Published online 18 May 2018

Nucleic Acids Research, 2018, Vol. 46, No. 14 ***e87***
doi: 10.1093/nar/gky411

Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH)

Tslil Gabrieli[1,†], Hila Sharim[1,†], Dena Fridman[1], Nissim Arbib[2], Yael Michaeli[1,*] and Yuval Ebenstein[1,*]

[1]School of Chemistry, Center for Nanoscience and Nanotechnology, Center for Light-Matter Interaction, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel and [2]Department of Obstetrics and Gynecology, Meir Hospital, Kfar Saba, Israel & Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Received November 19, 2017; Revised April 12, 2018; Editorial Decision May 01, 2018; Accepted May 04, 2018

ABSTRACT

Next generation sequencing (NGS) is challenged by structural and copy number variations larger than the typical read length of several hundred bases. Third-generation sequencing platforms such as single-molecule real-time (SMRT) and nanopore sequencing provide longer reads and are able to characterize variations that are undetected in NGS data. Nevertheless, these technologies suffer from inherent low throughput which prohibits deep sequencing at reasonable cost without target enrichment. Here, we optimized Cas9-Assisted Targeting of CHromosome segments (CATCH) for nanopore sequencing of the breast cancer gene BRCA1. A 200 kb target containing the 80 kb BRCA1 gene body and its flanking regions was isolated intact from primary human peripheral blood cells, allowing long-range amplification and long-read nanopore sequencing. The target was enriched 237-fold and sequenced at up to 70× coverage on a single flow-cell. Overall performance and single-nucleotide polymorphism (SNP) calling were directly compared to Illumina sequencing of the same enriched sample, highlighting the benefits of CATCH for targeted sequencing. The CATCH enrichment scheme only requires knowledge of the target flanking sequence for Cas9 cleavage while providing contiguous data across both coding and non-coding sequence and holds promise for characterization of complex disease-related or highly variable genomic regions.

INTRODUCTION

BRCA1 is one of the major breast and ovarian cancer susceptibility genes. Lifetime risk of breast cancer for a female carrier of BRCA1 mutation is >80% (1) and over 1800 distinct mutations have been reported for BRCA1 in the breast cancer information core (BIC) database (2). The genomic region of BRCA1 contains ∼50% repetitive DNA elements that contribute to genetic instability and genomic rearrangements in this area (3). The increasing demand for BRCA1 gene profiling has led to significant development in genomic enrichment and targeted sequencing of this locus. Most methods are based on PCR amplification or hybridization capture approaches, followed by Sanger or next generation sequencing (NGS). However, due to the large size of this gene (∼80 000 bp), current enrichment methods focus on the coding sequence, while neglecting intronic and regulatory regions. An exception is a method that uses biotinylated RNA probes transcribed from fragmented bacterial artificial chromosomes (BACs) for capture of long targets containing both coding and non-coding regions (4). This method requires a compatible BAC clone, and complex regions such as BRCA1 that do not match the BAC-derived bait may be lost. Furthermore, although most common mutations have a small size that can be characterized by ultra-deep NGS, mutations also include large rearrangements, deletions and duplications of large genomic regions as well as repetitive elements, which are not accessible to NGS (5,6). In fact, recent SMRT sequencing of a haploid human genome found that 89% of discovered variants have been missed by the 1000 Genomes Project (7). Long read nanopore sequencing may be an ideal tool for characterization of such large structural variations, given there is sufficient coverage to overcome its error rate. Recently, a whole human genome assembly using nanopore sequencing was reported (8). However, due to the limited throughput, 39

*To whom correspondence should be addressed. Tel: +972 3 640 8901; Fax: +972 3 640 5794; Email: uv@post.tau.ac.il
Correspondence may also be addressed to Yael Michaeli. Tel: +972 3 640 5452; Fax: +972 3 640 5794; Email: yaelmi@post.tau.ac.il
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

flow-cells were required for one human genome assembly at 30× coverage. Routine analysis of predetermined genomic loci for research and medical testing cannot justify such expensive and time consuming experiments. Hence, targeted enrichment is essential for harnessing the variant detection capabilities of long-read sequencing technologies.

CATCH (Cas9-assisted targeting of chromosome segments) was first utilized for targeted cloning of large intact genomic fragments (9). CATCH is based on *in-vitro* targeted fragmentation of high molecular weight genomic DNA by Cas9, followed by separation of the target region from the rest of the genomic DNA by pulsed field gel electrophoresis (PFGE). DNA is then isolated from the gel and further used for cloning or downstream analysis (Figure 1). We recently demonstrated the utility of CATCH for targeted nanopore sequencing and showed an enrichment factor of 21.7 for sequencing of a 200 kb target from *Escherichia coli* (bioRxiv https://doi.org/10.1101/110163). Bennett-Baker and co-workers have used CATCH, followed by NGS, and showed 39- and 174-fold enrichment for 2.3 Mb and 610 kb fragments from mouse *Srsx* locus, respectively (10). Here, we harness CATCH for nanopore and NGS sequencing of human BRCA1, a large clinically relevant target, in a facile and cost effective manner. Despite inferior SNP calling compared to NGS, nanopore sequencing resulted in up to 237-fold enrichment of the BRCA1 gene, potentially enabling characterization of mutations and large structural variations on a single MinION flow-cell.

