ORIGINAL INVESTIGATION

CrossMark

3C-PCR: a novel proximity ligation-based approach to phase chromosomal rearrangement breakpoints with distal allelic variants

Samantha L. P. Schilit[1,2,3] · Cynthia C. Morton[4,5,6]

Abstract

Recent advances in molecular cytogenetics highlight the importance of noncoding structural variation in human disease. Genomic rearrangements can disrupt chromatin architecture, leading to long-range alterations in gene expression. With increasing ability to assess distal gene dysregulation comes new challenges in clinical interpretation of rearrangements. While haplotyping methods to determine compound heterozygosity in a single gene with two pathogenic variants are established, such methods are insufficient for phasing larger distances between a pathogenic variant and a genomic rearrangement breakpoint. Herein, we present an inexpensive and efficient proximity ligation-based method called 3C-PCR for phasing chromosomal rearrangement breakpoints with distal allelic variants. 3C-PCR uses canonical chromosome conformation capture (3C) libraries for targeted distal phasing by implementing a novel nested PCR strategy with primers anchored across the rearrangement breakpoints and subsequent Sanger sequencing. As a proof of concept, 3C-PCR was used to phase a highly variable region 1.3 Mb upstream of a chromosomal rearrangement breakpoint in a balanced translocation. We found that the nested PCR approach amplified the derivative chromosome substrate exclusively and identified the same haplotype by Sanger sequencing reliably. Given its efficacy and versatility, 3C-PCR is ideal for use in phasing chromosomal rearrangement breakpoints with allelic variants located at a genomic distance over a megabase.

✉ Cynthia C. Morton
   cmorton@bwh.harvard.edu

1   Biological and Biomedical Sciences Program, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA

2   Program in Genetics and Genomics, Department of Genetics, Harvard Medical School, Boston, MA, USA

3   Leder Human Biology and Translational Medicine Program, Division of Medical Sciences, Harvard Medical School, Boston, MA, USA

4   Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA

5   Division of Evolution and Genomic Science, School of Biological Sciences, Manchester Academic Health Science Centre, Manchester, UK

6   Departments of Obstetrics and Gynecology and of Pathology, Brigham and Women's Hospital and Harvard Medical School, New Research Building, Room 160D, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

Introduction

In the past two decades, efforts to annotate the human genome have revealed a significant functional role for non-coding sequences. Genomic structural variations, such as copy-number variants and genomic rearrangements, have been shown to lead to genomic disorders (Stankiewicz and Lupski 2010). Many of these variants result in an abnormal phenotype by altering long-range control of gene expression (Kleinjan and van Heyningen 2005). This is mediated by the disruption of topologically associated domains (TADs) and subsequent promiscuous enhancer–promoter interactions that lead to pathogenic misexpression (Lettice et al. 2011; Lupiáñez et al. 2015; Redin et al. 2017). Given the clinical significance of long-range *cis* regulatory mutations, recent research has focused on predicting clinical outcomes for subjects with structural chromosomal rearrangements by considering dysregulation of genes that reside in the disrupted TADs (Ordulu et al. 2016; Zepeda-Mendoza et al. 2017).

If a dysregulated gene is associated with an autosomal recessive disease phenotype and subsequent sequencing of the gene reveals a second pathogenic variant, phasing is critical for clinical interpretation. While variants in *cis* may

not manifest in the disease phenotype, variants that reside in *trans* result in a compound heterozygote (Duzkale et al. 2013). The vast difference in clinical interpretation highlights a critical need for a method capable of deciphering large haplotypes across derivative chromosomes. There is great interest in applying this technology to de novo balanced chromosomal abnormalities (BCAs), because long-range position effects explain clinical phenotypes in a substantial proportion of subjects with BCAs (Redin et al. 2017).

