sequencing_software/igenome.html).
8. Mouse rDNA sequence (accession: BK000964).

3 Methods

3.1 Quality Control

Prior to any downstream analysis, the quality of the sequencing reads has to be assessed. This initial analysis provides an overview to what kind of biases and problems might occur during data analysis.

1. The quality of the sequencing reads has to be determined as a first step. For this, various software tools are available (FastQC, NGSQC, HTQC, htseq-qa, etc.). We use FastQC (available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ and installed on your local computer drive) a comprehensive and widely used software package to control the quality of high-throughput sequencing data, covering the most critical parameters. The results are summarized in a graphical overview and detailed reports.
2. Go through the FastQC reports and check for possible error prompts. An overview of the sequencing accuracy can be obtained in the "per base sequence quality" section. The sequencing accuracy is illustrated as Phred quality score giving the logarithmic probability that a base is called incorrectly (*see* **Note 1**). Several options are available to check for sequence biases and sequence duplication events during sample preparation. Check them carefully (*see* **Note 2**).

3.2 Custom Genome Build Including the Mouse rDNA

The custom sequence is integrated as an artificial chromosome into the reference genome. The addition of the reference genome in the alignment process decreases the likelihood of false annotations (reads originating from positions in the reference genome with high-sequence similarity) and allows proper normalization between various conditions, afterwards. Reference sequences of commonly used genomes as the mouse genome, can be downloaded from several servers (for example, from https://support.illumina.com/sequencing/sequencing_software/igenome.html (*see* **Note 3**))

1. Use a text editor to build a text file in fasta format containing your custom sequence (*see* **Note 4**). The fasta file should contain an empty line at the end of the sequence.
2. Concatenate the custom sequence and the public available reference genome in fasta format using the cat command in the terminal. Save the output to a new file (*see* **Note 5**).
3. To efficiently map the reads to a reference, the annotation file has to be converted to an appropriate index format. Build a bowtie2-index by running the bowtie2-build script of the Bowtie 2 package [5].

3.3 Mapping

1. Bowtie 2 is an efficient tool for aligning short sequencing reads to relatively long genomes as the mouse genome [5]. For the mapping of ChIP-seq data, the local alignment option of Bowtie 2 is highly recommended (*see* **Note 6**). Apart from that, the standard settings are sufficient to obtain a good mapping efficiency, which is displayed at the prompt after the alignment run.
2. After the alignment run, Bowtie 2 returns the mapping efficiency. Local alignments typically result in >90% overall alignment rates. However, reads might have multiple mapping sites and cannot be assigned confidently to a unique position in the genome. Use the view script as part of the SAMtools (Sequence Alignment/Map Tools) package [6] in order to filter reads with uncertain origin (*see* **Note 7**).

3.4 Visualization of Histone Modifications

1. The HOMER package [7] provides several useful scripts for further analysis purposes of ChIP-seq data. As a first step, the sequence alignment is transformed into a suitable data structure and the quality of the ChIP-seq experiment is assessed, by running the makeTagDirectory script.
2. Depending on the inherent GC-bias it might be required to normalize the data using the -normGC option.
3. To visualize the ChIP-fragment density at each position of the genome, create a bedGraph file using the makeUCSCfile script.
4. Load your custom build genome and the generated bedGraph file into the Integrative Genomics Viewer (IGV) [8] and zoom into the custom sequence, which is considered as a chromosome. The IGV is installed at your local computer drive and can handle customized genomes.

4 Notes

1. The Phred quality score Q is defined as the negative logarithmic error probability P ($Q = -10 \log_{10} P$). A quality score $Q = 20$ is interpreted as 1 error of 100 predictions and consequently, represents a 99% prediction accuracy.

2. Often an error is prompted for "per base sequence content" and "per base GC content" due to an observable sequence bias at the first bases at the 5′-end corresponding to positions exhibiting a lower median quality score. Since this manual uses a local alignment in the mapping step, where reads are trimmed automatically at both ends, no further preprocessing is required.
3. Downloading large genomes via the browser can cause problems and end up incomplete. Therefore, a direct download from the iGenomes repository is recommended using the command line (see step-by-step instruction http://support.illumina.com/help/SequencingAnalysisWorkflow/Content/Vault/Informatics/Sequencing_Analysis/swSEQ_IlluminaGenomes.htm).
4. Using the text editor with "plain text," setting is recommended to ensure that the text format is not converted. The first line of a fasta file should contain the name of the sequence and start with a ">" sign.
5. The cat command concatenates several files and can be used as: cat my_custom_sequence.fasta reference_genome.fasta > my_custom_build_genome.fasta. Afterwards the head command (head -n *m* my_custom_build_genome.fasta) can be used to display the first *m* lines of the generated file. Check that the ">" signs of the first two sequences appear at the beginning of a line, otherwise you forgot to press return at the end of your custom fasta file.
6. Bowtie 2 provides two alignment modes. In one method, called "end-to-end" alignment, all characters of the read will be used to find the best alignment, while the other method seeks after the best local alignment by clipping unmatched characters at the ends of the reads if necessary.
7. Bowtie 2 assigns a mapping quality score (MAPQ) to each mapped read, which is a measure of the probability that the alignment does not correspond to the read's true point of origin. The MAPQ is calculated as the Phred quality score (*see* **Note 1**). Using the view function of the SAMtools package with the settings -q 20 will remove all reads having a MAPQ < 20, i.e., reads having at least a 99 % mapping accuracy are kept.

Acknowledgment

Work in our laboratory is funded by the German Research Community (DFG, grants within the SFB960)

References

1. Grummt I, Längst G (2013) Epigenetic control of RNA polymerase I transcription in mammalian cells. Biochim Biophys Acta 1829:393–404. doi:10.1016/j.bbagrm.2012.10.004
2. Diermeier S, Grummt I, Németh A et al (2013) Chromatin-specific regulation of mammalian rDNA transcription by clustered TTF-I binding sites. PLoS Genet 9:e1003786. doi:10.1371/journal.pgen.1003786
3. Rye M, Saetrom P, Handstad T et al (2011) Clustered ChIP-Seq-defined transcription factor binding sites and histone modifications map distinct classes of regulatory elements. BMC Biol 9:80. doi:10.1186/1741-7007-9-80
4. Zentner GE, Saiakhova A, Manaenkov P et al (2011) Integrative genomic analysis of human ribosomal DNA. Nucleic Acids Res 39:4949–4960. doi:10.1093/nar/gkq1326
5. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi:10.1038/nmeth.1923
6. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352
7. Heinz S, Benner C, Spann N et al (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576–589. doi:10.1016/j.molcel.2010.05.004
8. Robinson JT, Thorvaldsdóttir H, Winckler W et al (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26. doi:10.1038/nbt.1754

Chapter 18

Deep Sequencing Analysis of Nucleolar Small RNAs: RNA Isolation and Library Preparation

Baoyan Bai and Marikki Laiho

Abstract

The nucleolus is a subcellular compartment with a key essential function in ribosome biogenesis. The nucleolus is rich in noncoding RNAs, mostly the ribosomal RNAs and small nucleolar RNAs. Surprisingly, also several miRNAs have been detected in the nucleolus, raising the question as to whether other small RNA species are present and functional in the nucleolus. We have developed a strategy for stepwise enrichment of nucleolar small RNAs from the total nucleolar RNA extracts and subsequent construction of nucleolar small RNA libraries which are suitable for deep sequencing. Our method successfully isolates the small RNA population from total RNAs and monitors the RNA quality in each step to ensure that small RNAs recovered represent the actual small RNA population in the nucleolus and not degradation products from larger RNAs. We have further applied this approach to characterize the distribution of small RNAs in different cellular compartments.

Key words Nucleolus, Small RNA purification, Library preparation, Deep sequencing

1 Introduction

The nucleolus is a major cellular compartment for RNA production, processing, and modification [1–3]. Its key function, ribosome biogenesis, is mediated by RNA polymerase I transcription, followed by cotranscriptional processing, cleavage, and maturation of the ribosomal RNAs [4–6]. These processes are mediated and regulated by small nucleolar RNAs (snoRNAs), needed for RNA folding and modification together with their cognate ribonucleotide proteins [7]. Multiple additional RNA species such as RNase P RNA, 7S RNA, and microRNAs have also been detected in the nucleolus [7]. Given the rich representation of RNAs in the nucleolus, especially rRNA and snoRNAs, and reports of abundant expression of small (s) RNA derivatives, we have addressed the identity and subcellular distribution of the sRNAs. We wanted to resolve what types of small RNAs are detectable in the nucleolus and asked whether their expression or localization is affected by pathways involved in sRNA processing or by

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_18,

rRNA transcription. We therefore used massively parallel sRNA sequencing (sRNA-seq) to investigate in an unbiased manner the expression of <30 nt RNAs isolated from nuclear, cytoplasmic, nucleolar, or total cellular fractions.

We have previously developed a TRIzol based method for sequentially recovery of the nucleolar macromolecular fractions, protein, RNA, and DNA from the nucleolus [8, 9]. We have shown that this method is suitable for isolation of the RNA component from the nucleolus at both high quality and quantity. However, the fraction of small RNAs (18–22 nt) in a total RNA extract varies significantly among different tissues and species, and in human is around 0.5–9.2% [10], resulting the majority of sequence reads generated from library constructed on total RNA to correspond to the most abundant RNA species. Although methods have been developed to enrich RNAs in the size range below 200 nucleotides, these are also abundant in RNA species, like the 5.8S RNA, the 5S RNA, tRNAs, snoRNAs, and other snRNAs which are not necessarily of interest to a study aiming at the smallest range of the RNA complement (e.g. the 15–30 nt fraction).

To increase the detection rate of small RNAs, we have developed a strategy that enriches the small RNA population in a stepwise fashion without loss of RNA integrity (Fig. 1). We first divide a total RNA extract into two fractions that contain RNA species above and below 200 nt, respectively. RNA integrity is monitered by RNA agarose gel analysis of the RNAs larger than 200 nt. We

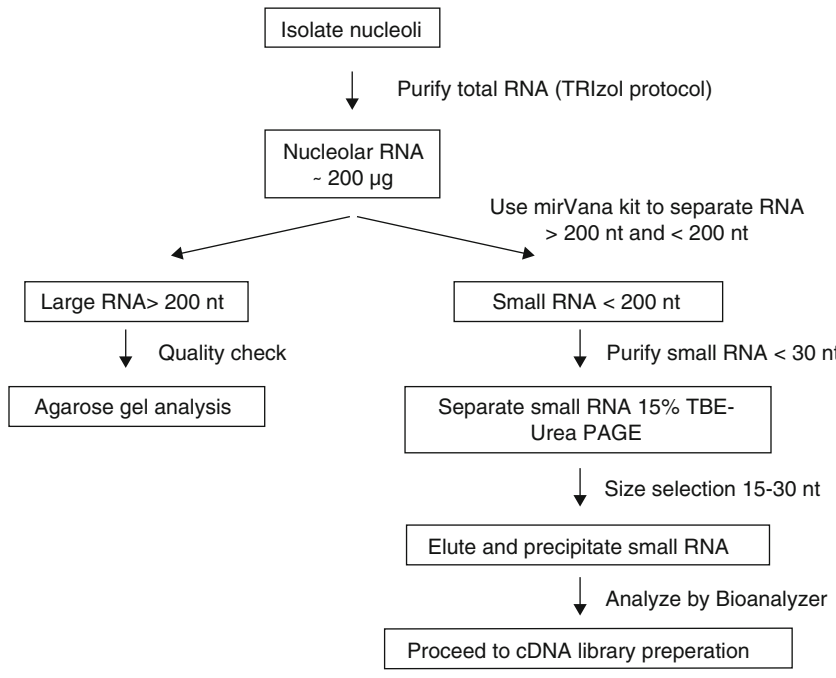

Fig. 1 Flow chart for small RNA preparation

further separate the RNA below 200 nt on a 15% TBE-Urea PAGE gel and excise gel slices containing RNAs in the size ranges 17–30 and 30–200 nt. The purified samples of RNAs in different size ranges are further analyzed using the Agilent small RNA Chip. By analysis of RNA quantity and quality in the sample at each step of purification, we can calculate loss during purification and estimate the percentage of small RNAs in the total RNA population. To monitor library construction and efficiency, we add two differently sized synthetic RNA spike-ins (18 and 21 nt) to the RNA samples. The synthetic RNAs are then detectable in the libraries. The addition of spike-ins before library construction is a useful tool for checking the efficiency of library construction and normalizing data between samples; however, the amount of spike-ins may have to be optimized according to the sequencing platform one operates on.

Here, the library has been prepared mainly using strategy based on the Ion Torrent Small RNA seq kit (Fig. 2). We have used this strategy to characterize the small RNAs in the nucleolus and to compare the small RNA distribution in different cellular compartments and have shown that the small RNome of the nucleolus has many unique features [11, 12]. Our studies include a comprehensive analysis of the subcellular distribution of sRNAs in HeLa cells featuring among other particulars a distinct and enriched complement of snoRNA-derived small RNAs.

2 Materials

All solutions should be prepared into ultrapure deionized water, and all chemicals should be of analytical, and when possible, nuclease-free grade. Avoid contamination of nucleases from the surfaces, equipment, or plasticware. Use gloves throughout every step. Appropriate handling, procedures, and safety protocols should be adhered to when handling biological materials and potentially hazardous reagents. Institutional policies should be strictly adhered to in this regard.

2.1 Reagents

1. TRIzol (Invitrogen).
2. Chloroform.
3. Isopropanol.
4. Ethanol.
5. 80% Ethanol (prepared in nuclease-free water), made fresh before use.
6. Nuclease-free water (Applied Biosystems).
7. MirVana™ miRNA isolation kit.
8. Novex TBE-Urea sample buffer (2×) (Invitrogen).