MATERIALS AND METHODS

Human subjects

The healthy donor sample used in this study was collected with informed consent for research use and approved by Institutional Review Boards in accordance with the declaration of Helsinki.

High-molecular-weight DNA extraction

For bacterial DNA extraction, *E. coli* K-12 MG1655 cells were grown in LB, and log phase cells were embedded in low melting agarose gel plugs at a concentration of ∼2 × 10^9 cells/ml as previously described (bioRxiv https://doi.org/10.1101/110163). For human high molecular weight DNA extraction, peripheral blood mononuclear cells (PBMCs) were isolated from peripheral blood of a healthy donor by density gradient centrifugation using Ficoll Paque Plus (GE Healthcare) according to manufacturer's instructions. For preparation of agarose plugs, 1 × 10^6 human cells were washed twice with PBS, resuspended in cell suspension buffer (CHEF mammalian DNA extraction kit, Bio-Rad) and incubated at 43°C for 10 min. Two percent low melting agarose (CleanCut agarose, Bio-Rad) was melted at 70°C followed by incubation at 43°C for 10 min. Melted agarose was added to the resuspended cells at a final concentration of 0.7% and mixed gently. The mixture was immediately cast into a plug mold and plugs were incubated at 4°C until solidified. Plugs were incubated twice (2 h incubation followed by an overnight incubation) at 50°C with freshly prepared 200 μl Proteinase K (20 mg/ml stock concentration, Sigma) in 2.5 ml lysis buffer (100 mM EDTA, pH 8.0, 0.2% (w/v) sodium deoxycholate, 1% (w/v) sodium lauryl sarcosine) with occasional shaking. Plugs were washed three times by adding 10 ml TE buffer (10 mM Tris, pH 8, 1 mM EDTA), manually shaking for 10 s and discarding the wash buffer before adding the next wash. Next, plugs were incubated with 50 μl RNaseA (Qiagen) in 2.5 ml TE buffer for 1 h at 37°C with occasional shaking. Plugs were then washed four times by adding 10 ml wash buffer (10 mM Tris, pH 8, 50 mM EDTA), and shaking for 15 min on a horizontal platform mixer at 180 rpm at room temperature. Following washes, plugs were stored at 4°C in wash buffer or used for Cas9 digestion.

Cas9 digestion

Preparation of gRNA for digesting E. coli genomic DNA was performed as described (bioRxiv https://doi.org/10.1101/110163). TracrRNA and crRNA for targeting human fragments were purchased from IDT. The following target sequences were used for isolation of BRCA1 locus: 5′: GCCATGACAACAACCCAGAC 3′: GCTTATTACATT CTCGGCCA. Prior to plug digestion, gRNAs were mixed according to the following conditions: 2 μl tracrRNA (200 μM stock), 2–4 μl of 5′ crRNA (100 μM stock), 2–4 μl of 3′ crRNA (100 μM stock), 12 μl duplexing buffer (IDT). For hybridization of tracrRNA and crRNA, the mixture was incubated for 5 min at 95°C and then for 10 min at room temperature. gRNAs and Cas9 were pre-assembled prior to the digestion by incubation of 0.33 μl Cas9 enzyme (20 μM stock, New England Biolabs) with the 4–6.6 μl hybridized gRNA mix, 4 μl Cas9 buffer (10× stock, New England Biolabs) and DDW to a final volume of 40 μl. Pre-assembly mix was incubated for 30 min at room temperature. Plugs were washed 4 times with 10 ml of 10 mM Tris–HCl, pH 8 followed by a single wash with 500 μl Cas9 reaction buffer (1× stock, New England Biolabs). Following pre-assembly, plugs were cut into three equal slices of ∼30 μl (∼2 μg genomic DNA) and each slice was digested with 40 μl pre-assembled Cas9-gRNA mix at 37°C for 2 h. Finally, 3 μl Proteinase K (Sigma, 20 mg/ml stock concentration) were added to each tube, and samples were incubated at 43°C for 3 h in order to remove excess Cas9 bound to the DNA.

Separation by PFGE

A 300 ml solution of 0.9% sea plaque low melting agarose gel (Lonza) was prepared in 0.3× TBE. One to six treated gel plugs (up to ∼35 μg DNA content in six plugs) were loaded onto the gel and run on a Rotaphor electrophoresis device (Biometra) for 24 h in 0.22× TBE buffer. The bands corresponding to ∼200 kb were cut out of the gel (Figure 1) and each band slice was placed in a separate tube. For recovery of DNA from the gel, agarose was washed 3 times in 10 ml TE buffer (10 mM Tris, 1 mM EDTA, pH 8) and melted at 70°C for 5 min followed by incubation at 43°C for 10 min. Next, 2 μl agarase (0.5 U/μl, Thermo Scientific) was added to each tube for digestion of the agarose and incubated for 1 h at 43°C. The DNA was purified from agarose by isopropanol precipitation. Briefly, 22 μl of 3 M NaAc and 220 μl of isopropanol were added to each tube and mixed by inverting the tubes several times. The tubes