While computational and experimental phasing has been used to identify haplotypes since the 1980s, current methods are insufficient to resolve a haplotype that spans megabase distances on derivative chromosomes, as requisite for a TAD-disrupting chromosomal rearrangement (Browning and Browning 2011). Computational haplotype phasing, which relies on genotype data from unrelated individuals using statistical approaches or from families using identity by descent (IBD), cannot be applied to nonrecurrent genomic rearrangements because they are not common in the population or may not be inherited (Browning and Browning 2011). While experimental techniques such as long-range polymerase chain reaction (PCR), Drop-Phase, and targeted locus amplification (TLA) do not require population or family genotyping data, they are limited by genomic distance, losing efficacy beyond 30, 200, and 400 kb, respectively (de Vree et al. 2014; McDonald et al. 2002; Regan et al. 2015). Other technologies that physically separate chromosomes before genotyping, such as by microdissection using a computer-directed laser beam or by dispersion using a microfluidic device, may span large enough distances (Fan et al. 2011; Ma et al. 2010); however, these techniques require specialized equipment and are labor intensive making them difficult to apply broadly. Even experimental techniques with straightforward protocols that can easily be translated to other laboratories, like HaploSeq, are still limiting in that they are costly and require substantial computational expertise due to the cost and subsequent analysis of next-generation sequencing (Selvaraj 2013).

In this study, we developed 3C-PCR, an inexpensive and efficient proximity ligation-based approach to phase chromosomal rearrangement breakpoints with distal allelic variants. Our method adapts the use of canonical chromosome conformation capture (3C) libraries by employing a novel nested PCR strategy with primers anchored across the rearrangement breakpoints and subsequent Sanger sequencing (Dekker et al. 2002). 3C has become a widely used method that can be performed in a matter of days using standard molecular biology equipment, and PCR and Sanger sequencing are routine in diagnostic laboratories (Miele et al. 2006). By combining these simple and accessible methods, 3C-PCR makes possible phasing variants at a distance of over a megabase from a chromosomal rearrangement without the expense of specialized equipment, next-generation sequencing or extensive computational analysis.

Materials and methods

Acquisition of lymphoblastoid cell lines

Subjects DGAP230, with 46,XY,t(20;22)(q13.3;q11.2), and DGAP278-02, a karyotypically normal age- and sex-matched control, were enrolled through the Developmental Genome Anatomy Project (DGAP, dgap.harvard.edu). DGAP obtained informed consent, medical records and blood samples under a protocol approved by the Partners HealthCare Systems Institutional Review Board. Epstein–Barr virus-transformed lymphoblastoid cell lines (LCLs) were generated at the Genomics and Technology Core in the Center for Human Genetic Research at Massachusetts General Hospital (Boston, MA, USA). Large-insert ("jumping library") whole-genome sequencing and subsequent Sanger sequencing identified the precise breakpoints of the DGAP230 chromosomal rearrangement as previously described and reported (Hanscom and Talkowski 2014; Redin et al. 2017; Talkowski et al. 2011). Two additional karyotypically normal age- and sex-matched control LCLs, GM20184 and GM20188, were obtained from the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Institute for Medical Research (Camden, NJ, USA).

Identification of a variable region on chromosome 20

TADs disrupted by the breakpoints in DGAP230 were identified according to human embryonic stem cell Hi-C domains from the Hi-C project (Dixon et al. 2012). The University of California Santa Cruz Genome Browser was used to delineate regions located over a megabase away from the t(20;22) breakpoints within the same TAD (Rosenbloom et al. 2015). These sequences were compared against the Database of Single Nucleotide Polymorphisms (dbSNP) to identify highly variable regions in the distal TAD-residing sequences (Sherry et al. 2001).

To assess heterozygosity of these candidate regions in DGAP230 and control LCLs, genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). PCR was performed using LongAmp Taq 2X Master Mix (New England Biolabs, [NEB]) and customized primers [Integrated DNA Technologies (IDT)] designed to amplify potential variable regions. After amplification confirmation with agarose gel electrophoresis, Sanger sequencing reactions of PCR products were carried out with an ABI3730xl DNA analyzer. Chromatograms were aligned and multiple single

nucleotide variants were called using Geneious (version 7.0, Biomatters). A target region was selected based upon the presence of several single nucleotide variants in the chromatograms for all experimental and control samples.