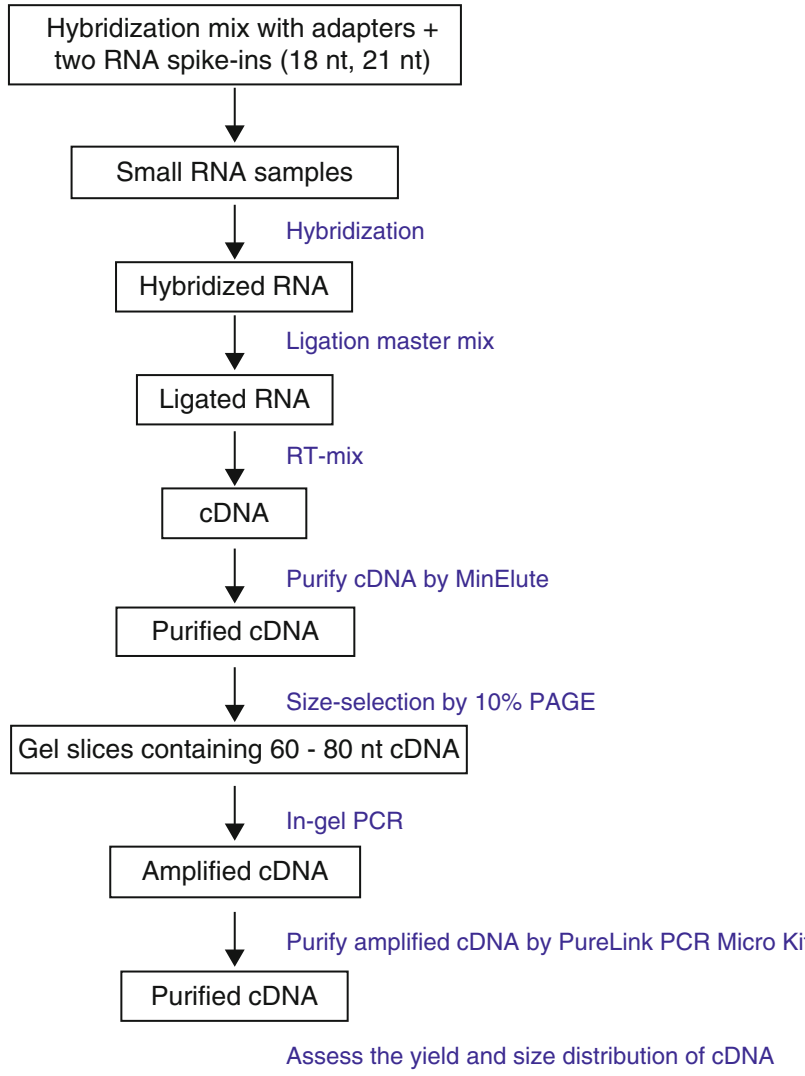

Fig. 2 Flow chart for small RNA library preparation

9. 10 bp DNA ladder (Invitrogen).
10. 1.5 mm thick 15% TBE-Urea gel (normally prepared in lab).
11. 1× TBE buffer: 89 mM Tris–HCl, 89 mM boric acid, 2 mM EDTA, pH 8.0.
12. Glycogen (Invitrogen).
13. SYBR Gold nucleic acid gel stain, 10,000× (Invitrogen).
14. Elution buffer: 0.3 M NaCl (prepared in nuclease-free water).
15. Spin-X cellulose acetate filter (Corning).
16. Agilent small RNA kit (Agilent).

17. Ion Total RNA-Seq Kit v1 or v2.
18. Spike-in RNA oligos (use commercial sources for synthesis): AGTAACTCTAGCGGCTTAGTC and ATACGTCGACACGGTTCA.
19. MinElute PCR purification Kit (Qiagen).
20. 1.0 mm thick Novex 10% TBE-Urea gel, 10 well (Invitrogen).
21. PureLink PCR Micro Kit, 50 prep (Invitrogen).
22. Agilent DNA 1000 Kit (Agilent).

3 Methods

3.1 Isolation of Small RNAs from the Nucleolus

3.1.1 Extraction of Total RNA from Nucleoli

1. Purified nucleoli are resuspended in 1 mL TRIzol reagent. Since the nucleolar pellet/precipitate solubilizes slowly, vortexing the suspension normally improves solubility (*see* **Note 1**).
2. Incubate the lysate for an additional 5 min at room temperature.
3. Add 200 μL cold chloroform (CH_3Cl) and mix by vortexing.
4. Incubate the mixture for 3 min.
5. Centrifuge the mixture for 15 min at 12,000 × *g* and 4 °C.
6. Carefully transfer the aqueous phase of the sample, which contains the RNA component, to a new tube (around 600 μL). DNA isolation and protein extraction should preferably be carried out within 24 h.
7. To the aqueous phase (the RNA containing fraction), add 500 μL 100% isopropanol (room temperature) to precipitate the RNA.
8. Mix the sample by inverting the tube several times and incubate at room temperature for 10 min.
9. Centrifuge for 10 min at 12,000 × *g* and 4 °C.
10. Prepare 80% ethanol and incubate on ice.
11. Carefully remove the supernatant from the tube. Do not disturb the RNA pellet.
12. Wash the pellet with 1 mL cold 80% ethanol. Invert the tube to resuspend the RNA pellet.
13. Centrifuge the tube for 5 min at 12,000 × *g* and 4 °C.
14. Remove the supernatant and air dry the RNA pellet for 10 min.
15. Dissolve the RNA pellet in 50 μL DEPC-treated water.
16. Quantify the RNA sample using Nanodrop or Qubit. Usually, the 260/280-ratio is around 2.0 and the 260/230-ratio around 2.1.
17. The RNA sample should preferably be stored at −80 °C.

18. Optional: Visualize RNA integrity by agarose gel electrophoresis.

3.1.2 Separation of Small RNAs (Size < 200 nt) from Total Nucleolar RNAs

1. Start with 200 μg total nucleolar RNA in 100 μL nuclease-free water (*see* **Note 2**).
2. Add 500 μL MirVana Lysis/Binding Buffer.
3. Add 1/10 volume of miRNA Homogenate Additive (60 μL) to the RNA mixture from **step 2**.
4. Cap the tube and mix the samples by inverting the tube several times.
5. Incubate the samples for 10 min at room temperature.
6. Add 1/3 volume of 100% ethanol (220 μL) to the RNA mixture.
7. Spin the mixture at 2000 × *g* for 5 min at room temperature to pellet large RNAs (>200 nt) (save the RNA pellet for quality check, *see* **step 18**).
8. Transfer the 880 μL supernatant containing the small RNAs into a 2 mL new tube.
9. Add an equal volume of isopropanol (880 μL) and mix well.
10. Precipitate the small RNAs in −80 °C for at least half hour.
11. Centrifuge the small RNAs at maximal speed (we use 16,400 × *g*) at 4 °C for 30 min.
12. Remove the supernatant.
13. Wash the RNA pellet once with 1 ml cold 80% ethanol.
14. Collect the RNA pellet by centrifugation for 5 min at 7500 × *g* at 4 °C. Discard the wash supernatant.
15. Air dry the RNA pellet containing the small RNAs for 10 min at room temperature on bench.
16. Dissolve the RNA pellet in 20 μL nuclease-free water.
17. Measure RNA concentration using Nanodrop.
18. (*Continue* from **step 7**) Wash the pellet containing large RNAs with 1 mL cold 80% ethanol.
19. Collect the RNA pellet by centrifugation for 5 min at 7500 × *g* at 4 °C. Discard the wash supernatant.
20. Dry the RNA pellet containing large RNAs for 10 min at room temperature on bench.
21. Dissolve RNA pellet in 100 μL nuclease-free water.
22. Measure RNA concentration by Nanodrop.
23. Optional: Visualize RNA integrity by agarose gel electrophoresis.

3.1.3 Extraction of 18–30 nt Small RNAs from the Larger ~200 nt RNAs

1. Mix 18 µl small RNAs (*continue* from Subheading 3.1.2, **step 17**) with 18 µL 2× urea-TBE loading buffer.
2. Mix 5 µl 10 bp DNA ladder (50 ng/µl) with 5 µL 2× urea-TBE loading buffer.
3. Denature the RNA sample and 10 bp DNA ladder at 70 °C for 5 min, and immediately incubate the sample on ice.
4. Load the 10 bp DNA ladder and the RNA samples on 1.5 mm thick 15% TBE-Urea PAGE gel (*see* **Note 3**).
5. Run the gel electrophoresis in 1× TBE buffer using a constant voltage of 200 V.
6. Stop the gel electrophoresis when the bromophenol blue reaches/has traveled ¾ of the whole gel, usually after about 40 min.
7. Stain the gel in 30 ml 1× SYBR Gold solution in a clean plastic box.
8. Image the gel using UV-transilluminator.
9. According to the 10 bp DNA ladder, cut out the gel fragments corresponding to the 18–30 nt bands and 30–200 nt bands using a clean scalpel.
10. Use a clean 21-gauge needle to make a small hole in the bottom of two 0.6 mL tubes.
11. Carefully transfer the gel fragments into a 0.6 mL tube and mark with the size of the respective bands.
12. Weigh the gel fragments.
13. Put the 0.6 mL tubes inside 1.7 mL tubes and centrifuge at 2000 × *g* for 3 min for the gel to disintegrate.
14. Add a volume of 0.3 M NaCl elution buffer corresponding to 3× the gel weight (µg = µL) to each gel and rotate overnight at room temperature to elute the RNAs from the gel.
15. Next day, centrifuge the eluates at 2000 × *g* for 5 min to precipitate the gel debris.
16. Transfer the supernatants containing the RNAs to the top of Spin-X cellulose acetate filters to remove any remaining gel debris.
17. Centrifuge the filter in the microfuge for 2 min at full speed.
18. Add 2 µL of glycogen (20 µg/µL) and 2.5 volumes of cold 100% ethanol.
19. Cap the tubes and mix well.
20. Precipitate the small RNAs at −80 °C for at least half an hour.
21. Centrifuge for 30 min at 16,400 × *g* at 4 °C.
22. Remove the supernatant.

23. Wash the RNA pellet with 1 mL cold 80% ethanol.
24. Collect the RNA pellet by centrifugation and discard the wash supernatant.
25. Air dry the RNA pellet at room temperature for 10 min.
26. Resuspend the RNA pellet containing the 18–22 nt small RNAs and the 30–200 nt small RNAs in 10 μL and 20 μL nuclease-free water, respectively.
27. Measure the small RNA purity and concentration using an Agilent small RNA chip (*see* **Note 4**).

3.2 Construction of the Small RNA Library (2–3 Days)

3.2.1 Preparation of cDNA

1. Prepare the hybridization master mix in 0.2 mL PCR tubes: 2 μL Ion Adaptor mix, 3 μL hybridization solution and 1 μL mixed RNA oligo spike-ins (*see* **Note 5**) for one library.
2. Add 2 μL small RNA sample (5 ng/μL) to 6 μL hybridization master mix, mix by pipetting up and down a few times.
3. Centrifuge briefly to collect the sample.
4. Incubate the hybridization reaction in a thermal cycler at 65 °C for 10 min followed by 16 °C for 5 min.
5. Prepare the ligation master mix on ice: 10 μL 2× ligation buffer (*see* **Note 6**) and 2 μL ligation enzyme.
6. Add 12 μL ligation master mix to 8 μL hybridization reaction for a total reaction volume of 20 μL, mix by pipetting.
7. Incubate the ligation reaction on a thermal cycler at 16 °C for 16 h.
8. If you use Ion Total RNA-Seq Kit v1, prepare the reverse transcription master mix: mix 11 μL nuclease-free water, 4 μL 10× RT buffer, 2 μL 2.5 mM dNTP mix and 2 μL Ion RT primer. Alternatively, if you use Ion Total RNA-Seq Kit v2, *move* to **step 13**.
9. Add 19 μL RT master mix to each 20 μL ligation reaction and mix well by pipetting.
10. Incubate on a thermal cycler with a heated lid at 70 °C for 5 min, then snap-cool on ice.
11. Add 1 μL ArrayScript reverse transcriptase to each ligated RNA sample and gently vortex to mix thoroughly, then centrifuge briefly.
12. Incubate on a thermal cycler with a heated lid at 42 °C for 30 min and hold at 4 °C.
13. (*Continue* from **step 7**) For Ion Total RNA-Seq Kit v2, prepare the reverse transcription master mix: mix 2 μL nuclease-free water, 4 μL 10× RT buffer, 2 μL 2.5 mM dNTP mix and 8 μL Ion RT primer V2.
14. Add 16 μL of master mix to each 20 μL ligation reaction and mix well by pipetting.

15. Incubate on a thermal cycler with a heated lid at 70 °C for 10 min and hold at 4 °C.
16. Add 4 μL 10× SuperScript III Enzyme mix.
17. Incubate on a thermal cycler with a heated lid at 42 °C for 30 min and hold at 4 °C.

3.2.2 cDNA Purification by the MinElute PCR Purification Kit

1. Transfer all of the prepared cDNA (40 μL) to a clean 1.5-mL microfuge tube.
2. Add 60 μL of nuclease-free water.
3. Add 500 μL of Buffer PB or Buffer PBI.
4. Mix well by pipetting.
5. Load the cDNA onto the MinElute column.
6. Centrifuge at 13,000 × *g* for 1 min. Discard the flow through.
7. Add 750 μL Buffer PE to the MinElute column to wash the column, centrifuge at 13,000 × *g*.
8. Discard the flow through. Return the MinElute column to the microfuge tube.
9. Spin the empty column at 13,000 × *g* for 1 min.
10. Place the MinElute column in a clean microfuge tube.
11. Add 12 μL of Buffer EB to the center of the MinElute column.
12. Wait 1 min, and spin the column at 13,000 × *g* for 1 min.

3.2.3 Size: Selection of cDNA Using a 10% Urea-TBE Gel

1. Prepare 1× TBE buffer and preincubate it at 55 °C in a water bath.
2. Prerun the 10% urea-TBE gel at least 20 min using a constant voltage of 180 V.
3. Prepare the cDNA sample and the DNA ladder: 6 μL cDNA/10 bp DNA ladder (50 ng/μL) and 6 μL loading dye.
4. Denature the DNA sample at 95 °C for 3 min. Snap-cool the tubes on ice for at least 1 min.
5. Load the cDNA sample and the 10 bp DNA ladder in to the gel (*see* **Note** 7). Run the gels at constant voltage of 180 V until the second dye (xylene cyanol, light blue) front just passes the middle of the gel (~45 min).
6. Transfer the gel to a clean plastic box containing 30 mL 1× SYBR Gold solution.
7. Stain the gel with gentle rotation for 10 min.
8. Illuminate the stained gel and excise the gel containing cDNA in the size range 60–80 nt.
9. Cut the gel slice (containing the 60–80 nt cDNA) further into four slices and place the two slices from the middle of the lane into two clean 0.2 mL PCR tubes. Place the remaining gel slices into a clean 1.5 mL tube and store at 4 °C.

3.2.4 In-Gel Amplification of the cDNA by PCR

1. Prepare the PCR master mix (100 μL/tube): mix 10 μL 10× PCR buffer, 8 μL 2.5 mM dNTP mix, 2 μL Ion 5′PCR primer, 2 μL Ion 3′ PCR primer, 1.2 μL AmpliTaqDNA, and 76.8 μL nuclease-free water.
2. Add 100 μL PCR master mix to each of the two 0.2 mL PCR tubes containing a gel slice.
3. Ensure the master mix entirely covers the gel slices.
4. Set up the PCR program in a thermal cycler: stage 1 95 °C for 5 min; stage 2 (18 cycles) 95 °C for 30 s, 62 °C for 30 s; and 72 °C for 30 s; stage 3 72 °C for 7 min.