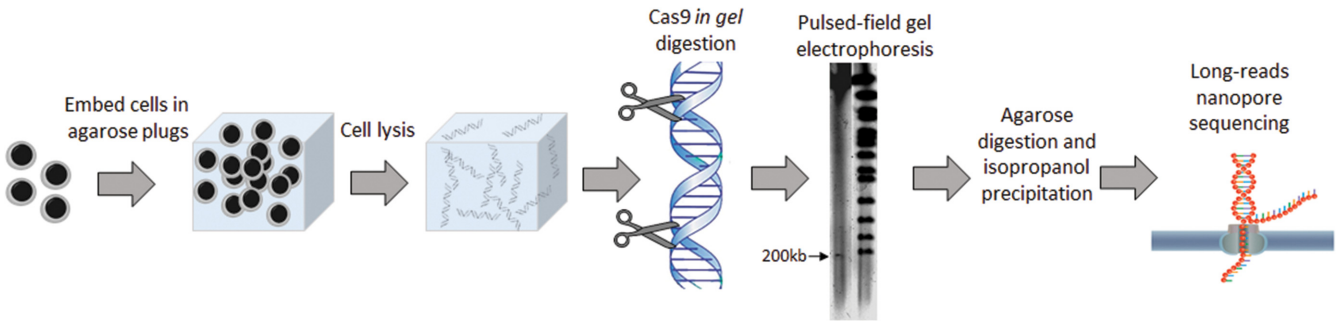

Figure 1. Schematic representation of the CATCH method. Peripheral blood mononuclear cells were embedded in an agarose gel-plug and lysed. Genomic DNA was cleaved in the plug using guided Cas9, and the target DNA was separated by PFGE. The desired band (indicated by an arrow) was excised from the gel, and the DNA was isolated, purified and analyzed.

were incubated at –20°C for at least 1 h and centrifuged for 30 min at 14 000 g at 4°C. The supernatant was carefully discarded, and the pellet was washed with 500 μl of 75% ethanol followed by 5 min of centrifugation at 14 000 g at 4°C. The supernatant was discarded. The pellet was briefly allowed to dry and resuspended in 20 μl RNase-free water.

Nanopore sequencing

For construction of bacterial DNA sequencing library, a low input expansion pack kit in combination with SQK-MAP007, R9 version kit (Oxford Nanopore Technologies) was used according to manufacturer's instructions. Library was constructed from 125 ng of bacterial DNA that was recovered from the gel without amplification, loaded on R9.4 flow-cells (Oxford Nanopore Technologies), and sequenced for 48 h. For sequencing of human DNA, 1.1–1.8 ng target DNA recovered from the gel was amplified using the multiple displacement amplification (MDA) REPLI-g midi kit (Qiagen) for long-range amplification. DNA was then purified with AMPure XP beads and 20 μg of DNA were recovered. Next, 1 μg of DNA was subjected to T7 endonuclease I fragmentation for linearization of branched amplicons as follows: DNA was incubated at 37°C for 15 min with 2 μl buffer 2 (New England Biolabs), 1 μl T7 endonuclease I (New England Biolabs) in a total volume of 20 μl, followed by an additional purification with AMPure XP beads. This process yielded 20–40 kb amplicons for nanopore sequencing. Next, we either performed a size selection step or continued directly to construction of a sequencing library. Size selection was performed by running unbranched DNA in a low melting agarose gel, extracting DNA from the upper band corresponding to 20–40 kb by agarase gel digestion (0.5 U/μl, Thermo Scientific) followed with AMPure XP beads purification. The $1D^2$ sequencing kit SQK-LSK308 (Oxford Nanopore Technologies) was used for construction of sequencing libraries with 500 ng input DNA, excluding DNA fragmentation and repair steps. R9.5 flow-cells were run for 24–48 h on a MinION sequencing device (Oxford Nanopore Technologies), and reads were base-called using the MinKNOW software. Reads which were shorter than 500 bp or had an average quality score lower than 10 were excluded from downstream analysis using Nanofilt (https://github.com/wdecoster/nanofilt). Alignment of reads to the human hg38 reference genome was performed using BWA-MEM (version 0.7.15, arXiv:1303.3997v2), with -x ont2d parameters, and coverage was calculated using the Galaxy wrapper for BEDTools genomecov (version 2.26.0) (11,12). The resulting bedGraph files were loaded into Circos (version 0.69-6) (13), the Integrative Genomics Viewer (IGV) (14), or the UCSC genome browser for visual evaluation of read coverage. *De novo* assembly was performed using Canu (version 1.4) (15) with default parameters, specifying a genome size of 200 kb.

Visualization of mutations and structural variations was performed using IGV. Variants were called with FreeBayes (version v1.0.2–29-g41c1313) (arXiv:1207.3907) using Simple diploid calling settings, and filtered with the VCFfilter tool in vcflib (https://github.com/vcflib/vcflib) using the following parameters: DP > 15 & QUAL > 1 & QUAL / AO > 10 & SAF > 0 & SAR > 0 & RPR > 1 & RPL > 1. For validation, specific locations were amplified (Supplementary Table S1) and PCR products were analyzed by Sanger sequencing.