Generation of 3C libraries

3C libraries were generated as previously described (Dekker et al. 2002; Gheldof et al. 2012; Miele et al. 2006; Splinter et al. 2012; van de Werken et al. 2012). In brief, 10 million cell aliquots of LCLs were crosslinked with 2% formaldehyde (Sigma-Aldrich) and lysed. Chromatin was digested with *Hin*dIII-HF (NEB), ligated with T4 DNA ligase (NEB) and reverse crosslinked by incubation with Proteinase K (NEB) and RNase A (EMD Millipore). DNA libraries were purified by phenol/chloroform/IAA extraction (Sigma-Aldrich), MaXtract High Density Tubes (Qiagen) and subsequent ammonium acetate precipitation (Sigma-Aldrich). 3C libraries were generated in triplicate, with three independent cultures for the DGAP230 LCL and three different control LCLs.

Design of primers for nested PCR approach

Primer design was adapted from 3C protocols, but with adjustments to accommodate target regions further away than 80–150 bp from the restriction enzyme digestion site and PCR amplicons longer than 160–300 bp, as previously described (Miele et al. 2006). Sequences were obtained for two predicted *Hin*dIII-digested fragments: one with the target region on chr20, and a second containing the sequence on chr22 most proximal to the der(20) breakpoint. A synthetic sequence of a potential ligation product from these two fragments was designed in SeqBuilder (version 14.1.0.118, DNASTAR) by concatenating the two sequences at their respective *Hin*dIII restriction sites. Primers spanning both fragments and the target variable region were designed in Primer3Plus and assessed for sequence specificity using BLAT (Kent 2002; Untergasser et al. 2007). Nested primer pairs were designed such that one primer pair flanked the entire substrate recognized by the second primer pair.

## Rearrangement-specific amplification and sequencing

Nested PCRs of breakpoint-spanning fragments were performed using LongAmp Taq 2X Master Mix (NEB). The first PCR reaction amplified ~ 300 ng of 3C libraries for all experimental and control samples using the outer primer pair and thermocycling conditions including a long extension time and low annealing temperature [3 min at 94 °C, 35 cycles × (30 s at 94 °C, 30 s at 56 °C, 2.5 min at 65 °C), 10 min at 65 °C, hold 4 °C]. Amplicons were purified using a QIAquick PCR purification kit (Qiagen). After quantification, ~ 100 ng of purified amplicons were used as substrates for a second PCR reaction using the inner primer pair and more stringent conditions with a shorter extension time and higher annealing temperature [3 min at 94 °C, 45 cycles × (30 s at 94 °C, 2 min at 65 °C), 10 min at 65 °C, hold 4 °C]. Nested PCR amplicon specificity was evaluated using agarose gel electrophoresis. Amplicons were purified using a QIAquick PCR purification kit (Qiagen) and Sanger sequenced with an ABI3730xl DNA analyzer using the same sequencing primer as used for the genomic DNA samples. 3C-PCR chromatograms were aligned to genomic DNA chromatograms for comparison and nucleotide variants were called using Geneious (version 7.0, Biomatters).

Results

To develop an assay capable of phasing allelic variants over a megabase away from a breakpoint of a chromosomal rearrangement within the same TAD, we searched for an LCL that has a BCA with at least one breakpoint located over a megabase away from a TAD boundary. Through DGAP, we selected the DGAP230 LCL, with 46,XY,t(20;22)(q13.3;q11.2) and a distance of more than 1.4 Mb between the chromosome 20 (chr20) breakpoint and the upstream boundary of the TAD in which it resides (Fig. 1a) (Redin et al. 2017). To ensure assay specificity, we also selected three karyotypically normal age- and sex-matched control LCLs: DGAP278-02, GM20184 and GM20188. As a source for allelic variation, we identified a highly variable region 1.3 Mb upstream of the chr20 breakpoint. Sanger sequencing of this target region showed heterozygosity at several bases in DGAP230 as well as in all control cell lines (Fig. 1b).