Stage	Temp (°C)	Time
Hold	95	5 min
18 cycle	95	30 s
	62	30 s
	72	30 s
Hold	72	7 min

3.2.5 Purification of the Amplified DNA by the PureLink PCR Micro Kit

1. Combine the two 100 μL PCR reactions in a new 1.5 mL tube, and add 800 μL binding buffer (B2), mix well.
2. Load 500 μL the mixture to the column, spin the column at 10,000 × *g* for 1 min, and discard the flow through.
3. Repeat **step 2** once.
4. Add 600 μL Wash buffer to the column and spin the column at 10,000 × *g* for 1 min, discard the flow through.
5. Centrifuge the empty column at 14,000 × *g* for 1 min.
6. Place the column in a clean elution tube.
7. Add 12 μL of Elution buffer to the center of the membrane.
8. Incubate 1 min at room temperature, and centrifuge the column at 14,000 × *g* for 1 min.
9. Measure cDNA concentration and quality with Agilent DNA 1000 Kit.

4 Notes

1. Isolation of nucleoli was performed as described in ref. [9]. Five minute incubation time is sufficient for the cell lines we have tested (HeLa, HCT116, WS1). Longer incubation time may be required for other cell lines.
2. To ensure that we have enough starting material, we have combined the RNA samples from two independent nucleolar puri-

fications (100–300 μg total). The purity of each RNA extract is checked by agarose gel analysis before pooling the samples.

3. Leave one empty well between DNA ladder and RNA samples, as this makes it easier to excise a clean gel slice.
4. The size of small RNAs can also be visualized by running the small RNA chip in the Agilent Bioanalyzer.
5. We use two different RNA oligo sizes (18 and 21 nt) as spike-ins.
6. If the 2× Ligation Buffer contains a white precipitate, warm the tube at 37 °C for 2–5 min until the precipitate is dissolved. The 2× Ligation Buffer is very viscous; pipet slowly to dispense it accurately.
7. If you are preparing several libraries at the same time, we recommend that you load no more than two cDNA samples in one gel to avoid cross-contamination.

Acknowledgments

The original work leading to the development of this protocol has been supported by NIH P30 CA006973 and Johns Hopkins University start-up funds.

References

1. Pederson T (2011) The nucleolus. Cold Spring Harb Perspect Biol 3:pii:a000638.28
2. Fatica A, Tollervey D (2002) Making ribosomes. Curr Opin Cell Biol 14:313–318
3. Boisvert FM, van Koningsbruggen S, Navascués J, Lamond AI (2007) The multifunctional nucleolus. Nat Rev Mol Cell Biol 8:574–585
4. Haag JR, Pikaard CS (2007) RNA polymerase I: a multifunctional molecular machine. Cell 131:1224–1225
5. Russell J, Zomerdijk JC (2006) The RNA polymerase I transcription machinery. Biochem Soc Symp 73:203–216
6. Turowski TW, Tollervey D (2015) Cotranscriptional events in eukaryotic ribosome synthesis. Wiley Interdiscip Rev RNA 6:129–139
7. Watkins NJ, Bohnsack MT (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. Wiley Interdiscip Rev RNA 3:397–414
8. Bai B, Laiho M (2012) Efficient sequential recovery of nucleolar macromolecular components. Proteomics 12:3044–3048
9. Bai B, Laiho M (2015) Sequential recovery of macromolecular components of the nucleolus. Methods Mol Biol 1228:43–51
10. Tissot C (2008) Analysis of miRNA content in total RNA preparations using the Agilent 2100 bioanalyzer. Agilent Technologies, Palo Alto, CA
11. Bai B, Liu H, Laiho M (2014) Small RNA expression and deep sequencing analyses of the nucleolus reveal the presence of nucleolus-associated microRNAs. FEBS Open Bio 4:441–449
12. Bai B, Yegnasubramanian S, Wheelan SJ, Laiho M (2014) RNA-Seq of the nucleolus reveals abundant SNORD44-derived small RNAs. PLoS One 9:e107519

Chapter 19

Deep Sequencing Analysis of Nucleolar Small RNAs: Bioinformatics

Baoyan Bai and Marikki Laiho

Abstract

Small RNAs (size 20–30 nt) of various types have been actively investigated in recent years, and their subcellular compartmentalization and relative concentrations are likely to be of importance to their cellular and physiological functions. Comprehensive data on this subset of the transcriptome can only be obtained by application of high-throughput sequencing, which yields data that are inherently complex and multidimensional, as sequence composition, length, and abundance will all inform to the small RNA function. Subsequent data analysis, hypothesis testing, and presentation/visualization of the results are correspondingly challenging. We have constructed small RNA libraries derived from different cellular compartments, including the nucleolus, and asked whether small RNAs exist in the nucleolus and whether they are distinct from cytoplasmic and nuclear small RNAs, the miRNAs. Here, we present a workflow for analysis of small RNA sequencing data generated by the Ion Torrent PGM sequencer from samples derived from different cellular compartments.

Key words Nucleolus, Small RNA sequencing, Ion torrent PGM, snoRNA, miRNA

1 Introduction

The nucleolus is the site for transcription and maturation ribosomal RNAs (rRNAs). The ribosomal DNA is first transcribed to a 47S precursor rRNA, which is further processed to the mature 28S, 18S, and 5.8S rRNAs by multiple steps that require the proper cleavage, processing, and modification of the rRNAs. The modifications are guided by small nucleolar RNAs (snoRNAs). Two main classes of snoRNAs, referred to as box C/D and box H/ACA snoRNAs, are distinguished based on their short (3–7 nt) and specific consensus motifs that direct their nucleolar localization and protein interactions [1, 2].

Previous reports indicating the nucleolus may contain specific miRNAs [3, 4] and that certain snoRNAs may possess miRNA-like properties [5, 6], prompted us to undertake a comprehensive study of the nucleolar small RNAome content. For this purpose, we

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_19,

developed a strategy to enrich high-purity small RNA samples from various cellular compartments, including the nucleolus, and used these to prepare libraries for deep sequencing [7, 8]. The details of the small RNA library construction method are described in Chapter 18.

The sequencing was conducted on a 314 Ion Torrent PGM platform [7]. Reads generated by this platform require removal of the 3′ end adaptors used in the sample preparation as well as additional filtering of very short or low quality reads, which is carried out using the proprietary Ion Torrent software. For our specimens, after trimming and quality control, 0.1–0.4 million reads were recovered [7]. The number of reads derived from the nucleolar library was in the lower end of this range, possibly due to particular features of the nucleolar small RNAs. Next, all high-quality reads from the libraries were mapped to the human reference genome sequence (hg19). In general, at least 60% of the reads from each library were mapped to the human genome and are available in the GEO as GSE50057. The coordinates for each read mapped onto the genome are stored in a BED file format. Considering that the human reference genome (hg19) does not include annotated rDNA sequences, all the reads that could not be mapped to the human genome were aligned to rDNA in a separate step. During library construction, RNA oligos may be added to the RNA sample as "spike-ins" to control for RNA loss during sample handling, and these must also be identified and quantified in a separate step. Next, we used HTseq-count [9] to identify whether the stored genomic coordinates for each read contained any known transcript information. Reads were annotated by extracting the annotation from a known RefGene file downloaded from the UCSC browser. Our results showed that for total cellular, cytoplasmic, and nuclear RNA samples, 90% reads aligned to miRNA loci, while more than 75% reads in nucleolar library mapped to snoRNA loci, suggesting that the snoRNAs or their loci give rise to a subset of small RNAs that is different from those found in other cellular compartments [7, 8]. The mapping information can also be visualized using the IGV software [10].

In the final steps, we focused on the composition of the small RNAs in the nucleolus relative to other compartments, concentrating on miRNAs and snoRNAs. We investigated whether the expression of the miRNAs in the nucleolus correlated with the expression in other cellular compartments. To this end, we aligned all mapped reads with mature miRNA sequences downloaded from miRBase and counted the reads for each detected miRNA [8]. For the nucleolar library, identical sequences mapped onto small RNA loci were collapsed into a single file of nonredundant sequences with an associated abundance. We aligned nonredundant sequences with known human snoRNA sequences retrieved from snoRNA database and calculated the frequency of nonredundant sequences mapping to the 5′ and 3′ ends of specific snoRNAs as well as their frequency of snoRNA-derived features (e.g., C/D or H/ACA box motifs) [7].

2 Material

2.1 Local Hardware and Software

1. System Linux and PC station.
2. Software installed in Linux.
   (a) Torrent Suite v1.5 or newer.
   (b) FASTX-toolkit.
   (c) FastQC.
   (d) TMAP.
   (e) HTseq.
   (f) Python.
   (g) Perl.
   (h) Blast.
   (i) CD-HIT.
3. IGV installed on PC.

2.2 Databases

1. Human reference genome sequence (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/).
2. rDNA (accession: U13369 and AL592188).
3. Two spike-in sequences: AGTAACTCTAGCGGCTTAGTC and ATACGTCGACACGGTTCA [10].
4. Gene annotation file (http://hgdownload.cse.ucsc.edu/downloads.html#human).
5. miRNA sequence (http://www.mirbase.org/).
6. snoRNA sequence (https://www-snorna.biotoul.fr/).

3 Methods and Pipeline

We focus on major steps in the handling of small RNA libraries and the identification of small RNAs from these data: trimming and mapping of the data to the reference genome, annotation, and visualization (Fig. 1). Other massive parallel RNA sequencing platforms and their proprietary softwares are equally compatible with this approach.

3.1 Preparing the Database for Alignment

1. Establish the database work directory in the Linux server.
2. Download the human reference genome from the UCSC browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/) and store the data in the database directory.
3. Download the rDNA sequences from NCBI (http://www.ncbi.nlm.nih.gov) and store sequences in a FASTA format in the database directory.
4. Create a file containing spike RNA sequences in the database directory.

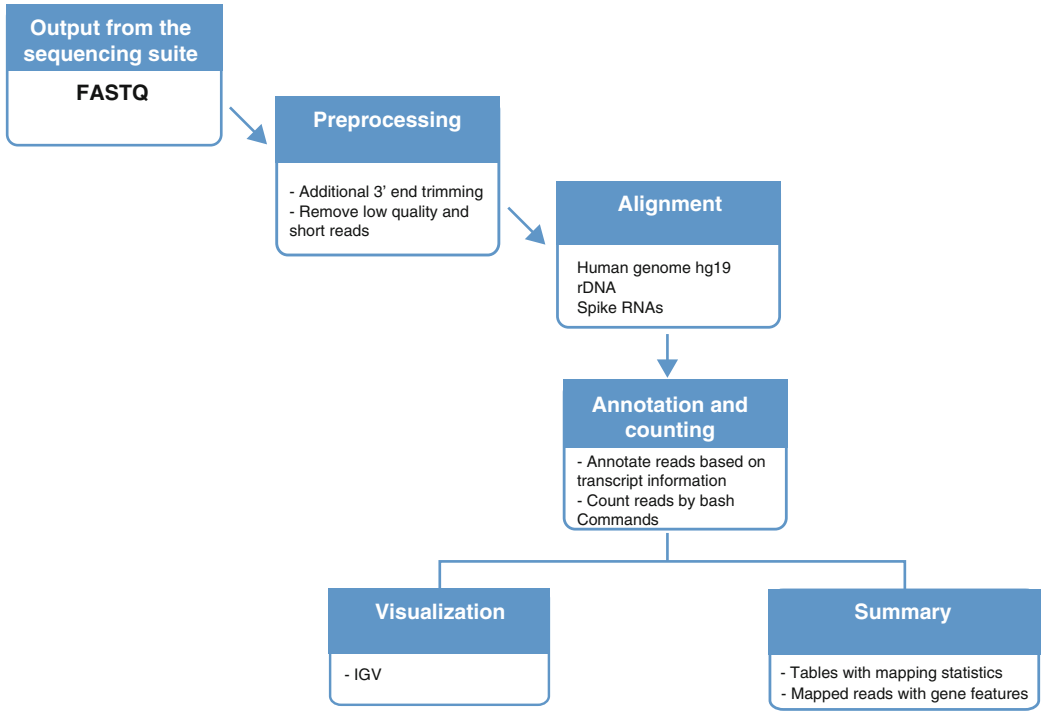

Fig. 1 Pipeline for data management and small RNA analysis

3.2 Preprocessing the Data and Quality Control

1. Prepare the work directory on the Linux server.
2. Copy/transfer the data (Nucleolar.fastq) from the Ion Torrent PGM sequencer machine to the directory.
3. Remove low quality and short reads by using Fastq_quality_trimmer tool (FAST-toolkit): *fastq_quality_trimmer -t 17 -15 -i Nucleolar.fastaq -o Nucleolar_trimmed.fastaq*.
4. (optional) To assess sequence quality and potential biases detectable in a RNA-seq library, the preprocessed reads were used as input to generate plots to assess quality values of the nucleotide composition by the FastQC tools: *fastq -o qc -f Nucleolar.fastq*. All output files will be written in the qc directory.

3.3 Mapping the Reads with Reference Sequences (Human Genome, rDNA, and Spike-in RNA)

1. Create index reference genome: *tmap index -f hg19.fasta*.
2. Map the reads to the human reference genome using TMAP program: *mapall -f hg19. fa -r Nucleolar_trimmed.fastq -v map1 map2 map3*.
3. Considering that rDNA is not included in hg19.fasta, align unmapped reads with rDNA (U13369 and AL592188) using blast software.
4. Parse the blast result using perl script.

5. Align the reads with two spike RNAs and count how many reads mapped onto spike RNAs in each library.
6. Calculate the length of all mapped reads and plot the size distribution in each library.

3.4 Counting the Reads by Known Transcripts

1. Download the reference gene annotation file (RefGene) written in a GTF format from UCSC genome Browser.
2. Count how many reads map onto known transcripts using hyseq-count.py script in HTseq python package: *htseq-count nucleolar_reads.sam RefGene*.

3.5 Viewing Alignments

1. Download IGV Java web and install it on PC.
2. Load the human genome hg19.
3. Sort and index the alignment file received from Subheading 3.3, **step 2**.
4. Load the sorted and indexed alignment file.