For enrichment factor calculation we used the following formula: EF = (total coverage of the target region / size of the target region)/(total coverage of the off-target region/size of the off target region) = mean coverage of the target region / mean coverage of the off-target region.

Next generation sequencing

Sequencing Library was prepared using Nextera DNA Library Prep Kit (Illumina) with 100 ng of unbranched purified DNA as template according to manufacturer's protocol. Single-end sequencing of 1 nM library was performed on an Illumina Mini-Seq instrument using MiniSeq High Output Reagent Kit (75 cycles). Sequencing reads were trimmed with TrimGalore! (version 0.4.3.1) (https://github.com/FelixKrueger/TrimGalore) using default parameters to trim Illumina adapter sequences, and then aligned to the hg38 human reference using bowtie2 (version 2.3.4) (16). Following alignment, reads with mapping quality (MAPQ) less than 30 were filtered with SAMtools (version 1.2) (17). Finally, genomic coverage of reads was calculated using BEDTools (version 2.25.0) genomecov. Variants were called with FreeBayes and filtered as described above for nanopore sequencing.

RESULTS

Targeted enrichment of a large *E. coli* genomic region

For method development and as a proof of concept, we used CATCH for targeted enrichment of a ∼200 kb fragment from the *E. coli* k 12 MG1655 genome. Due to the small size of bacterial cells, over 10^8 cells may be treated in a single reaction, resulting in up to 10^8 cleaved copies of the target region. This in turn results in a clear and distinct band during PFGE isolation, ideal for optimization purposes. The bacterial genome was cleaved at two sites flanking the target according to the complementary gRNAs, and the fragment was isolated by PFGE. The large amounts of target DNA extracted from bacterial cells allowed sequencing library preparation from native DNA without pre-amplification. Nanopore sequencing was used to analyze the isolated fragments and measure the level of enrichment. Figure 2 shows pronounced elevated coverage across the target region relative to the rest of the genome, indicating a significant enrichment of the target. Several repeated experiments yielded a median read length of 7424 bp and showed a consistent ∼21-fold enrichment. In addition, we used optical mapping to verify that the enriched fragments were intact. The combination of long-read sequencing and optical mapping may allow targeted characterization of large SVs and genetic rearrangements as well as long-range epigenetic profiling ((18,19), bioRxiv https://doi.org/10.1101/110163, bioRxiv https://doi.org/10.1101/113522). DNA molecules extracted from the pulsed-field gel were fluorescently labeled at specific sequence motifs, loaded onto a nanochannel array chip and imaged. The resulting molecule images contain a fluorescent barcode along the molecules that unambiguously identifies them as originating from the target region. As shown in Figure 2, optical mapping showed significant coverage enrichment in the target region compared to the rest of the genome. Furthermore, 35% of the detected molecules were longer than 75 kb, indicating we were able to isolate the target region and maintain high molecular weight DNA, which can be further used for studying genomic aberrations at various length scales or for targeted BAC/fosmid library construction.

Isolation and characterization of the human BRCA1 gene locus

Next, we used CATCH to isolate and analyze the human breast and ovarian cancer-related BRCA1 gene. First, two gRNAs were designed to target ∼200 kb region that contains the ∼80 kb BRCA1 gene, along with its regulatory and flanking regions. High molecular weight DNA was isolated from human PBMCs by lysing the cells within gel plugs, thus protecting the genomes from shearing. DNA was cleaved by soaking the gel plug with Cas9 and BRCA1-specific gRNAs, followed by PFGE isolation and isopropanol purification. Up to 4.6 ng of target DNA were recovered from each plug. Prior to construction of nanopore sequencing library, DNA was amplified using a multiple displacement amplification (MDA) protocol for long-range amplification, purified, and subjected to endonuclease fragmentation for linearization of branched amplicons. DNA was sequenced on the MinION portable sequencer (Oxford Nanopore Technologies) and a total of 964 050 ± 180 263 reads yielding 2.37 ± 0.32 Gb were acquired in a single flow-cell experiment within 48 h. 50% of the reads were obtained within the first 10 h (Supplementary Figure S1), sufficient data for discovery of most SNPs identified with the full dataset. Data was filtered according to length and quality score of the reads (see methods), maintaining 0.64 ± 0.09 Gb with a mean read length of 2981 ± 252 for further analysis. Over 98% of the reads were aligned to the human genome with 35× ±10 mean sequencing depth in the target region. About 1% of the reads were aligned to the target, yielding an enrichment factor of 194 ± 14. When adding a size selection step prior to library preparation, 99.95% of the reads aligned to the human genome, with ∼70× mean sequencing depth in the target region, mean read length of ∼4922, and enrichment factor of 237. NGS analysis of the same sample resulted in ∼90x mean coverage of the target region and yielded an enrichment factor of 192 (Figure 3, Supplementary Figures S2–S5, Supplementary Table S2). We note that all sequencing runs showed a similar, highly variable coverage distribution, most likely due to the target amplification step. Despite the use of the presumably unbiased multiple displacement amplification (MDA), the coverage across the target region ranges from 0 to 176, where 0.5–1.5-fold of the mean coverage was obtained in 94% of the target region. This is in contrast to the experiments performed on unamplified *E. coli* target DNA that showed relatively uniform coverage across the target region (Figure 2).