We next set out to develop a method capable of determining the haplotype of the target variable region on the derivative chromosome 20 (der(20)). If the target region and chr20 breakpoint were located only a few kb apart, phasing could be accomplished by selectively amplifying the der(20) allele using primers that span the translocation junction to produce an amplicon containing the target region in *cis*, which could be assessed by Sanger sequencing. However, the 1.3 Mb distance between the breakpoint and the target region render this strategy unsuccessful, because PCR performs at distances three orders of magnitude smaller. To overcome this technical challenge, we developed a strategy called 3C-PCR. This method capitalizes on principles underlying 3C technologies developed by Dekker and Kleckner in 2002, which show that when crosslinked DNA is enzymatically digested into genomic fragments and then ligated to other fragments in close physical proximity, sequences in *cis* have a higher interaction frequency than those in *trans* (Dekker et al. 2002;
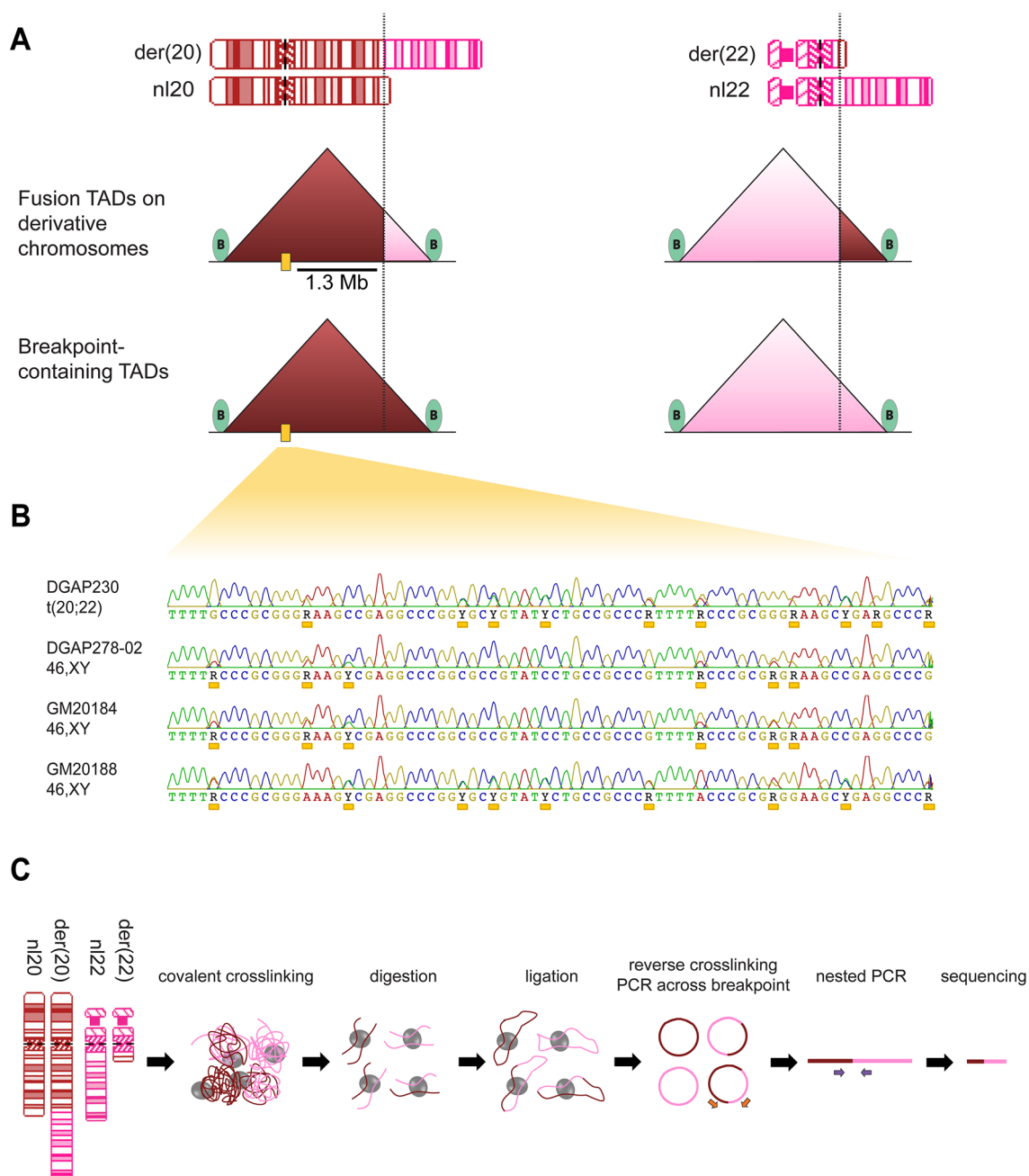
Fig. 1 Experimental system. **a** The lymphoblastoid cell line, designated DGAP230, has a balanced translocation (top) between the long (q) arms of chromosomes 20 (mahogany color) and 22 (light pink color). Translocation breakpoints reside near the boundaries (green "B" circles) of predicted TADs (triangular shapes), enabling assessment of a distal region with multiple single nucleotide variants (yellow box) within the same chromatin loop (bottom). **b** Chromatograms from Sanger sequencing of the target region reveal a highly variable region in DGAP230 and control cell lines. Single nucleotide variants are indicated by a small orange box below the correspond-