3.6 Downstream Analysis

1. Download all human mature miRNA sequences from miRBase (http://www.mirbase.org/) and store the sequences in a FASTA file.
2. Index the miRNA_database.fa file.
3. Align all mapped reads with miRNA_database using blast software.
4. Parse the result and count the reads for each miRNA in each library.
5. Calculate the correlation coefficient between the libraries.
6. Take the reads mapped onto snoRNA loci from the nucleolar library and store them in FASTA format in No_snoRNA_reads.fa file.
7. Remove the redundant sequences and cluster the reads in No_snoRNA_reads.fa using CD-hit software. The output file contains all unique sequences and their abundance.
8. Manually copy all human known snoRNA sequences from snoRNABase (https://www-snorna.biotoul.fr//) and store them in a FASTA format file (snoRNA_database.fa).
9. Index snoRNA_database.fa.
10. Align all unique sequences with snoRNA_database using blast software.
11. Parse the result and count the reads for each snoRNA.
12. Calculate the mapping frequency of the reads at the 5′ end and 3′ end of each snoRNA.
13. Inspect the result manually.
14. Search for C/D and H/ACA box motifs in the reads mapped to known snoRNA sequences and calculate the frequency.
15. Inspect the result manually.

References

1. Kiss T (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell 109:145–148
2. Kiss T, Fayet-Lebaron E, Jady BE (2010) Box H/ACA small ribonucleoproteins. Mol Cell 37:597–606
3. Politz JC, Hogan EM, Pederson T (2009) MicroRNAs with a nucleolar location. RNA 15:1705–1715
4. Li ZF, Liang YM, Lau PN et al (2013) Dynamic localisation of mature microRNAs in Human nucleoli is influenced by exogenous genetic materials. PLoS One 8:e70869
5. Ender C, Krek A, Friedlander MR et al (2008) A human snoRNA with microRNA-like functions. Mol Cell 32:519–528
6. Taft RJ, Glazov EA, Lassmann T et al (2009) Small RNAs derived from snoRNAs. RNA 15:1233–1240
7. Bai B, Yegnasubramanian S, Wheelan SJ, Laiho M (2014) RNA-Seq of the nucleolus reveals abundant SNORD44-derived small RNAs. PLoS One 9:e107519
8. Bai B, Liu H, Laiho M (2014) Small RNA expression and deep sequencing analyses of the nucleolus reveal the presence of nucleolus-associated microRNAs. FEBS Open Bio 4:441–449
9. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169
10. Robinson JT, Thorvaldsdottir H, Winckler W et al (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26

Chapter 20

Quantitative Proteomic Analysis of the Human Nucleolus

Dalila Bensaddek, Armel Nicolas, and Angus I. Lamond

Abstract

Recent years have witnessed spectacular progress in the field of mass spectrometry (MS)-based quantitative proteomics, including advances in instrumentation, chromatography, sample preparation methods, and experimental design for multidimensional analyses. It is now possible not only to identify most of the protein components of a cell proteome in a single experiment, but also to describe additional proteome dimensions, such as protein turnover rates, posttranslational modifications, and subcellular localization. Furthermore, by comparing the proteome at different time points, it is possible to create a "time-lapse" view of proteome dynamics. By combining high-throughput quantitative proteomics with detailed subcellular fractionation protocols and data analysis techniques it is also now possible to characterize in detail the proteomes of specific subcellular organelles, providing important insights into cell regulatory mechanisms and physiological responses. In this chapter we present a reliable workflow and protocol for MS-based analysis and quantitation of the proteome of nucleoli isolated from human cells. The protocol presented is based on a SILAC analysis of human MCF10A-Src-ER cells with analysis performed on a Q-Exactive Plus Orbitrap MS instrument (Thermo Fisher Scientific). The subsequent chapter describes how to process the resulting raw MS files from this experiment using MaxQuant software and data analysis procedures to evaluate the nucleolar proteome using customized R scripts.

Key words Quantitative proteomics, Mass spectrometry, Nucleolus, Subcellular fractionation

1 Introduction

Nucleoli represent a major compartment within the nucleus that play a crucial role in ribosome subunit biogenesis and in other important mechanisms, including regulation of molecules involved in the control of cell cycle progression, tumorigenesis, apoptosis, and stress responses. Nucleoli are dynamic compartments that are characterized by a constant flux of macromolecules. Therefore, the precise composition of the nucleolus can vary, both between cell types and within the same cell type under different growth conditions. Defining the protein content of the nucleolus in different cell types and under various growth conditions is thus of great importance for understanding the mechanisms regulating gene expression and cell physiology.

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_20,

Mass spectrometry (MS)-based technology is now the method of choice for large-scale protein identification and quantitation [1–5]. The typical "bottom-up" MS workflow involves digesting the proteins isolated in cell extracts into peptides. Digestion is most commonly performed using trypsin, either alone or in combination with other proteases. The resulting peptide mixture is then either analyzed directly (sometimes called "single shot" analysis), or, to increase the depth of protein coverage, the resulting peptides can optionally be further fractionated by HPLC. For example, chromatographic methods such as hydrophilic strong anionic exchange, hydrophilic interaction liquid chromatography (HILIC), or high pH reverse phase chromatography, can be used prior to MS analysis. In either case, the peptide mixtures are subsequently fractionated again, usually using a reverse phase nano-HPLC system linked to an electrospray device in front of a mass spectrometer. The peptides are eluted from the reverse phase column using a gradient of organic solvent, resulting in the gradual delivery of different peptide species from the complex mixture into the MS instrument. The elution gradient is typically run over a 2–4 h period, with a primary "survey" MS scan (also called the "MS1" scan) performed approximately once per second. With current generation MS instruments this setup can identify ~10–20 of the most abundant peptide peaks in the eluted fraction every second, with subsequent MS/MS scans (also called MS2 scans) then performed to sequence and reliably identify the peptides.

We have previously employed this so-called "bottom-up" MS proteomics approach (Fig. 1) to characterize the protein composition of nucleoli isolated from multiple human cell types. This has identified several thousand different proteins that co-purify with highly enriched preparations of nucleoli [6, 7]. We have also employed the Stable Isotope Labeling with Amino acids in Cell culture (SILAC) methodology, developed by Mann and colleagues [8–11], to enhance the quantitative analysis of nucleolar proteins and to facilitate studies of nucleolar dynamics using a "time-lapse" proteomic approach [6].

In the SILAC workflow for analyzing cultured cell lines, cells are grown for several division cycles in media containing specific amino acids which have the usual (referred to as "light") isotopes of either carbon, nitrogen and/or hydrogen, substituted with heavier isotopes. For example, ^{12}C replaced with ^{13}C, ^{14}N replaced with ^{15}N, and ^{1}H replaced with ^{2}H. This results in complete replacement of the cellular protein pool with the heavy-substituted forms of amino acids, which can be readily detected and quantified in the MS instrument. Typically only the amino acids arginine and/or lysine are substituted with heavy isotopes because these are the sites of trypsin cleavage, thereby ensuring a single heavy-isotope substituted amino acid is present in each peptide in a tryptic hydrolysate of the protein extract.

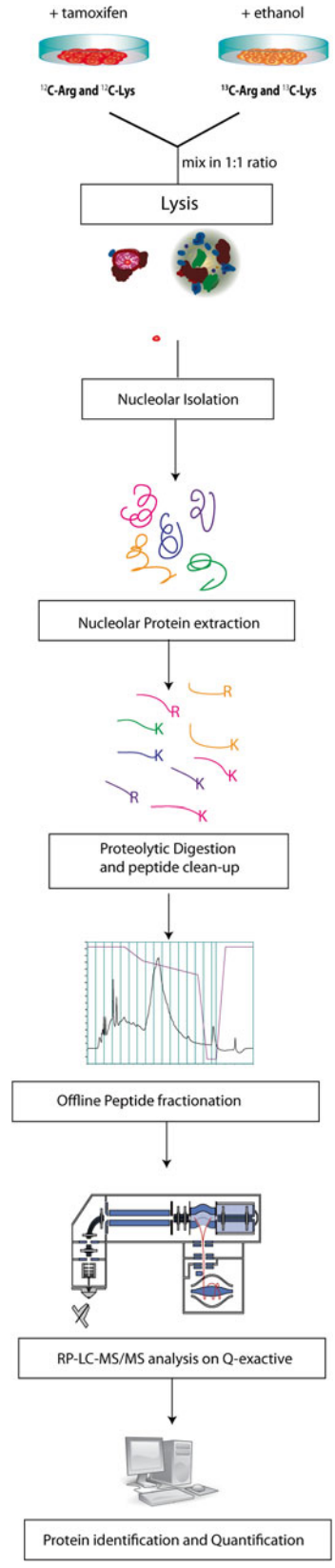

Fig. 1 Schematic description of a typical quantitative proteomics "bottom-up" workflow

Here, we present a protocol for the quantitative, MS-based proteomic analysis of nucleoli, which combines SILAC with nucleolar purification. This is illustrated here with a SILAC experiment comparing the protein composition of nucleoli isolated from the immortalized but untransformed human breast epithelial cell line MCF10A-Src-ER, and the same cell line after the cells have been transformed by activation of v-Src. This cell line constitutively expresses the transforming viral oncogene v-Src fused to the hormone-binding domain of the estrogen receptor. The tyrosine kinase activity of this fusion protein is activated upon addition of hormone, resulting in rapid transformation of the cells accompanied by major phenotypic changes affecting their morphology and behavior. We describe an experiment to study the changes in the nucleolar proteome that occur in this MCF10A-Src-ER cell line 48 h after cell transformation resulting from activation of v-Src. The MS analysis is performed using a Q-Exactive Plus MS instrument (Thermo Fisher Scientific) and in-line nano-HPLC reverse phase chromatography. The following chapter will describe how the resulting raw MS data files can be analyzed to identify and quantitate peptides and proteins and how these data can further be analyzed to characterize the nucleolar proteome (Chapter 21 Nicolas et al.).

2 Materials

1. The MCF10A-Src-ER cell line was established from the MCF10A cell line (*see* **Note 1**).
2. The labeled amino acids for SILAC can be purchased from Cambridge Isotope Laboratories: L-arginine: HCl (13C6, 99%, 15N4, 99%), L-arginine: HCl unlabeled, L-Lysine: 2HCl (13C6, 99%; 15N2, 99%), L-lysine: 2HCl, unlabeled.
3. Standard Cell culture medium: Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F-12) (500 mL), heat inactivated horse serum (25 mL), Penicillin Streptomycin mixture (5000 U/mL, 5.5 mL), insulin from porcine pancreas (Sigma, 500 μL of a 10 mg/mL stock solution), Cholera Toxin (Sigma 50 μL of 1 mg/mL stock solution), human EGF (100 μL, of 100 μg/mL stock solution), hydrocortisone (50 μL of 5 mg/mL solution in ethanol), puromycin (Roche, 0.5 μg/mL final concentration). The medium is filter sterilized using a Stericup and Steritop (Merck).
4. Tamoxifen was purchased from Sigma and was dissolved in ethanol at a final concentration of 10 mM.
5. SILAC media: DMEM/F-12 without arginine or lysine (Dundee Cell Products, 500 mL), 3 kDa cut-off dialyzed horse serum (Dundee Cell Products, 25 mL), Pen Strep (5000 U/mL, 5.5 mL), insulin from porcine pancreas (500 μL of a 10 mg/mL stock solution), Cholera Toxin (50 μL of 1 mg/mL

stock solution), human EGF (100 μL, of 100 μg/mL stock solution), hydrocortisone (50 μL of 5 mg/mL solution) and 0.5 μg/mL puromycin (Roche). Arginine and lysine were added to the media to final concentrations of 0.699 mM and 0.499 mM, respectively. The SILAC media were prepared in two flavors; amino acids K0 and R0 were used for the light media and K8 and R10 were used for the heavy media. The medium was filter sterilized using a Stericup and Steritop (Merck).

6. Sterile PBS.
7. Automated cell counter.
8. Lysis buffer: 20 mM Tris–HCl pH 7.4, 10 mM KCl, 3 mM $MgCl_2$, 10% Glycerol, 1× protease inhibitors cocktail (Roche), 1× PhosStop (Roche), 0.1% IGEPAL CA-630.
9. Sucrose buffer S1: 0.25 M Sucrose, 10 mM $MgCl_2$.
10. Sucrose buffer S2: 0.35 M Sucrose, 0.5 mM $MgCl_2$.
11. Sucrose buffer S3: 0.88 M Sucrose, 0.5 mM $MgCl_2$.
12. Ultrasonic probe sonicator.
13. EZQ-assay (Life technologies).
14. CBQCA fluorescent assay (Life Technologies).
15. Nucleolar protein extraction buffer: 8 M urea, 100 mM triethyl ammonium bicarbonate (pH 8.5).
16. Tris(2-carboxyethyl)phosphine (TCEP, Thermo Fisher Scientific).
17. Iodoacetamide (Sigma).
18. Trifluoroacetic acid (TFA) (Sigma).
19. Formic acid (FA) (Sigma).
20. Urea.
21. Triethyl ammonium bicarbonate (TEAB) (Sigma).
22. HPLC grade acetonitrile and HPLC grade water are purchased from Fluka.
23. MS-grade trypsin (Thermo Fisher Scientific).
24. Lys-C (Wako, Japan).
25. Sep-Pak vac 3 cm^3 (200 mg) cartridges (Waters).
26. Vacuum manifold (Waters).
27. Diaphragm vacuum pump (Knf Lab).
28. TSKgel Amide-80, 3 μm particle, 4.6 mm × 150 mm column (TOSOH).
29. LC–MS compatible polypropylene screw-cap vials (MicroSolv Technology Corporation).
30. Acclaim PepMap 100, 100 μm × 2 cm nanoviper C18, 5 μm, 100 Å precolumns (Thermo Fisher Scientific).

31. Integrated column-emitter EasySpray PepMap RSLC C18, 2 μm particle, 75 μm × 50 cm (Thermo Fisher Scientific).
32. All LC–MS/MS analysis described here are performed on a Q-Exactive Plus Orbitrap (Thermo Fisher Scientific) connected to an RSLC nano-flow system (Thermo Fisher Scientific) via a zero dead volume connection to the EASY-spray column emitter.
33. Genevac centrifugal evaporator EZ-2 plus.
34. This protocol uses corrosive solvents so proper health safety measures should be observed such as wearing eye protection and appropriate gloves, e.g., SHIELDskin CHEM gloves (Shield Scientific).

3 Methods

3.1 Metabolic Labeling of Proteins in Cell Culture

1. Seed cells at appropriate density in 10 cm dishes (in standard media), at this stage one 10 cm plate per condition is sufficient.
2. The following day, replace the normal media in the dishes with the SILAC media.
3. Monitor the cells to avoid over growth, cells should be split when confluency is >80 %.
4. Change the media and split the cells, using one 10 cm plate per condition.
5. Grow the cells through at least three further passages to ensure full labeling of the proteins before expanding the cells to carry out the proteomic study (*see* **Note 2**).