We next performed *de-novo* assembly using canu, which yielded three contigs. Two of the contigs were generated from less than 10 reads and did not map to the target region. The third contig was based on 734 reads, with a length of 179 539 bp and 98% identity to the target region (Supplementary data S1). Aligned nanopore reads were used for analysis of BRCA1 mutations and structural variations. In addition, NGS analysis was performed on the same sample for verification of SNP results (Figure 4 and Supplementary Table S3). 177 SNPs were detected in the NGS data, 11 of them homozygous and the rest heterozygous. Of these 177 SNPs, 55 (31%) were detected in the nanopore sequencing data, using the same analysis pipeline. Another 55 of them could be identified by going over the alignment results manually (31%). Two SNPs were not present in the nanopore data (1%), and the rest (37%) were inconclusive, mostly due to low coverage and the presence of homopolymers. A list of all SNPs detected is presented in Supplementary Table S3.

40 SNPs that were identified in the nanopore data were not detected in the NGS data. 24 (60%) of these SNPs were not supported by the NGS reads, and the rest either had insufficient coverage (4 SNPs, 10%), were inconclusive (7 SNPs, 17.5%), or could be identified manually (5 SNPs, 12.5%). Small deletions and insertions with a length of 1–2 bp were disregarded from nanopore data, as they appeared in homopolymer regions and were possibly the result of basecalling errors.

As shown in Figure 4, in heterozygous SNPs, the two variants were represented almost evenly in the aligned reads (red/green), while for homozygous SNPs the majority of reads are converted (red/blue). In addition, we could identify a deletion of 3 out of 22 CA dinucleotide repeats that