ing nucleotide ($R = A/G$; $Y = C/T$). **c** In 3C-PCR, coupling proximity ligation with breakpoint-spanning nested PCR can capture *cis* sequences distant from the chromosomal rearrangement. Chromatin conformation capture libraries are generated by covalent crosslinking of chromatin, enzymatic digestion and ligation of proximal genomic fragments to bring high-frequency three-dimensional interactions into two-dimensional linear space. Reverse crosslinked ligation products are then subjected to two rounds of nested PCR to select for specific amplicons that cross the breakpoint junction and include the *cis* target region for subsequent Sanger sequencing (color figure online)

Denker and de Laat 2016). We hypothesized that we could use 3C to bring fragments containing the translocation junction and der(20) target region closer together, thus enabling PCR across the junction of a ligation product

including the *cis* target region. Given the strong possibility of amplifying nonspecific sequences from a complex 3C library with diverse ligation products, we pursued a nested PCR step to improve specificity (Fig. 1c) (Dekker 2006).

Using the predicted ligation product as a substrate, we designed nested primers that would span the target region on chr20, the enzymatic digestion and ligation site, and the chr22 genomic fragment near the breakpoint (Fig. 2a, b; Supplemental Table S1). As expected, the first amplification resulted in several nonspecific PCR products for all DGAP230 and control LCL 3C libraries (Fig. 2c). However, after performing nested PCR on products purified from the first amplification, we produced DNA fragments of predicted size from all DGAP230 samples but from none of the controls, suggesting that nested PCR recognized the predicted proximity ligation product from the *cis*-interacting der(20) chromosome present only in DGAP230 samples. As evidence that the predicted proximity ligation product is the substrate for amplification, nested PCR on negative control genomic libraries without crosslinking, digestion or ligation yielded no PCR-amplified products (Supplemental Fig.

S1a). Additionally, *Hin*dIII digestion and subsequent agarose gel electrophoresis of the amplicon from the DGAP230 3C library-nested PCR confirmed derivation from the predicted ligation product (Supplemental Fig. S1b-c). Sequencing of all three amplicons revealed a single identical sequence, providing evidence that this is the haplotype of the target region on der(20) (Fig. 2d).