3.2 Treatment of Cells: Oncogenic Transformation of MCF10-Src-ER Cells

1. Subculture the cells so that you have five 10 cm plates of heavy labeled cells and five 10 cm plates of light cells.
2. The cells grown in light media are treated with tamoxifen to a final concentration of 1 μM (2 μL of ethanol per 20 mL of media) to transform the cells, while the cells grown in heavy media are treated with the same volume of ethanol.
3. The cells are grown for a further 48 h at 37 °C prior to harvesting them.

3.3 Harvesting the Cells

1. The cells are washed with PBS, using 10 mL per dish.
2. In order to detach the adherent cells, they are incubated with 2 mL of Accutase (Life Technologies) at 37 °C for 10 min (*see* **Notes 3** and **4**).
3. The cells are spun down at 200 × *g* for 5 min at room temperature.
4. The supernatant is decanted and discarded.

From this point onwards, all steps are carried out at 4 °C unless stated otherwise.

5. The cell pellet is resuspended in ice cold PBS and pelleted by centrifugation at 200 × *g* for 5 min. The supernatant is discarded. This step is repeated twice.
6. The cells are counted using a cell counter (Countess, Invitrogen) to determine the number of cells in each condition.
7. The two cell populations are mixed in a 1:1 ratio, based on cell number, before starting the subcellular fractionation.
8. The mixed cell population is centrifuged to remove PBS.
9. The cell pellet is lysed in lysis buffer by passing them three times through a 21 G syringe (to dissociate cell clusters) and once through a 27 G syringe (to break the cells open and shear the cytoplasm away from the sturdier nuclei). In our hands, this method is more reproducible and efficient without the need for Dounce homogenization.

3.4 Nucleolar Isolation

In this Subsection we will only give a brief description of the nucleolar isolation protocol. The procedure yields highly pure nucleoli from several cultured mammalian cell lines, but it should be adapted to specific cell types, if required.

1. The lysed cells are centrifuged at 228 × *g* for 5 min at 4 °C to pellet the nuclei. The supernatant can be retained as the cytoplasmic fraction.
2. The nuclear pellet is resuspended in 3 ml of buffer S1 and layered over a 3 ml cushion of buffer S2 and centrifuged at 1430 × *g* for 5 min at 4 °C to obtain a cleaner nuclear pellet.
3. The supernatant is discarded and the clean, pelleted nuclei are resuspended in 3 ml of S2 and sonicated for 6 × 10 s (with 10 s rest between each sonication) at full power in an ice bath to prevent overheating of the sample (*see* **Note 5**).
4. Sonicated nuclei are layered over 3 ml of S3 and centrifuged at 2800 × *g* for 10 min, yielding a pellet containing highly pure nucleoli and a supernatant containing the nucleoplasmic fraction.
5. The nucleoli can be further washed by resuspending them in 0.5 ml of S2 prior to centrifugation at 2800 × *g* for 5 min at 4 °C.
6. They can be stored in any volume of this buffer at −80 °C by snap freezing them in liquid nitrogen.

3.5 Protein Extraction from the Nucleolus

1. Before using the nucleoli, they are pelleted by centrifugation at 2800 × *g* for 5 min at 4 °C to remove buffer S2.
2. The protein content of the purified nucleolar samples is extracted in 50 μL of 8 M urea, 100 mM triethyl ammonium

bicarbonate (pH 8.5) with sonication (6 × 30 s on and 30 s off with a pulse amplitude of 50%) at 4 °C (*see* **Note 6**).

3. The protein concentration is determined using a fluorescent assay EZQ, following the manufacturer's instructions.
4. The proteins are reduced using TCEP (20 mM) for 15 min at room temperature prior to alkylation with iodoacetamide (50 mM) at 45 min in the dark.
5. The protein solution is diluted fourfold with TEAB prior to adding Lys-C, in a substrate-to-enzyme ratio of 50:1, w/w. The digestion is carried out at 37 °C overnight.
6. The digests are further diluted so that the final concentration of urea is less than 1 M.
7. Trypsin is added in a substrate-to-enzyme ratio of 50:1, w/w. The reaction is incubated at 37 °C for 8 h.
8. The digests are acidified with trifluoroacetic acid (TFA) prior to desalting.
9. The peptides are desalted using C18 sepak following the manufacturer's instructions.
10. Finally the desalted peptides are dried in vacuo.

3.6 Peptide Fractionation

1. To increase the proteome coverage, peptides are fractionated prior to LC–MS/MS analysis using HILIC.
2. The dried peptides are dissolved in 80% acetonitrile, 20% water incorporating 0.01% TFA.
3. Peptides are separated using a reverse gradient of aqueous on an TSK amide column. The mobile phases used are solvent A: water incorporating 0.1% TFA and solvent B: acetonitrile containing 0.1% TFA. The flow rate is set to 0.4 ml/min, the peptides are separated and eluted using a reverse gradient of solvent B from 80% B to 0% B: from retention time RT = 0 to RT = 20 min hydrophobic peptides are washed out using 80% B, the hydrophilic peptides are gradually eluted by reducing %B from 80 to 70% (from RT = 20 min to RT = 30 min), then to 0% B at RT = 65 min. The column is washed with 100% A for 5 min, then re-equilibrated at 80%B from RT = 75 to RT = 90 min.
4. Fractions are collected in time steps of 126 s per fraction, dried in vacuo and stored at -20° C.

3.7 LC–MS/MS Analysis

1. The peptide fractions are dissolved in 5% formic acid (50 μL).
2. The peptide concentration is determined using a fluorescent assay, CBQCA to determine the volume of the sample to be injected on the analytical column to avoid overloading the column. For a 50 cm column, inject 1–2 μg of digested peptide mixture.
3. The desired volume is placed in HPLC vials and placed in the autosampler.

4. The peptides are analyzed using a nano-Liquid Chromatography (LC) (RSLC) coupled to a Q-Exactive Orbitrap mass spectrometer.
5. The mobile phases used in this separation are solvent A: HPLC grade water containing 0.1 % formic acid (v/v) and solvent B: 80 % acetonitrile in water containing 0.1 % formic acid (v/v). The solvents are volatile and should be prepared fresh at least once a week, with regular monitoring of their viscosity.
6. The instrument is operated in data-dependent acquisition (DDA) mode to automatically switch between MS and MS/MS events using the Tune and Xcalibur software package (Thermo Fisher Scientific), which controls both the LC and the mass spectrometer.
7. The timing between the MS and the LC system is achieved with a standard contact closure, which triggers MS data acquisition after the sample has been injected onto the analytical column.
8. The peptides are resolved on a 50 cm RP- C18 EASY-Spray temperature controlled integrated column-emitter using a 2 h-multistep gradient of solvent B (5 % B to 35 % B), hydrophobic peptides are washed by increasing the B content to 99 %, finally re-equilibrate the column for the next injection using 5 % B for 20 min. Keep the flow rate constant at 200 nl/min.
9. Initiate the spray by applying 2.5 kV to the EASY-Spray needle. The data are acquired under the control of Xcalibur software in a data-dependent mode selecting the 15 most intense ions for HCD-MS/MS (*see* **Note 7** for definitions).
10. Create the data acquisition method in the Xcalibur Instrument Setup, in which full scan spectra, typically in the m/z range (350–1400) are acquired in the Orbitrap with a resolution ($R = 70{,}000$), the accumulation target value is 1E6 ions; maximum fill time is 20 s. For accurate mass measurement, enable Lock Mass to internally calibrate the mass spectra using the polydimethylcyclosiloxane (PCM) ions generated at m/z 445.120025.
11. The top 15 ions are selected from the MS scan to undergo Higher energy C-trap Dissociation (HCD). The fragment ions are measured in the Orbitrap with a resolution ($R = 17{,}500$), the accumulation target is 1E5, maximum fill time is 60 ms, isolation window is 1.6 m/z, normalized collision energy (NCE = 27). Singly charged peptides, as well as peptides carrying charges >6 are excluded. Peptide match and dynamic exclusion are enabled. Figure 2 is a screen capture showing a typical data-dependent method.
12. Source settings were optimized prior to analysis and saved in a tune file used in the method; sheath gas and auxiliary gas flow

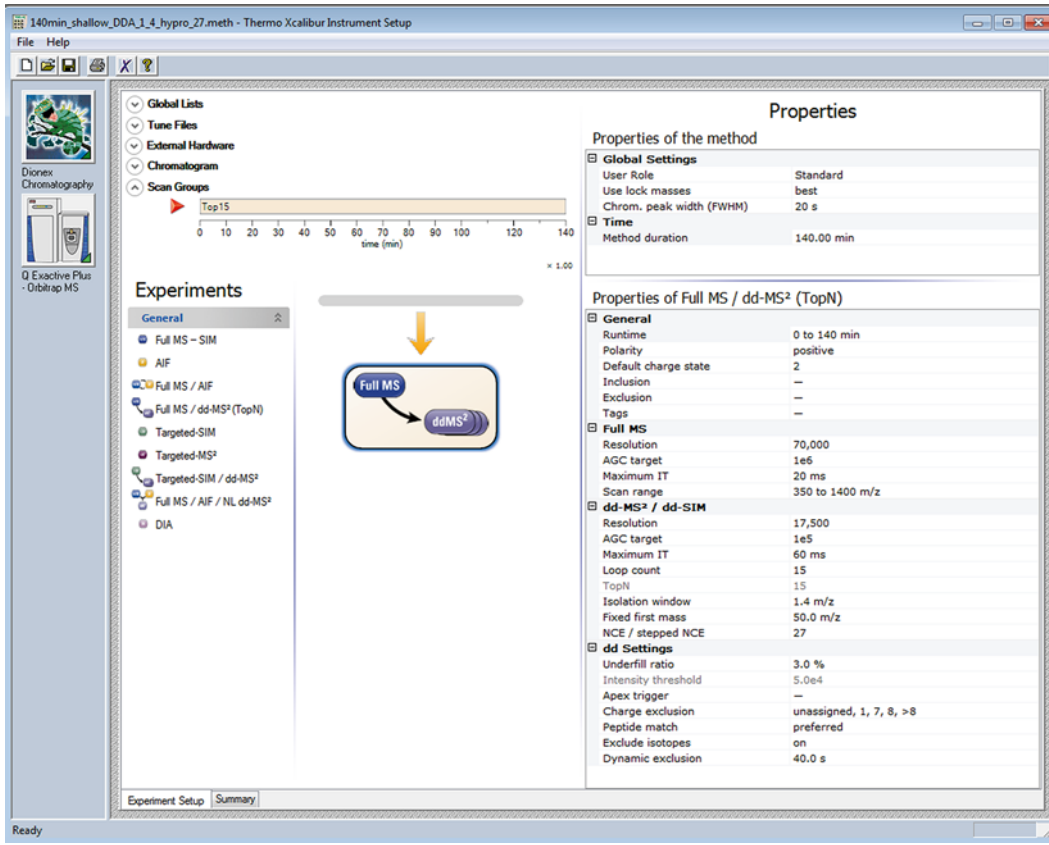

Fig. 2 Screen shot showing a typical data-dependent acquisition method used on a Q-Exactive plus. (**a**) By clicking on the mass spectrometer tab, one can change the MS parameters used during the data acquisition; (**b**) by clicking on the HPLC tab (with the chameleon logo) one can change the HPLC parameters such as gradient length, mobile phase composition, and steepness of the gradient

rates are set to 0, spray voltage is set to 2.5 kV, capillary temperature is set to 250 °C and S-lens RF level is set to 50. Figure 3 is a screen capture showing an example of a tune file.

13. The raw files are analyzed as described in Chapter 21 (Nicolas et al.).

4 Notes

1. The MCF10A-Src-ER was established from MCF10A cell line, which is a spontaneously immortalized, non-cancerous human mammary epithelial cell line derived from the breast tissue of a patient with fibrocystic disease [12]. The MCF10A-Src-ER line expresses an exogenous gene coding for ER-v-Src, a fusion protein with the ligand-binding domain from Estrogen Receptor (ER) fused to v-Src [13]. Treatment of MCF10A-Src-ER cells with the hormone 4-hydroxytamoxifen (4-OHT)

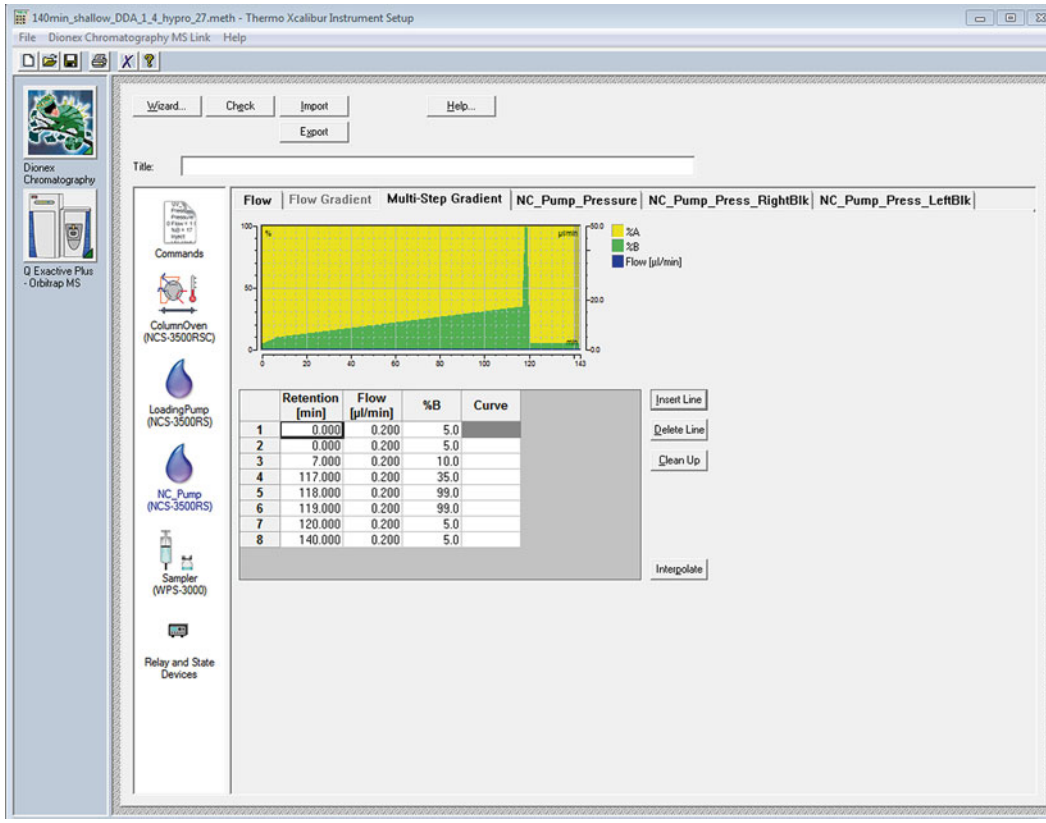

Fig. 2 (continued)