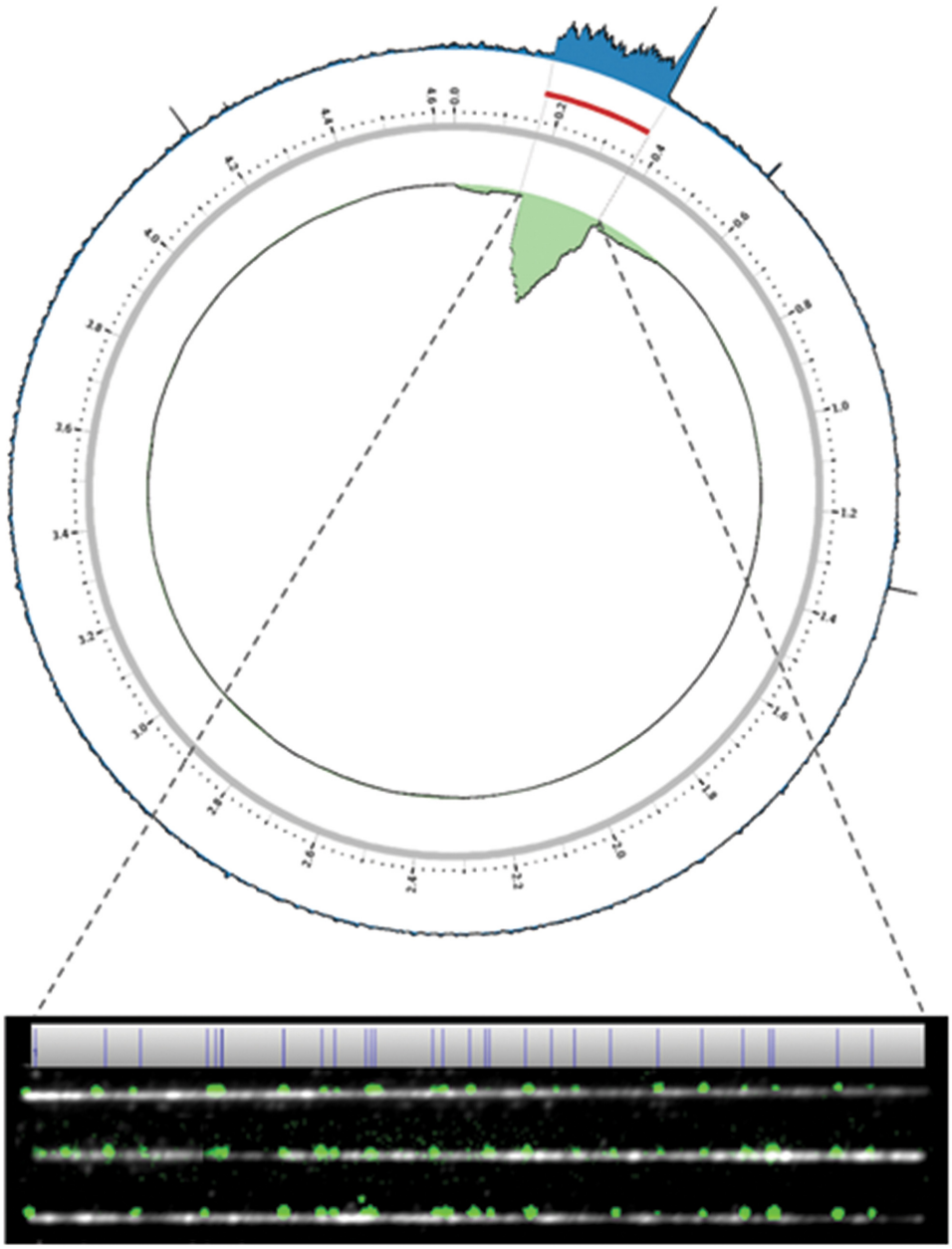

Figure 2. Circos plot showing coverage analysis of the enriched 200 kb target region from the *E. coli* genome analyzed by nanopore sequencing (external circle) and optical mapping (internal circle). The red line indicates the target region. Bottom panel shows intact molecules analyzed by optical mapping. Blue vertical lines represent the expected label positions.

NGS data could not reveal. All the variants shown in Figure 4 were also verified by amplification of the specific location followed by Sanger resequencing.

DISCUSSION

Long-read nanopore sequencing may potentially characterize a comprehensive spectrum of structural variations that are challenging for NGS (20,21). Nevertheless, the high cost of nanopore sequencing prevents such variation studies on a genome scale, calling for targeted analysis of particular targets of interest. Here, we demonstrate targeted enrichment of the human BRCA1 gene, followed by long-read sequencing. CATCH allowed us to use the portable MinION nanopore sequencer for analysis of genetic aberrations in the entire gene region including regulatory elements and non-coding regions.

BRCA1 is an 80 kb gene, notorious for its wide spectrum of SVs at all length scales. These include SNPs, deletions and duplications ranging up to several kb, complex rearrangements and multiple tandem repeat elements. Although this work was performed on BRCA1 from a healthy donor, the 4.9 kb mean read length achieved in this experiment demonstrates the potential of detecting such aberrations. In addition, 2–10 bp repeats, also called microsatellites or short tandem repeats (STR), were shown to influence many

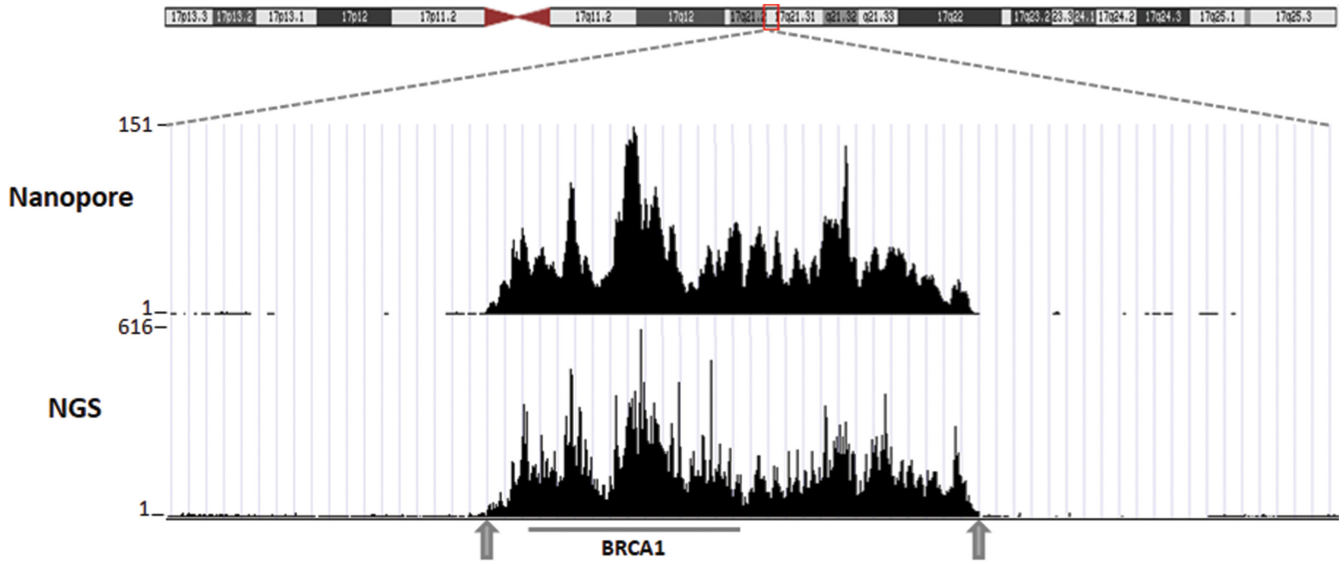

Figure 3. Nanopore and NGS sequencing coverage analysis of the ∼200 kb target region isolated from the human BRCA1 gene locus.