Discussion

We present 3C-PCR, an inexpensive and efficient proximity ligation-based approach to phase chromosomal rearrangement breakpoints with distal allelic variants. We anticipate that the simplicity of this approach will expedite its adoption in future clinical practice to determine compound
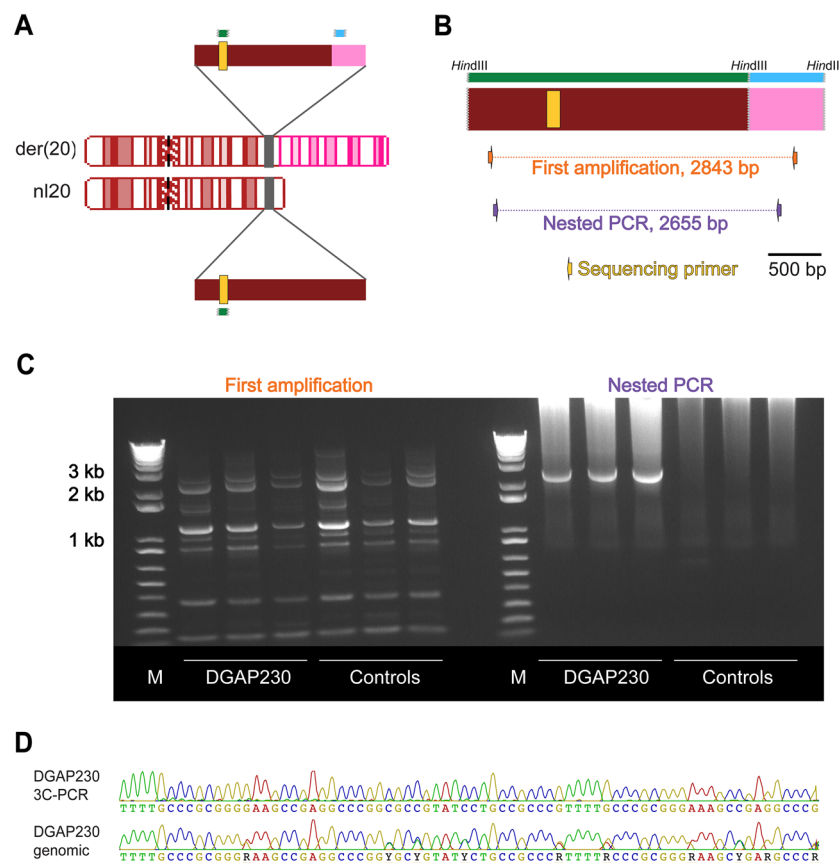
Fig. 2 Assay validation. **a** The goal of the assay in the DGAP230 experimental system is to differentiate the target region (yellow box) on the der(20) chromosome (top) from the target region on the normal chr20 (bottom). The small green bar represents the 3C genomic fragment that contains the target region, and the small blue bar represents the digested genomic fragment containing a breakpoint-proximal region from the segment of chr22 translocated to the der(20). Rough gray edges reflect enzymatic digestion at flanking *Hin*dIII restriction sites. **b** Schematic of nested PCR amplifications

for the predicted ligation product with the target region (green bar above mahogany rectangle) and the chr22 fragment (blue bar above light pink rectangle). **c** Gel electrophoresis displays products from the first PCR across the breakpoint for experimental and control 3C libraries (left), and the second nested PCR (right, $N = 3$). Key DNA fragment sizes of the markers (M) are indicated on the left. **d** Sanger sequencing traces of the target variable region from the nested PCR amplicon (top) and genomic DNA from the same cell line (bottom; $N = 3$) (color figure online)

heterozygosity in cases where a gene dysregulated by a disrupted TAD harbors a second pathogenic variant.

3C-PCR serves as a novel application to the widely used 3C method and differentiates itself from other adaptions of 3C in its ease, technical capabilities and versatility (Dekker et al. 2002). 3C-PCR targets the allele of a variable locus in *cis* with a chromosomal rearrangement on a derivative chromosome by a simple nested PCR strategy on 3C libraries, eliminating the need for costly and time-consuming next-generation sequencing and computational analysis used in other proximity ligation-based phasing methods (de Vree et al. 2014; Selvaraj 2013). In addition, these other phasing methods are also technically inferior to 3C-PCR, in that HaploSeq has a sparse ascertainment density resulting in less than a 25% chance of detecting the distal allelic variant of interest as opposed to 100% for 3C-PCR, and TLA can only haplotype distances of up to 300 kb, less than a third of the capabilities of 3C-PCR (Snyder et al. 2015).

In our system, nonspecific amplification of 3C libraries is ameliorated by a two-step nested PCR. This differs from standard PCR of 3C libraries to determine semi-quantitative interaction frequencies, because primers can be designed to flank closely the restriction enzyme digestion sites of the two genomic fragments in question, allowing for short PCR extension times that select for a small 160–300 bp amplicon (Miele et al. 2006). In our assay, resulting amplicons must include the target region residing anywhere in the enzymatically digested genomic fragments (e.g., at a distance of 2 kb, when considering that restriction endonucleases with six-base pair recognition sequences produce genomic fragments about 4 kb in size). Our optimized nested PCR strategy compensates for the nonspecific amplicons produced from longer extension times. The first PCR amplifies all possible products, with conditions including a long extension time and low annealing temperature. To prevent biased overamplification of certain products, the number of cycles allows for amplification within the linear range. The subsequent nested PCR applies more stringent conditions with a shorter extension time and a much higher annealing temperature to select for the specific amplicon of interest. Additional cycles are used to compensate for the less efficient PCR.