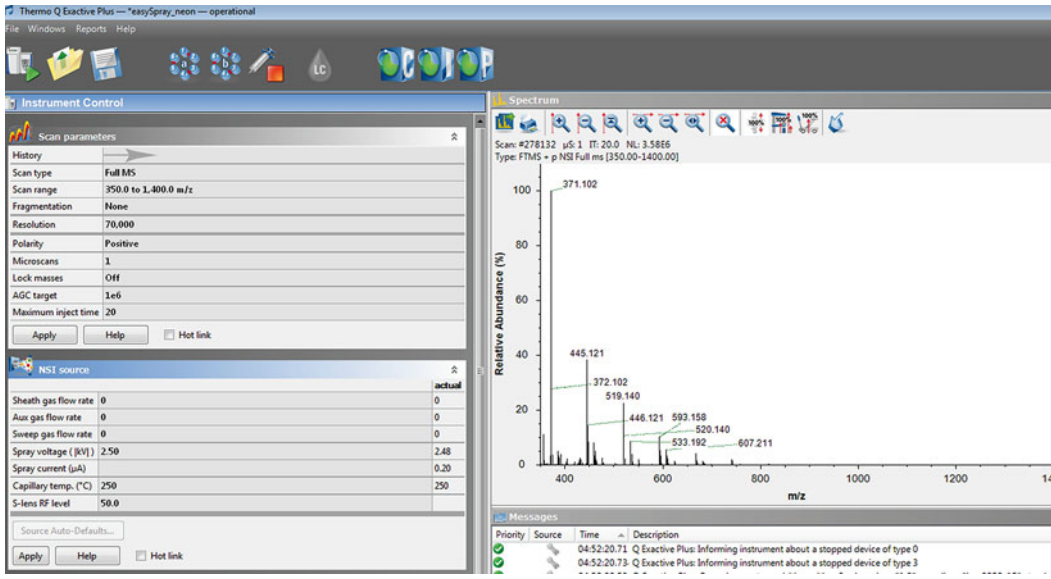

Fig. 3 Screen shot showing a tune file, in which the calibration and the instrument parameters are stored

activates ER-v-Src kinase activity through a conformational change followed by self-phosphorylation on tyrosine 419, triggering phenotypic transformation. MCF10A-Src-ER cells transformed with 4-OHT show phenotypes associated with metastatic tumors that are absent before v-Src activation [13, 14] making this cell line a good model to study oncogenic transformation by v-Src.

2. When labeling cells in culture for the first time, it is useful to take an aliquot of the heavy labeled cells, digest the protein, and analyze by MS to determine the extent of labeling, using cells when labeling is confirmed as >99 %.
3. To maximize extraction of proteins from adherent cells with minimal damage or loss of surface proteins, we avoided scraping the cells (which induces structural damage to cells and leaves a significant portion of the plasma membranes attached to the dish). For obvious reasons in a classic proteomics workflow, we also avoided detaching cells with trypsin.
4. Accutase is a proprietary enzyme preparation, which has been shown to be less toxic to cells than trypsin. It is compatible with FACS detection of most cell surface antigens and thus is likely to preserve much more of the surface motifs of transmembrane proteins.
5. Check the sonicate using phase contrast microscopy to ensure that no intact nuclei remain, and that the nucleoli are readily observable as dense, refractile bodies. If you still see intact nuclei, continue to sonicate (adding 2×10 s steps and checking again).
6. During the sonication of nucleoli, make sure the mixture does not over heat (above 37 °C) to avoid carbamylation of proteins.
7. C18: octadecylsilyl used as packing material for reverse phase columns.

   DDA: Data-dependent acquisition is a mode of data acquisition in which a fixed number of precursor ions whose m/z values are measured in a first MS survey scan are subjected for MS/MS analysis.

   ES or ESI: Electrospray ionization, a soft ionization process in which the ions are transferred from the solution phase to the gas phase at the sampling region of the mass spectrometer (the source).

   HCD: High energy C-trap dissociation, it is a process in which ions are activated through collisions with an inert gas to convert their kinetic energy into internal energy to cause them to fragment giving rise to product ions which can be used to sequence the ion undergoing dissociation.

   HILIC: Hydrophilic interaction liquid chromatography. Analytes are separated based on their hydrophilicity, where

hydrophilic analytes are retained longer on the stationary phase and are eluted with higher aqueous content.

HPLC: High performance liquid chromatography, it is a form of liquid chromatography in which small particle stationary phases are used at relatively high pressures (hundreds of bars).

LC: Liquid chromatography is an analytical technique that is used to separate mixtures of analytes. It uses a stationary phase (column) and mobile phases (liquid).

MS: Mass spectrometry, also refers to a single scan event.

MS/MS: Tandem MS analysis, usually refers to a tandem MS event in which an ion is selected for fragmentation and the fragment ions, or product ions, are detected. This scanning mode is known as product ion scanning.

m/z: Mass-to-charge ratio, in MS it is not the mass that is measured but rather the mass-to-charge ratio of ions.

RP: Reverse-phase, is a form of chromatography in which the stationary phase, the column in this case, is hydrophobic. In this form of chromatography, the analytes are separated based on their hydrophobicity where the most hydrophobic analytes are retained longer on the stationary phase and are eluted later with a higher organic content.

Acknowledgments

D. B. is funded by the Human Induced Pluripotent Stem Cells Initiative (098503/E/12/Z). A.N. is funded by the BBSRC sLoLa grant (BB/K003801/1). This work was funded by a Wellcome Trust Programme grant to A. I. L (108058/Z/15/Z) and supported by a Wellcome Trust Strategic award (105024/Z/14/Z).

References

1. Larance M, Lamond AI (2015) Multidimensional proteomics for cell biology. Nat Rev Mol Cell Biol 16(5):269–280. doi:10.1038/nrm3970
2. Rodríguez-Suárez E, Whetton AD (2013) The application of quantification techniques in proteomics for biomedical research. Mass Spectrom Rev 32(1):1–26
3. Mann M, Kulak Nils A, Nagaraj N, Cox J (2013) The coming age of complete, accurate, and ubiquitous proteomes. Mol Cell 49(4):583–590
4. Altelaar AFM, Munoz J, Heck AJR (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14(1):35–48
5. Jünger MA, Aebersold R (2014) Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. Wiley Interdiscipl Rev 3(1):83–112
6. Lam YW, Lamond AI, Mann M, Andersen JS (2007) Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. Curr Biol 17(9):749–760

7. Andersen JS, Lyon CE, Fox AH, Leung AKL, Lam YW, Steen H, Mann M, Lamond AI (2002) Directed proteomic analysis of the human nucleolus. Curr Biol 12(1):1–11
8. Ong SE, Kratchmarova I, Mann M (2002) Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). J Proteome Res 2(2):173–181
9. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1(5):376–386
10. Ong S-E, Foster LJ, Mann M (2003) Mass spectrometric-based approaches in quantitative proteomics. Methods 29(2):124–130
11. Mann M (2014) Fifteen years of stable isotope labeling by amino acids in cell culture (SILAC). In: Warscheid B (ed) Stable isotope labeling by amino acids in cell culture (SILAC), vol 1188, Methods in molecular biology. Springer, New York, NY, pp 1–7
12. Soule HD, Maloney TM, Wolman SR, Peterson WD, Brenz R, McGrath CM, Russo J, Pauley RJ, Jones RF, Brooks SC (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res 50(18):6075–6086
13. Iliopoulos D, Hirsch HA, Struhl K (2009) An epigenetic switch involving NF-κB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. Cell 139(4):693–706
14. Hirsch HA, Iliopoulos D, Tsichlis PN, Struhl K (2009) Metformin selectively targets cancer stem cells, and acts together with chemotherapy to block tumor growth and prolong remission. Cancer Res 69(19):7507–7511

Chapter 21

Analysis of Mass Spectrometry Data for Nucleolar Proteomics Experiments

Armel Nicolas, Dalila Bensaddek, and Angus I. Lamond

Abstract

With recent advances in experiment design, sample preparation, separation and instruments, mass spectrometry (MS)-based quantitative proteomics is becoming increasingly more popular. This has the potential to usher a new revolution in biology, in which the protein complement of cell populations can be described not only with increasing coverage, but also in all of its dimensions with unprecedented precision. Indeed, while earlier proteomics studies aimed solely at identifying as many as possible of the proteins present in the sample, newer, so-called Next Generation Proteomics studies add to this the aim of determining and quantifying the protein variants present in the sample, their mutual associations within complexes, their posttranslational modifications, their variation across the cell-cycle or in response to stimuli or perturbations, and their subcellular distribution. This has the potential to make MS proteomics much more useful for researchers, but will also mean that researchers with no background in MS will increasingly be confronted with the less-than trivial challenges of preparing samples for MS analysis, then processing and interpreting the results. In Chapter 20, we described a workflow for isolating the protein contents of a specific SILAC-labeled organelle sample (the nucleolus) and processing it into peptides suitable for bottom-up MS analysis. Here, we complete this workflow by describing how to use the freely available MaxQuant software to convert the spectra stored in the Raw files into peptide- and protein-level information. We also briefly describe how to visualize the data using the free R scripting language.

Key words Quantitative proteomics, Mass spectrometry, Nucleolus, Subcellular fractionation, Data analysis, MaxQuant, R, Rstudio, ggplot2

1 Introduction

This tutorial is intended as a practical guide for the steps and procedures involved in the analysis of mass spectrometry (MS) proteomics data. We illustrate the process here by reviewing as a practical example the proteomic analysis of nucleoli isolated from untransformed, human MCF-10A ER-Src cells [1–3], grown in SILAC medium either with (heavy label), or without (light label), activation of the v-Src protein, which results in rapid cell transformation. The nucleolar protein samples were isolated and MS analysis performed using protocols described in Chapter 20 (Bensaddek

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9_21,

et al.). The data analysis procedures described here illustrate the analysis of the Raw MS files (*see* **Note 1**) from a Q-Exactive Plus Orbitrap MS (Thermo Fisher Scientific), as used in Chapter 20 (Bensaddek et al.). We describe how to analyze and quantify the MS spectra using the freely available MaxQuant software suite [4, 5].

The most widely used MS-based proteomics workflow is based on data-dependent, bottom-up peptide analysis. This involves matching experimentally measured MS/MS spectra (*see* **Note 1**) of peptides, which are derived from digested proteins isolated from cells and/or tissues, with a search database of all the potential spectra that can theoretically arise from the expected, observable peptides encoded by the genome of the cells being studied. The peptide sequence in the database that generates the predicted MS spectrum providing the best match to the experimentally observed spectrum is assumed to be the correct identification, subject to various quality control criteria. Overall, this workflow has been established as a robust and effective procedure for high-throughput peptide identification for samples isolated from human cells and well-characterized model organisms. Inevitably, however, this approach can also generate some false peptide assignments. To cope with this, the MaxQuant software uses a "decoy" database to estimate a false discovery rate (FDR, *see* **Note 1**) for peptide identifications within an experiment. This provides an objective way to estimate the reliability of the resulting data set. Thus, each peptide–spectrum match (PSM) is assigned a score—which estimates the quality of the match—and then a posterior error probability (PEP) calculated—which corresponds to the probability that a PSM with a similar score is a false discovery. In line with common practice, we recommend here setting the PEP value to 0.01, discarding peptide identifications with a higher PEP score as being unreliable, i.e., too likely to be false hits. In principle, this results in a data set where we expect at least 99 % of the peptide identifications to be correct.

MaxQuant generates several output tables; the three most important tables contain data corresponding to "evidences," "peptides," and "protein groups."

In overview, the "protein groups" file organizes identified proteins into groups based on them having shared peptides. In each protein group, the leading protein(s) explain(s) all of the peptides assigned to the group, while other proteins in the group (typically isoforms and/or closely related homologues) may only explain a fraction. It is also possible for a peptide to be a part of different protein groups; for instance, a peptide belonging to a common domain found in two or more proteins, where these proteins also have additional, divergent domains, from which signature peptides were also identified in the MS analysis. A peptide assigned to a single protein group is called a "unique" peptide. A peptide shared by several protein groups is assigned as a "razor" peptide (a reference to Occam's razor) to the protein group with the largest num-

ber of observed peptides, on the basis that this provides the most likely explanation for which protein the peptide was derived from.

Each peptide detected in the MS spectra is identified based on a number of "evidences," with each individual evidence representing an instance of that peptide being identified in a single sample. As explained in Chapter 20 (Bensaddek et al.), high-throughput MS-based proteomic experiments are typically performed on peptide mixtures that are fractionated by reverse phase nano-HPLC and gradually eluted with a gradient of acetonitrile and sprayed into the MS instrument. Therefore, each peptide will elute from the nano-LC column over a given time period, typically 30–40 s for a 2 h reversed phase liquid chromatography gradient, and it will be detected by the sequential scans in the MS instrument in a pattern that initially increases, peaks and then decreases again during the time period when it is eluted from the column. For each specific peptide, the time point during the gradient when it elutes and the duration of the peak detection time depends upon both the amino acid sequence of the peptide (reflecting its hydrophobicity) and its abundance in the extract that is being analyzed. Normally, each peptide will only be selected for fragmentation and an MS/MS spectrum measured once or twice during the peak of its elution period (i.e., when its concentration in the MS instrument is at its highest value). We typically observe one evidence per posttranslationally modified sequence per file, but there may be more, especially when a peptide is detected in several charge states. Each evidence listed in the file corresponds to a single successful PSM and thus has its own PEP score derived from the score of the PSM (*see* **Note 1**). Each peptide's PEP score is the lowest of the PEP scores of its evidences, i.e., derived from its spectrum with the best PSM score. By contrast, because all of the peptides identified for a protein collectively contribute to the identification of that protein, the PEP score assigned to each protein is the product of the PEP scores of all the identified peptides that map to that protein sequence.

2 Materials

This section explains how to install and configure the software required for this analysis, namely: MaxQuant, which analyses Raw files and outputs identified peptides, proteins, and their intensities; and the data analysis scripting language R, its graphical interface Rstudio, and its packages ggplot2 and reshape.