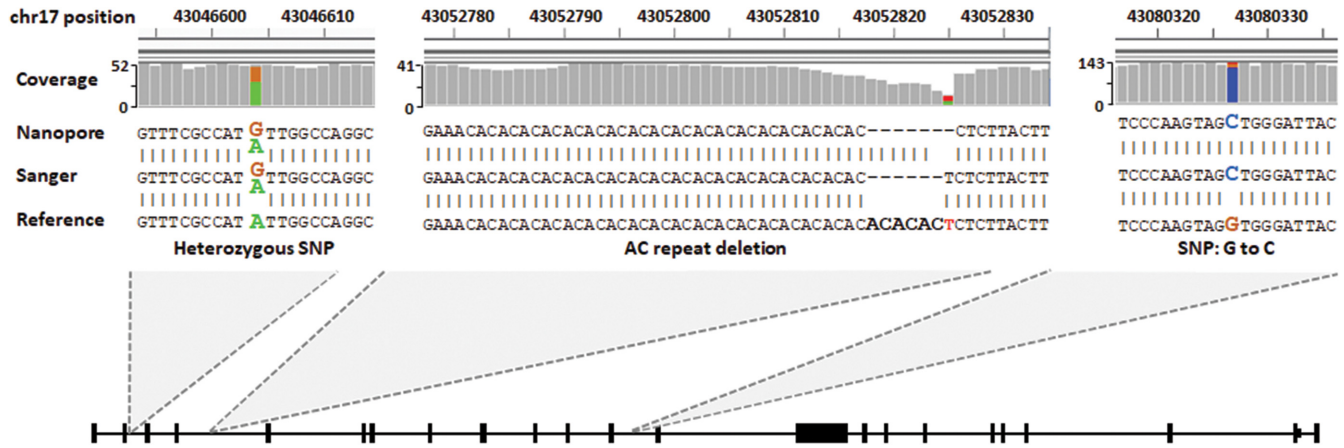

Figure 4. Detection of homozygous SNP, heterozygous SNP and deletion of repetitive sequence in the enriched BRCA1 gene. Coverage was visualized by IGV. Alterations between the *de-novo* consensus sequence and the reference are shown as a colored bar. The height of each color represents the number of reads that contained the indicated base shown below.

processes in disease-related loci, including chromatin state, RNA transcription, splicing and translation as well as post-translational modifications (22). NGS sequencing errors increase exponentially with STR length and saturate around 15–20 bp in di-nucleotide repeat arrays (23). Our nanopore data revealed a deletion of three di-nucleotide blocks in a 44 bp repeat array. The array was not fully recovered in the corresponding NGS data, validating that it is challenging to accurately characterize STRs by NGS. In terms of SNP detection, NGS discovered 177 SNPs. Nanopore data revealed 62% of the SNPs with only two SNPs not present at all in the nanopore data and the remaining 37% inconclusive. We note that the same SNP detection pipeline was used for both data types. Improvements in dedicated SNP callers for nanopore data as well as increased coverage will most likely enhance the ability to extract reliable SNPs.

Due to the low throughput of the MinION, target enrichment is essential for utilizing its long reads for variant analysis. Current enrichment methods for this gene mostly include direct amplification or hybridization capture of select short gene fragments of the coding regions (Supplementary Table S4). Nevertheless, it was shown that mutations in non-coding regions of this gene can induce significant reduction of transcription (24), or exon skipping resulting in truncated BRCA1 protein (25). Furthermore, the CATCH scheme reported here is immune to genetic complexities in the target region. The targeting Cas9 guide-RNAs are designed to cut at non-variable sites flanking BRCA1, and the entire gene is isolated, including intronic and regulatory regions. This approach avoids the complications associated with primer design needed for amplification or hybridization capture. These require knowledge of the target sequence and thus can be hindered by the complex nature of BRCA1. CATCH offers isolation of native high molecular weight DNA that retains its epigenetic state. Future developments will allow targeted epigenetic profiling by direct sequencing of epige-

netic modifications and multiplexed optical mapping ((26–29), bioRxiv https://doi.org/10.1101/113522). We envision using CATCH for isolation of selected disease-related gene sets, offering comprehensive multi-scale genetic and epigenetic analysis. This approach can be integrated into diagnostic pipelines to complement current methods, and in parallel, may shed light on the mechanisms of disease onset and progression.

DATA AVAILABILITY

Data for CATCH-nanopore and CATCH-NGS have been deposited at the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under accession number SRP133472. Coverage data of chromosome 17 have been uploaded to the UCSC public sessions under the name 'CATCH BRCA1' (https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=hilasha&hgS_otherUserSessionName=CATCH%20BRCA1)

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Chris Boles (Sage Science) for sharing the design of BRCA1 gRNAs. We thank Prof. Oded Rechavi for providing NGS support.

FUNDING

I-Core program of the Israel Science Foundation [1902/12]; European Research Council Starter Grant [337830]. Funding for open access charge: I-Core program of the Israel Science Foundation [1902/12]; European Research Council Starter Grant [337830].
Conflict of interest statement. None declared.