Of note, this technique relies on the assumption that sequences in *cis* will have higher interaction frequencies than those in *trans*. While ligation products containing the *trans* target region and the breakpoint-proximal fragment would be much less common, they may still be present. To alleviate these concerns, PCR products detected in the DGAP230 cell line with the *t*(20;22) substrate are expected more frequently than in karyotypically normal cells. Indeed, our results identified an amplicon of the predicted size from the nested PCR in three independent 3C libraries performed on the experimental cell line and no products in three different 3C libraries derived from karyotypically normal LCLs

(Fig. 2c). Sanger sequencing of the same haplotype in all three replicates provides evidence of detection of the higher-frequency *cis* interaction event (Fig. 2d).

Our novel method does have some limitations. 3C-PCR targets a specific region, so customized primers must be designed and synthesized to probe the region of interest. The breakpoint of interest must also be resolved to near-nucleotide resolution (on the order of a couple kilobases), as is done by mate-pair or large-insert jumping libraries, to identify a genomic region known to reside on the derivative chromosome close to the breakpoint. If breakpoint information is only available at the resolution level of a karyotype, 3C-PCR will be successful if (1) there is a genomic region known with certainty to reside in *cis* with the breakpoint and (2) if this region is less than 30 Mb away from the allelic variant, as a higher interaction frequency for *cis* sequences compared to *trans* sequences persists for genomic distances of up to 30 Mb in proximity ligation assays (only ~ 0.6% for *trans* interactions, but increasingly to 2% at larger distances) (Selvaraj 2013). This strong bias for *cis* interactions also provides versatility in 3C-PCR, as indels, which may alter genomic distances on the order of 1–10,000 bp between the breakpoint and the allelic variant, would not significantly influence interaction frequencies (Mills et al. 2011). Similarly, due to this long-spanning *cis* interaction bias relative to the 880 kb median size of TADs, the variant of interest is not required to reside in the same TAD as the rearrangement breakpoint (Dixon et al. 2012).

Due to dependence of this technology on discriminating *cis* versus *trans* by proximity ligation, 3C-PCR will inherently work better for balanced translocations than for balanced inversions, in which both sides of the breakpoint derive from the same chromosome. The efficacy will depend on the difference in interaction frequency of the breakpoint-proximal genomic region and the variant of interest on the inverted and normal chromosomes, which will be affected by many factors including linear distance and the presence of TADs, enhancer–promoter interactions and insulator elements (Denker and de Laat 2016).

Due to the requirement to make proximity ligation libraries, another limitation is that 3C-PCR requires intact chromatin from tissue or cultured cells. Finally, the assay is also dependent upon successful PCR, which may be impacted by the specific ligation product's GC or AT content, predicted secondary structure or length. However, these limitations are less prohibitive than other technologies capable of phasing at distances over a megabase, including targeted haplotyping by dilution, single-chromosome sequencing and HaploSeq, all of which are labor intensive and require next-generation sequencing (Kaper et al. 2013; Ma et al. 2010; Selvaraj 2013; Snyder et al. 2015). 3C-PCR can phase distal variants with low cost and limited labor, using standard molecular biology reagents and equipment. As clinical diagnostic laboratories

enter the era of "next-gen cytogenetics", determining allelic nucleotide variant(s) of the sequence of a gene dysregulated by a structural chromosomal rearrangement will become essential. In these cases, 3C-PCR will be integral to clinical interpretation and prediction of disease phenotypes.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Research involving human participants and/or animals All procedures performed in studies involving human participants were in accordance with the ethical standards of the Partners HealthCare Systems Institutional Review Board and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent, medical records and blood samples from DGAP230 and DGAP278-02 were obtained through the Developmental Genome Anatomy Project (DGAP) protocol approved by the Partners HealthCare Systems Institutional Review Board.