2.1 PC Workstation with Windows Vista or Newer OS

This is required to run MaxQuant, which is not cross-platform (see section below).

2.2 MaxQuant

1. Install MaxQuant on your computer (*see* **Note 2**). MaxQuant comes as a .zip archive, which we recommend extracting to the root directory (C:).
2. You will also need to install MSFileReader to allow MaxQuant to read Raw files output from Thermo Fisher Scientific MS instruments (*see* **Note 3**).

2.3 R, Rstudio, and R Packages

1. Download and install the latest version of R (*see* **Note 4**).
2. Download and install the latest version of Rstudio (*see* **Note 5**).
3. Install the R packages "ggplot2" and "reshape": in Rstudio, go to "Tools" and click "Install packages." In the pop-up window, select "Repository (CRAN, CRANextra)" in the "Install from:" drop menu and enter the name of the first package, "ggplot2," in the field below, then click "Install." Repeat the operation with package "reshape" (*see* also **Note 6**).

3 Methods

3.1 Setting Up a MaxQuant Search Database

As a preliminary step before starting the MaxQuant search, a search database must be defined and MaxQuant setup to parse it, as follows.

1. In MaxQuant, go to the "Andromeda configuration" tab, click on "Sequence databases" then click "Add."
2. Click on "Fasta file" and browse to your FASTA database (for sources of FASTA-formatted proteome databases, *see* **Note 7**; for a description of the FASTA format, *see* **Note 1**).
3. Click "Modify table" (*see* **Note 8**).
4. Click "*taxonomy*" then enter either taxon id (left), or species Linnaean name (right).
5. Click on "Name parse rule" to modify the FASTA header parse rule so that it correctly parses the sequence ID in the header of each entry of the FASTA database. MaxQuant does this by using regular expressions (regex), which define a set of complex rules for pattern matching in character strings. The MaxQuant default, ">([^]*)," would match any string starting with ">" followed by any number of characters, as long as they do not contain spaces. Replace the default with the regex ">.*\|(.*)\|" which will match UniProt IDs (*see* **Note 9**).
6. Click "Save changes."

3.2 Setting Up the MaxQuant Analysis

1. In the "Raw Files" tab, click "load" and browse on your computer to your Raw files, which were derived from the MS analysis performed on the Orbitrap Q-Exactive Plus instrument (Chapter 20 (Bensaddek et al.)). It is convenient to have the raw files available on a local drive mounted on your computer. The files can instead be located on a remote drive that is avail-

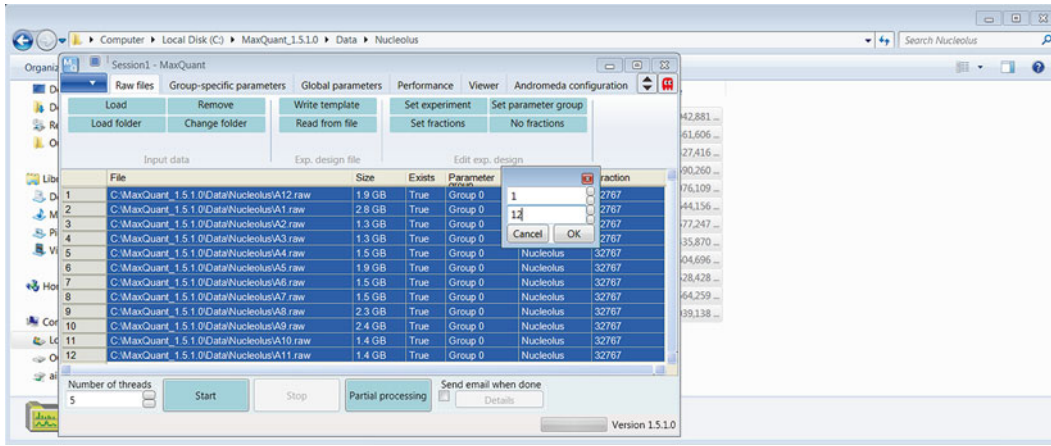

Fig. 1 Setting up a MaxQuant search. Screenshot of the "Raw files" tab

able on the network but this may slow down the analysis, depending on the network bandwidth and connection speed.

2. Assign experiments ("Nucleolus," "Total Cell Lysate," *see* **Note 10**) and fractions (in this example, the HILIC fraction numbers) using the "Set experiment" and "Set fractions" buttons in the ribbon menu (Fig. 1). *See* **Note 10** for more details on what MaxQuant understand under the terms "experiment" and "fraction."
3. Next, set group-specific parameters, i.e., parameters that will only affect specific Raw files (*see* **Note 11**). In this experiment, we have performed a double digestion of the protein samples with LysC (which cuts after lysines) followed by trypsin (which cuts after lysines and arginines), so the final digestion pattern is that of trypsin. If some samples had either been digested instead with different enzymes (e.g., GluC or chymotrypsin), or enriched for specific forms of posttranslational modification (e.g., phosphorylation of Ser, Thr, or Tyr residues), they would be assigned to different parameter groups. In this analysis, since all samples require the same parameters, they all belong to parameter group 0.
4. Click on the "Group-specific parameters" tab, then on the "General" button. Set "Multiplicity" to 2, and under "Heavy labels" tick "Arg10" and "Lys8." Do not tick any "Light labels." In the "Variable modifications" field, keep the two default modifications—"Acetylation (Protein N-term)" and "Oxidation (M)"—and add "Deamidation (NQ)," "Gln ->pyro-Glu" and "Phosphorylation (STY)" (*see* **Note 12**). Keep all other fields at their default values (Fig. 2; also *see* **Note 13** on modifications).
5. Leave all other parameters in the "Group-specific parameters" tab at their default values.

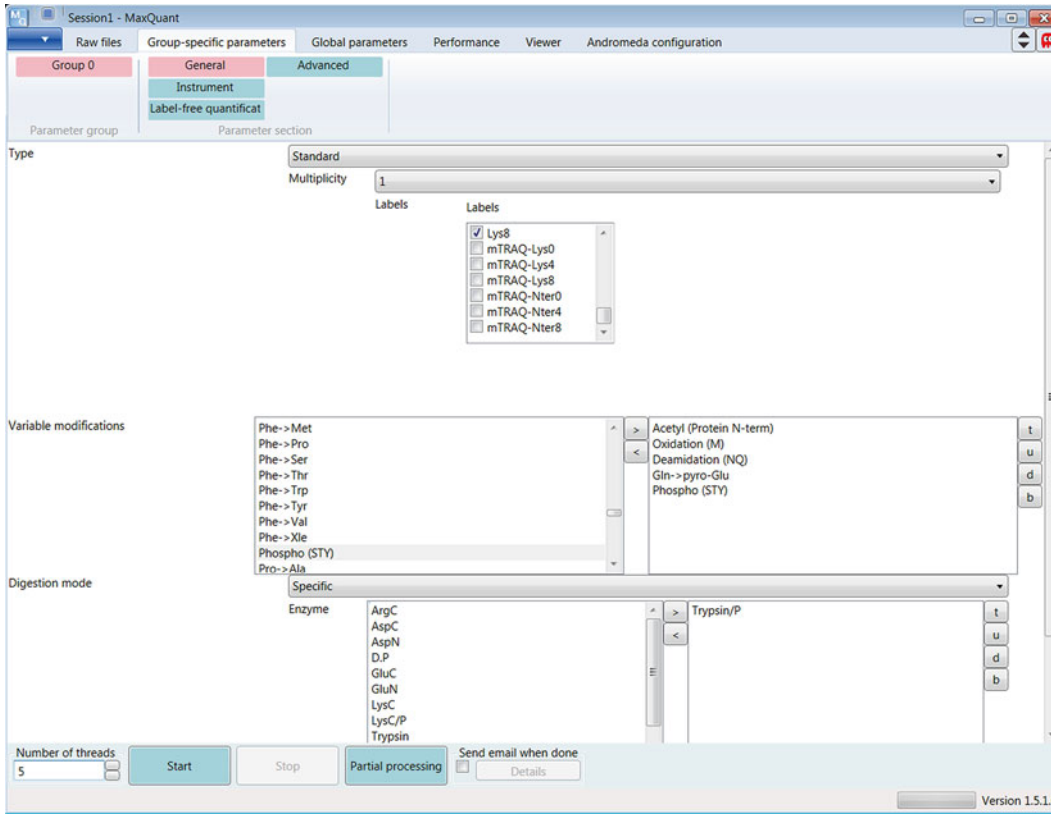

Fig. 2 Setting up a MaxQuant search. Screenshot of the "Group-specific parameters" tab

6. Now click on the "Global parameters" tab to define parameters that will affect all Raw files. Click on the "General button," then on "Add file," and in the pop-up window browse to the location of the proteome database we setup above. Leave the fixed modifications (i.e., modifications affecting all copies of compatible peptides; for details on modifications, *see* **Note 14**) at its default setting of "Carbamidomethyl (C)," which is the modification created on cysteine residues by the alkylating agent Iodoacetic acid (IAA) (*see* Chapter 20 (Bensaddek et al.); if your samples have instead been alkylated with *N*-ethylmaleimide (NEM), the most common alkylating agent beside IAA, *see* **Note 14**).
7. As this experiment involves SILAC labeling, tick the "requantify" option to ensure a ratio value is calculated even for peptides for which conjugated SILACpeptides could not be detected.
8. Tick "match between runs" to allow peptide identifications to be transferred between Raw files (Fig. 3; for more details on "match between runs," *see* **Note 15**).

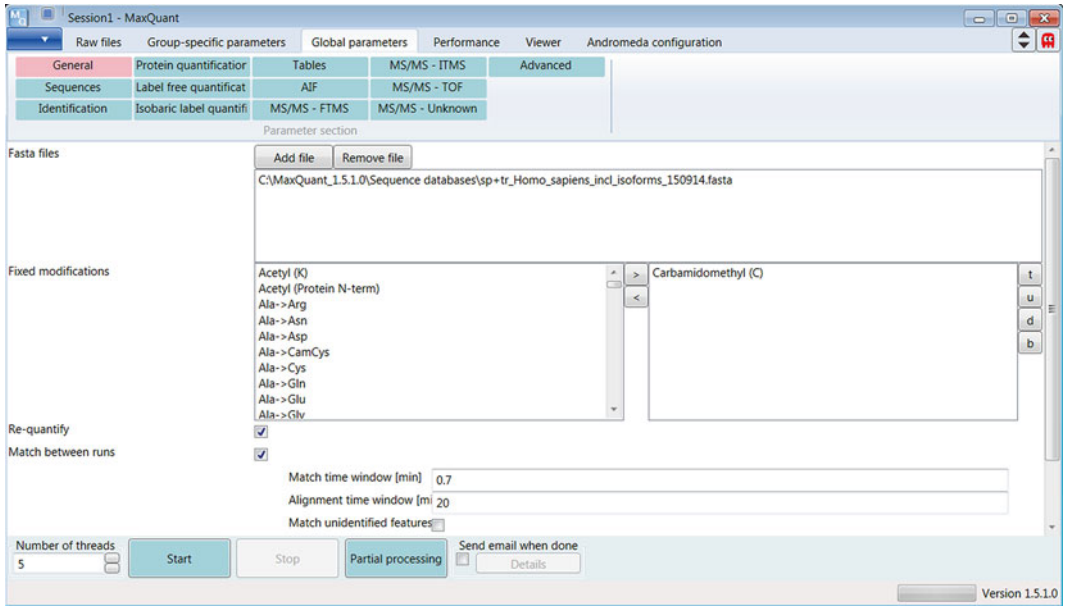

Fig. 3 Setting up a MaxQuant search. Screenshot of the "Global parameters" tab

9. Click on the "Sequences" button to add a common contaminants file for the first search. Browse to the "bin\conf" subfolders of the MaxQuant installation directory and select "contaminants.fasta."
10. At the bottom left of the window, select the number of threads you want the search to run on, which will depend on the hardware configuration of your computer. Never select more logical processors than are available on your computer. We advise leaving at least one processor free for other tasks. Thus, if 8 processors are available, select up to 6 or 7. The number of logical processors can be checked by running the Task Manager and navigating to the Performance tab.
11. Click on the arrow at the left of the tabs, select "Save parameters" and save them to a location of choice in case you need to either interrupt, or rerun, the search.
12. Hit the "Start" button at the bottom left of the window. The search can take anything from hours, to a few days to weeks, depending on computer performance, the number of Raw files and the parameters selected (especially the number of modifications searched). For example, a search on a set of 12 Raw files obtained from MS analysis of fractions from a single nucleolus preparation should be expected to complete within 12–24 h on a computer with average specifications. The progress of the ongoing analysis can be monitored in the "Performance" tab (for definitions of all terms, *see* **Note 1**).

3.3 Data Visualization in R

Once the search has completed, MaxQuant will write output tables to a "txt" subfolder. You will need the protein groups (proteinGroups.txt) file, which includes normalized SILAC ratios for each fraction (MaxQuant "experiment") and each protein group. We will demonstrate how to use normalized SILAC ratios to draw a volcano plot (x: $\log_2$ of normalized ratios; y: $-\log_{10}$ of p-values), using the R scripting language. R is freely available and is especially versatile because of the large number of packages provided and online support information that can be accessed.