REFERENCES

1. King,M.-C., Marks,J.H. and Mandell,J.B. (2003) Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*, **302**, 643–646.
2. Couch,F.J., Nathanson,K.L. and Offit,K. (2014) Two decades after BRCA: setting paradigms in personalized cancer care and prevention. *Science*, **343**, 1466–1470.
3. Welcsh,P.L. and King,M.C. (2001) BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum. Mol. Genet.*, **10**, 705–713.
4. Day,K., Song,J. and Absher,D. (2014) Targeted sequencing of large genomic regions with CATCH-Seq. *PLoS One*, **9**, e111756.
5. Cheeseman,K., Rouleau,E., Vannier,A., Thomas,A., Briaux,A., Lefol,C., Walrafen,P., Bensimon,A., Lidereau,R., Conseiller,E. *et al.* (2012) A diagnostic genetic test for the physical mapping of germline rearrangements in the susceptibility breast cancer genes BRCA1 and BRCA2. *Hum. Mutat.*, **33**, 998–1009.
6. Sluiter,M.D. and van Rensburg,E.J. (2011) Large genomic rearrangements of the BRCA1 and BRCA2 genes: review of the literature and report of a novel BRCA1 mutation. *Breast Cancer Res. Treat.*, **125**, 325–349.
7. Huddleston,J., Chaisson,M.J.P., Steinberg,K.M., Warren,W., Hoekzema,K., Gordon,D., Graves-Lindsay,T.A., Munson,K.M., Kronenberg,Z.N., Vives,L. *et al.* (2017) Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.*, **27**, 677–685.
8. Jain,M., Koren,S., Miga,K.H., Quick,J., Rand,A.C., Sasani,T.A., Tyson,J.R., Beggs,A.D., Dilthey,A.T., Fiddes,I.T. *et al.* (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.*, **36**, 338–345.
9. Jiang,W., Zhao,X., Gabrieli,T., Lou,C., Ebenstein,Y. and Zhu,T.F. (2015) Cas9-Assisted targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.*, **6**, 8101.
10. Bennett-Baker,P.E. and Mueller,J.L. (2017) CRISPR-mediated isolation of specific megabase segments of genomic DNA. *Nucleic Acids Res.*, **45**, e165–e165.
11. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
12. Goecks,J., Nekrutenko,A., Taylor,J. and Galaxy Team,T. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
13. Krzywinski,M., Schein,J., Birol,I., Connors,J., Gascoyne,R., Horsman,D., Jones,S.J. and Marra,M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.
14. Thorvaldsdottir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.
15. Koren,S., Walenz,B.P., Berlin,K., Miller,J.R., Bergman,N.H. and Phillippy,A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*, **27**, 722–736.
16. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
17. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
18. Gilboa,T., Torfstein,C., Juhasz,M., Grunwald,A., Ebenstein,Y., Weinhold,E. and Meller,A. (2016) Single-Molecule DNA methylation quantification using Electro-optical sensing in Solid-State nanopores. *ACS Nano*, **10**, 8861–8870.
19. Levy-Sakin,M. and Ebenstein,Y. (2013) Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy. *Curr. Opin. Biotechnol.*, **24**, 690–698.
20. Sudmant,P.H., Rausch,T., Gardner,E.J., Handsaker,R.E., Abyzov,A., Huddleston,J., Zhang,Y., Ye,K., Jun,G., Hsi-Yang Fritz,M. *et al.* (2015) An integrated map of structural variation in 2,504 human genomes. *Nature*, **526**, 75–81.
21. Cretu Stancu,M., van Roosmalen,M.J., Renkens,I., Nieboer,M.M., Middelkamp,S., de Ligt,J., Pregno,G., Giachino,D., Mandrile,G., Espejo Valle-Inclan,J. *et al.* (2017) Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.*, **8**, 1326.
22. La Spada,A.R. and Taylor,J.P. (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.*, **11**, 247–258.
23. Fungtammasan,A., Ananda,G., Hile,S.E., Su,M.S.-W., Sun,C., Harris,R., Medvedev,P., Eckert,K. and Makova,K.D. (2015) Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.*, **25**, 736–749.
24. dos Santos,E.S., Caputo,S.M., Castera,L., Gendrot,M., Briaux,A., Breault,M., Krieger,S., Rogan,P.K., Mucaki,E.J., Burke,L.J. *et al.* (2017) Assessment of the functional impact of germline BRCA1/2 variants located in non-coding regions in families with breast and/or ovarian cancer predisposition. *Breast Cancer Res. Treat.*, **168**, 311–325.
25. Yoon,K.-A., Kong,S.-Y., Lee,E.J., Cho,J.N., Chang,S. and Lee,E.S. (2017) A Novel germline mutation inBRCA1Causes exon 20 skipping in a korean family with a history of breast cancer. *J. Breast Cancer*, **20**, 310–313.
26. Michaeli,Y., Shahal,T., Torchinsky,D., Grunwald,A., Hoch,R. and Ebenstein,Y. (2013) Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases. *Chem. Commun. (Camb).*, **49**, 8599–8601.
27. Rand,A.C., Jain,M., Eizenga,J.M., Musselman-Brown,A., Olsen,H.E., Akeson,M. and Paten,B. (2017) Mapping DNA

methylation with high-throughput nanopore sequencing. *Nat. Methods*, **14**, 411–413.
28. Simpson,J.T., Workman,R.E., Zuzarte,P.C., David,M., Dursi,L.J. and Timp,W. (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods*, **14**, 407–410.
29. Zirkin,S., Fishman,S., Sharim,H., Michaeli,Y., Don,J. and Ebenstein,Y. (2014) Lighting up individual DNA damage sites by in vitro repair synthesis. *J. Am. Chem. Soc.*, **136**, 7771–7776.