References

Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. Nat Rev Genet 12:703–714. https://doi.org/10.1038/nrg3054

de Vree PJ et al (2014) Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nat Biotechnol 32:1019–1025. https://doi.org/10.1038/nbt.2959

Dekker J (2006) The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods 3:17–21. https://doi.org/10.1038/nmeth823

Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295:1306–1311. https://doi.org/10.1126/science.1067799

Denker A, de Laat W (2016) The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev 30:1357–1382. https://doi.org/10.1101/gad.281964.116

Dixon JR et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376–380. https://doi.org/10.1038/nature11082

Duzkale H et al (2013) A systematic approach to assessing the clinical significance of genetic variants. Clin Genet 84:453–463. https://doi.org/10.1111/cge.12257

Fan HC, Wang J, Potanina A, Quake SR (2011) Whole-genome molecular haplotyping of single cells. Nat Biotechnol 29:51–57. https://doi.org/10.1038/nbt.1739

Gheldof N, Leleu M, Noordermeer D, Rougemont J, Reymond A (2012) Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. Methods Mol Biol 786:211–225. https://doi.org/10.1007/978-1-61779-292-2_13

Hanscom C, Talkowski M (2014) Design of large-insert jumping libraries for structural variant detection using illumina sequencing. Curr Protoc Hum Genet 80:7–22, 21–29. https://doi.org/10.1002/0471142905.hg0722s80

Kaper F et al (2013) Whole-genome haplotyping by dilution, amplification, and sequencing. Proc Natl Acad Sci USA 110:5552–5557. https://doi.org/10.1073/pnas.1218696110

Kent WJ (2002) BLAT–the BLAST-like alignment tool. Genome Res 12:656–664. https://doi.org/10.1101/gr.229202

Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76:8–32. https://doi.org/10.1086/426833

Lettice LA et al (2011) Enhancer-adoption as a mechanism of human developmental disease. Hum Mutat 32:1492–1499. https://doi.org/10.1002/humu.21615

Lupiáñez DG et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161:1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Ma L et al (2010) Direct determination of molecular haplotypes by chromosome microdissection. Nat Methods 7:299–301. https://doi.org/10.1038/nmeth.1443

McDonald OG, Krynetski EY, Evans WE (2002) Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation. Pharmacogenetics 12:93–99

Miele A, Gheldof N, Tabuchi TM, Dostie J, Dekker J (2006) Mapping chromatin interactions by chromosome conformation capture. Curr Protoc Mol Biol 21:21-11. https://doi.org/10.1002/0471142727.mb2111s74

Mills RE et al (2011) Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res 21:830–839. https://doi.org/10.1101/gr.115907.110

Ordulu Z et al (2016) Structural chromosomal rearrangements require nucleotide-level resolution: lessons from next-generation sequencing in prenatal diagnosis. Am J Hum Genet 99:1015–1033. https://doi.org/10.1016/j.ajhg.2016.08.022

Redin C et al (2017) The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet 49:36–45. https://doi.org/10.1038/ng.3720

Regan JF et al (2015) A rapid molecular approach for chromosomal phasing. PLoS One 10:e0118270. https://doi.org/10.1371/journal.pone.0118270

Rosenbloom KR et al (2015) The UCSC Genome Browser database: 2015 update. Nucl Acid Res 43:D670–D681. https://doi.org/10.1093/nar/gku1177

Selvaraj S (2013) J RD, Bansal V, Ren B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing Nat Biotechnol 31:1111–1118. https://doi.org/10.1038/nbt.2728

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. Nucl Acids Res 29:308–311

Snyder MW, Adey A, Kitzman JO, Shendure J (2015) Haplotype-resolved genome sequencing: experimental methods and applications. Nat Rev Genet 16:344–358. https://doi.org/10.1038/nrg3903

Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W (2012) Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58:221–230. https://doi.org/10.1016/j.ymeth.2012.04.009

Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61:437–455. https://doi.org/10.1146/annurev-med-100708-204735

Talkowski ME et al (2011) Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements

for clinical diagnostics and genetic research. Am J Hum Genet 88:469–481. https://doi.org/10.1016/j.ajhg.2011.03.013

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucl Acid Res 35:W71–W74. https://doi.org/10.1093/nar/gkm306

van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W (2012) 4C technology: protocols and data analysis. Methods Enzymol 513:89–112. https://doi.org/10.1016/B978-0-12-391938-0.00004-5

Zepeda-Mendoza CJ et al (2017) Computational prediction of position effects of apparently balanced human chromosomal rearrangements. Am J Hum Genet. https://doi.org/10.1016/j.ajhg.2017.06.011