1. Import the data in Rstudio, a graphical interface for R.

```
> options(stringsAsFactors = FALSE)
> setwd("" )
```

Use the "setwd" function to set the work directory to the folder where the MaxQuant output files are located (replace back slashes with forward slashes).

```
> PG <- read.delim("proteinGroups.txt")
```

2. Find the normalized heavy to light ratio columns for the nucleolus fraction (the text in quotes may have to be adjusted depending on the exact column names; here the text searches for a string starting with "Ratio.H.L.normalized.Nucleolus.", using regex syntax (*see* **Note 9**)):

```
> Ratios <- grep("^Ratio\\.H\\.L\\.normalized\\.Nucleolus\\.", colnames(PG))
```

To view the matches:

```
> grep("^Ratio\\.H\\.L\\.normalized\\.Nucleolus\\.", colnames(PG), value = TRUE)
[1] "Ratio.H.L.normalized.Nucleolus.1" "Ratio.H.L.normalized.Nucleolus.2"
"Ratio.H.L.normalized.Nucleolus.3"
```

3. Check that the ratios are normally distributed:

```
> PG.2 <- PG[, c(which(colnames(PG) == "Protein.IDs"), Ratios)]
> require(reshape)
```

The line above checks that the relevant package (here, "reshape") is installed and loads its library.

```
> PG.2 <- melt(PG.2)
```

Using Protein.IDs as id variables

```
> PG.2$value <- log10(PG.2$value)
> PG.2 <- PG.2[is.na(PG.2$value) == F,]
> require(ggplot2)
> plot <- ggplot(PG.2) + geom_density(aes(x = value, colour = variable))
> x11()
> print(plot)
```

4. Calculate a *p* value from the ratios:

```
> PG.3 <- PG[, c(which(colnames(PG) == "Protein.IDs"), Ratios)]
> PG.3[,2:(length(Ratios)+1)] <- apply(PG.3[,2:(length(Ratios)+1)], MARGIN =
2, FUN = function(x) {log2(x)})
> PG.3$P.value <- apply(PG.3[,2:(length(Ratios)+1)], MARGIN = 1, FUN = function(x) {
  x <- x[is.na(x) == F]
  if (length(x) > 1) {
    return(t.test(x, y = NULL, alternative = "two.sided")$p.value)
    } else {
    return(NA)
    }
})
```

5. Draw a volcano plot (Fig. 4):

```
> PG.3$log2.Ratio <- apply(PG.3[,2:(length(Ratios)+1)], MARGIN = 1, FUN =
function(x) {median(x, na.rm = T)})
> PG.3 <- PG.3[,c("Protein.IDs","P.value","log2.Ratio")]
> PG.3$minus.log10.P.value <- -log10(PG.3$P.value)
> PG.3 <- PG.3[is.na(PG.3$minus.log10.P.value) == F,]
> PG.3$Colour <- "non significant"
> PG.3$Colour[which(PG.3$minus.log10.P.value > -log10(0.05))] <- "p < 0.05,
abs(ratio) > 2"
> PG.3$Colour[which(PG.3$minus.log10.P.value > -log10(0.01))] <- "p < 0.01,
abs(ratio) > 2"
> PG.3$Colour[which((PG.3$log2.Ratio <= log2(2))&(PG.3$log2.Ratio >= -log2(2)))]
<- "non significant"
> myColors <- c("red","orange","black")
> names(myColors) <- c("p < 0.01, abs(ratio) > 2","p < 0.05, abs(ratio) >
2","non significant")
> colScale <- scale_colour_manual(name = "colour", values = myColors)
> plot <- ggplot(PG.3) + geom_point(aes(x = log2.Ratio, y = minus.
log10.P.value, colour = Colour)) +
  geom_vline(x = -log2(2), colour = "yellow") + geom_vline(x = log2(2),
colour = "yellow") +
  geom_hline(y = -log10(0.05), colour = "orange") + geom_hline(y =
-log10(0.01), colour = "red") +
  colScale
> x11()
> print(plot)
```

Et voilà!

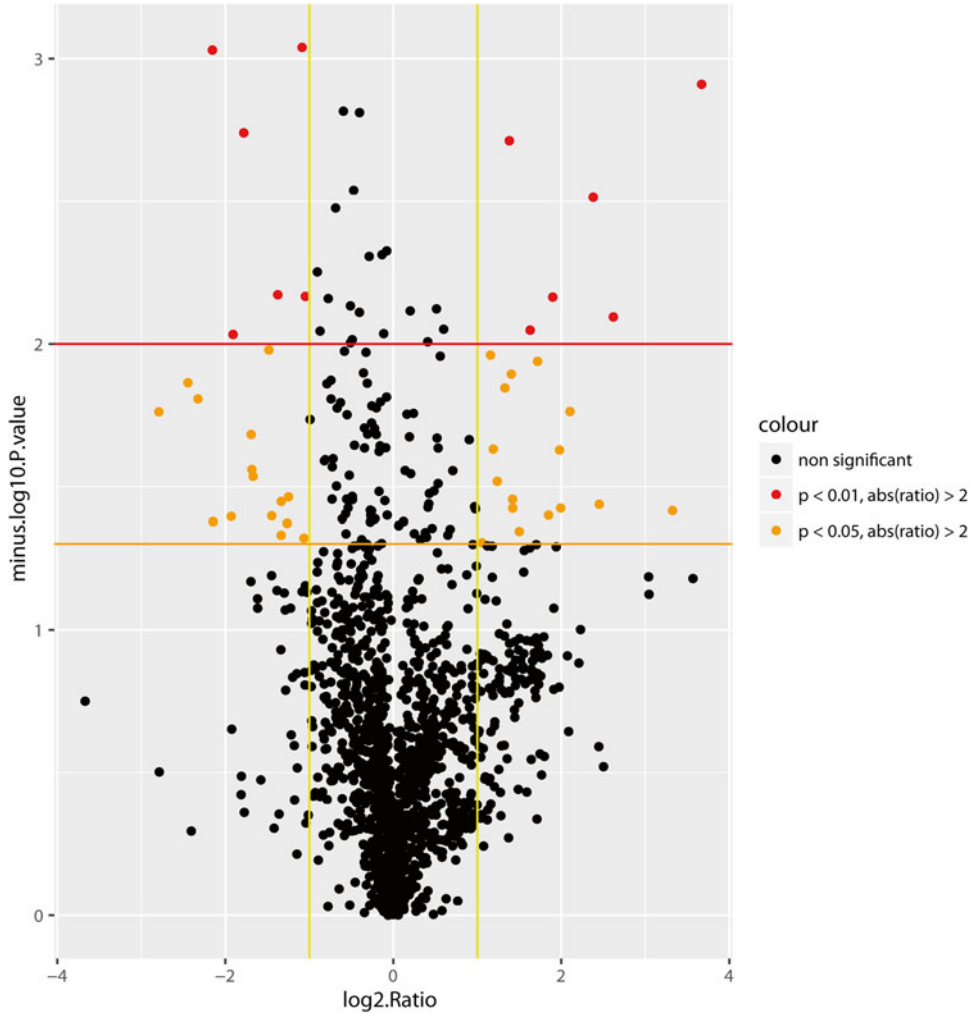

Fig. 4 Example of a volcano plot created using H/L SILAC ratios measured in three biological replicates. The $\log_2$ of median ratios were plotted on the *x*-axis and the $-\log_{10}$ of *P* values, calculated using a two-sided, one sample *T*-test, were plotted on the *y*-axis. Thus, proteins corresponding to points to the *left* and to the *right* of the *x*-axis, are decreased and increased after transformation, respectively. *P* values provide a measure of reproducibility between the replicate data sets. A threshold of +/− 2 fold change in protein abundance is typically used to identify significantly affected proteins (*vertical yellow lines* on plot), while those with highest statistical significance (*p* value > 0.01) are shown in *red*, while proteins of potential interest but lower significance (*p* value > 0.05) are shown in *orange*

4 Notes

1. Glossary:
   Duty cycle: The program which the mass spectrometer cycles through during the analysis run. Each duty cycle consists of one MS1 scan followed by (commonly) up to 10 MS2 scans. MS1 scans are full scans always covering the same large *M/Z*

range across the run and are used to monitor ions eluting from the LC and to identify candidate ions for fragmentation. After each MS1 scan, the peptides which are currently at their elution peak and have not recently been fragmentation are picked for isolation and fragmentation. MS2 spectra are scans of the results of the fragmentation of each peptide. On a Q-Exactive Plus, this duty cycle lasts about 1 s depending on the method used.

False Discovery Rate: Estimated number of false peptide discoveries for a given parameter. This provides a measure of the confidence that the peptide sequence identified is correct.

FASTA: A widely used, compact protein or nucleic acid sequence format where each entry consists of a header starting with ">," immediately followed by the sequence identifier and optional annotations on the same line, then, after a line break, by the molecule's sequence.

Modification: Posttranslational peptide modification; only modifications that are explicitly searched for will be detected. There are two types of modifications: fixed modifications, which are assumed to affect all copies of compatible peptides, and variable modifications, which will be searched for but will not be assumed to affect all copies of compatible peptides.

Peptide: Protein fragment, typically the result of enzymatic digestion; among mass spectrometrists, the term "peptide" in the singular form is often used to refer to all of the copies of a peptide sequence present in a sample.

Peptide–Spectrum Match (PSM): The way in which the peaks present in MS1 spectra are connected to a theoretical peptide sequence. In the method used in this example, up to 10 MS2 spectra are generated per second. An ideal MS2 spectrum would allow complete, de novo sequencing of the parent MS1 peak. However, MS2 spectra are often too complex or incomplete. The strategy for assigning a parent peptide sequence to observed MS2 spectra is called "Peptide–Spectrum Matching." Briefly, each MS2 spectrum is compared with all possible theoretically observable spectra from the database of expected proteins (typically, the proteome of the organism analyzed). Each peptide/spectrum pair is assigned a quality score. The MS2 spectrum, and its parent MS1 peak, are assigned the sequence of the best matching theoretical peptide.

Posterior Error Probability (PEP) score: Probability that a PSM with a given score is incorrect. This is a form of local FDR, i.e., the FDR for a given match score. During the data analysis, MaxQuant attempts to match the observed spectra not just to genuine predicted peptides, but also to a "decoy" database, which is usually made of sequence reversed versions of predicted peptides. These matches, which are known as

'false discoveries', are assigned a quality score just like real matches. The PEP score of a PSM with a given quality score is the ratio between the number of PSMs from the decoy database with the same quality score and the total number of PSMs with the same quality score.

PSM: See "Peptide–Spectrum Match."

Raw file: The mass spectrometer output file containing all of the spectra acquired in a single run.

Spectrum (plur. "spectra"): The output of MS analysis. A measure of current intensity at the detector as a function of mass over charge (M/Z), at a given time point in the run. In the common, data-dependent acquisition mode used in the example here, spectra can be MS1 or MS2 level scans (see "Duty cycle").

2. The latest version of MaxQuant (currently 1.5.3.30) can be downloaded, after registration, at the following link:
   http://141.61.102.17/maxquant_doku/doku.php?id=common:download_and_installation
3. Instructions for registering on the Thermo website and downloading the free MSFileReader software can be found at the following address:
   http://141.61.102.17/maxquant_doku/doku.php?id=common:msfilereader
4. The latest distribution of R can be downloaded from: https://cran.r-project.org/mirrors.html *Instructions for installation on Unix-family can be found at*: https://cran.r-project.org/doc/manuals/r-release/R-admin.html
5. The latest version of Rstudio can be downloaded at: https://www.rstudio.com/products/rstudio/download/
6. Additional, biology-related packages can be installed from Bioconductor (https://www.bioconductor.org/).
7. FASTA proteome files for common organisms can be downloaded from public sequence databases. The Human proteome can be obtained from UniProt at:
   http://www.uniprot.org/uniprot/?query=proteome:UP000005640
8. You have to click "Modify table" after each subsequent step, or changes will not be saved.
9. For more info on Regex, look at the following website: http://www.regular-expressions.info/; regular expressions can be tested at the following page: https://regex101.com/
10. By "experiments," MaxQuant understands different types of conditions within a broader experiment, for instance "treated" and "control" would be two different experiments. For a subcellular fractionation experiment, each subcellular fraction would be treated as a different "experiment." MaxQuant's "fractions" are the different fractions available for each "experiment," e.g., gel slices for an in-gel digest, or HILIC fractions.

When dealing with multiple repeats (i.e., biological and/or technical replicates) of a single experiment, it is advisable to perform the analysis in a single search to leverage the power of the 'match between runs' function (see below). In this case, assign each repeat a different "experiment" name (e.g., "nucleolus_repeat_1" and "nucleolus_repeat_2"). Note this should only be done when the replicates were performed using identical conditions.

11. Parameter groups are groups of Raw files that will be processed using specific parameters, as discussed immediately below.
12. Sometimes you may wish either to search for peptide modifications, or to digest with proteases that are not available in the MaxQuant default set. You would also create rules for these in Andromeda, using the "Modifications" or "Proteases."
13. Each additional modification included in the search will increase the search space exponentially and as a consequence increase the time taken to perform the search. Thus, we recommend not including more than a maximum of 6–7 variable modifications in a single search.
14. For samples alkylated with *N*-ethylmaleimide (NEM), you will have to create a dedicated modification in the "Andromeda configuration" tab. To do this, click on the "Andromeda configuration" tab, then click the "Modifications" button. Click "Add," click on the "Name" button and enter "NEM," then click the "Modify table" button. As before, you will have to do this after each subsequent modification. Click the "Description" tab and enter "Cys- > Cys-NEM." Click the "Composition" tab and enter "H(7) O(2) C(6) N," i.e., the number of atoms of each specific element added by the modification. Negative values can be entered for atoms removed from the unmodified peptide by the modification. Click successively on the "Position," "Type," and "New terminus" buttons and check that their values are set to "Anywhere," "Standard," and "None," respectively. Type on "Specificities," then on the "+" button next to the "Specificities" text, select "C" in the drop-down menu and click "OK." Click on the "K" field below "AA" and click on the "-" button above to remove the default lysine from the list of amino acids affected by the modification. Click on "Correction factors" and check that all values are set to 0 (default). Click again on each of the buttons where you made modifications to check that the modifications have been entered (don't forget to click on "Modify table" after each change). Click on the "Save changes" button.
15. It may happen that a given peptide only generated a good quality, assignable spectrum in some, but not all, of the runs in an experiment. MaxQuant is able to align the retention times from different runs and to correct for errors in mass measurements, so that it is possible to link peptide peaks across multiple

runs and to transfer identifications if required. This feature, called "match between runs," is particularly useful in subcellular fractionation experiments, as it can provide more complete cross-fraction profiles. Only use the "match between runs" feature when the different samples were analyzed on the same system under identical conditions.

Acknowledgments

A.N. is funded by the BBSRC sLoLa grant (BB/K003801/1). D. B. is funded by the Human Induced Pluripotent Stem Cells Initiative (098503/E/12/Z). This work was funded by a Wellcome Trust Programme grant to A. I. L (108058/Z/15/Z) and supported by a Wellcome Trust Strategic award (105024/Z/14/Z).

References

1. Soule HD, Maloney TM, Wolman SR, Peterson WD, Brenz R, McGrath CM, Russo J, Pauley RJ, Jones RF, Brooks SC (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res 50(18):6075–6086
2. Iliopoulos D, Hirsch HA, Struhl K (2009) An epigenetic switch involving NF-kB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. Cell 139(4):693–706
3. Hirsch HA, Iliopoulos D, Tsichlis PN, Struhl K (2009) Metformin selectively targets cancer stem cells, and acts together with chemotherapy to block tumor growth and prolong remission. CancerRes69(19):7507–7511.doi:10.1158/0008-5472.can-09-2994
4. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372
5. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10(4):1794–1805

INDEX

Attila Németh (ed.), *The Nucleolus: Methods and Protocols*, Methods in Molecular Biology, vol. 1455,
DOI 10.1007/978-1-4939-3792-9, © Springer Science+Business Media New York 2016

T

U

W

Y

Z