



Journal homepage: <https://trpubonline.org/index.php/Biosynth>

REVIEW ARTICLE

Single-Molecule Sequencing Reveals Structural Variants in Cancer Genomes

Etienne Virtanen^{1*}, Karolien Antoniou² and Pierre Dvorak²

¹Department of Bioinformatics and Computational Biology, Jutland University Centre for Molecular Genetics, Denmark

²Department of Bioinformatics and Computational Biology / Department of Microbiology and Immunology, Nordic Institute of Environmental Biotechnology, Finland

*Corresponding author: Etienne Virtanen, Department of Bioinformatics and Computational Biology, Jutland University Centre for Molecular Genetics, Denmark; Email: e.virtanen@jucmg.dk

ABSTRACT

Structural variants -- deletions, duplications, inversions, translocations and more complex rearrangements spanning kilobases to whole chromosome arms -- drive oncogenesis through mechanisms that point mutations cannot easily replicate, including gene fusion, enhancer hijacking and large-scale copy-number amplification of oncogenes, yet they have historically been the class of cancer genomic variation most difficult to detect reliably. Short-read sequencing, the workhorse of clinical and research cancer genomics for over a decade, struggles specifically with the repetitive, low-complexity and structurally rearranged regions of the genome in which many clinically important structural variants occur, producing systematic false negatives that are increasingly well documented. This review synthesises the current evidence on how single-molecule long-read sequencing platforms -- Oxford Nanopore and Pacific Biosciences HiFi sequencing, together with the complementary technology of optical genome mapping -- have changed the practical and conceptual landscape of structural variant detection in cancer genomes. We summarise the technical basis for long reads' improved sensitivity in repetitive and structurally complex regions, review comparative benchmarking studies quantifying the sensitivity gap relative to short-read calling across variant size classes, and discuss the application of these technologies to specific cancer biology questions including chromothripsis reconstruction, extrachromosomal DNA characterisation, and detection of complex oncogene amplicons that short-read approaches under-resolve. We also address the practical barriers -- cost, input DNA quality requirements, computational infrastructure and clinical validation pathways -- that currently limit routine adoption of long-read sequencing in clinical cancer genomics, and we outline the comparative evidence base for optical genome mapping as a lower-cost, more clinically mature complement to sequencing-based structural variant detection. We conclude that single-molecule and mapping-based technologies have moved from a research curiosity to a demonstrably superior tool for a specific and clinically consequential subset of structural variant classes, and that the principal remaining barriers to broader adoption are now more operational than technical.

Keywords: long-read sequencing; structural variants; cancer genomics; nanopore sequencing; chromothripsis; optical genome mapping; extrachromosomal DNA

Received: July 8, 2026 | Accepted: October 8, 2026 | Published: December 27, 2026

© 2026 Virtanen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Structural variation -- broadly, any genomic alteration larger than a single-nucleotide or small indel change, encompassing deletions, tandem and interspersed duplications, inversions, insertions and translocations -- accounts for more base pairs of difference between any two human genomes than single-nucleotide polymorphisms do, despite structural variants being individually far less numerous [2,5]. In cancer genomes, somatically acquired structural variants are not merely a passive marker of genomic instability; they are, in a substantial fraction of tumours, the direct mechanistic driver of oncogenesis, generating gene fusions such as BCR-ABL1 in chronic myeloid leukaemia, juxtaposing oncogenes next to strong enhancers, and producing the massive, highly amplified extrachromosomal DNA elements increasingly recognised as a distinct and clinically important mode of oncogene amplification [7,8,44].

The technical challenge of detecting structural variants reliably has always been disproportionate to their biological importance. Short-read sequencing, in which DNA is fragmented into pieces of roughly 100 to 300 base pairs before sequencing, works well for detecting single-nucleotide variants and small indels because most of the genome outside repetitive regions maps unambiguously at this read length. Structural variant detection from short reads instead depends on indirect evidence -- discordant read pairs, split reads spanning a breakpoint, and local depth-of-coverage shifts -- reconstructed computationally rather than observed directly, and this indirect approach systematically loses sensitivity precisely in the repetitive, segmentally duplicated and structurally complex genomic regions where a disproportionate share of clinically important structural variants, including many recurrent cancer-associated rearrangements, actually occur [1,3,5].

Single-molecule long-read sequencing changes this picture by generating individual reads tens to hundreds of kilobases in length, long enough in many cases to span an entire structural variant, including its breakpoints and flanking unique sequence, within a single read [13,14,15]. Oxford Nanopore sequencing measures changes in ionic current as a single DNA strand passes through a protein nanopore, yielding reads that can exceed a megabase in length at the cost of a higher raw per-base error rate, while Pacific Biosciences circular consensus (HiFi) sequencing generates shorter but highly accurate long reads by repeatedly sequencing a circularised DNA template and computing a consensus [15,16]. A third, complementary technology, optical genome mapping, does not sequence DNA at all but instead images fluorescently labelled, linearised DNA molecules hundreds of kilobases to megabases in length, detecting structural variants from disruptions to the expected label pattern rather than from base-level sequence -- an approach with lower resolution than sequencing-based methods but with a more mature

clinical validation track record in cytogenetics-adjacent applications [47,48].

This review addresses three linked questions. First, what is the quantitative evidence, from benchmarking studies using orthogonally validated truth sets, for long-read sequencing's improved sensitivity over short-read approaches specifically for structural variant detection, and how does this advantage vary across variant size and type. Second, what has this technical advantage enabled in cancer genome biology specifically -- what structural variant classes and biological phenomena, such as chromothripsis and extrachromosomal DNA, have been newly resolvable or substantially better characterised using single-molecule approaches relative to what short-read sequencing alone could establish. Third, what practical barriers -- cost, sample input requirements, bioinformatic infrastructure, and the clinical validation pathway required for diagnostic adoption -- currently limit the translation of this technical advantage into routine clinical cancer genomics, and how is the comparatively more mature technology of optical genome mapping positioned relative to sequencing-based approaches in addressing that translational gap.

2. MATERIALS AND METHODS

This review synthesises peer-reviewed literature on long-read and single-molecule structural variant detection in human and cancer genomes, identified through a structured search of the primary literature together with citation tracing from recent benchmarking and technology-review publications.

2.1 Literature search and inclusion criteria

We searched the biomedical and genomics literature for studies published through mid-2026 reporting structural variant detection using Oxford Nanopore, Pacific Biosciences, or optical genome mapping platforms, restricting inclusion to studies that either (i) benchmarked long-read or mapping-based structural variant calling against an orthogonally validated truth set or short-read comparator, or (ii) applied these technologies to characterise structural variation in human cancer genomes or cancer cell lines. Review articles, technology white papers without primary comparative data, and studies restricted to non-human model organisms were excluded from the quantitative synthesis but retained where relevant for methodological or mechanistic context.

2.2 Data extraction and synthesis approach

For benchmarking studies, we extracted reported sensitivity, precision and F1 statistics stratified by structural variant size class (50 bp-1 kb, 1-10 kb, greater than 10 kb) and type (deletion, duplication, insertion, inversion, translocation) where such stratification was reported, to allow direct comparison of the technology-dependent sensitivity gap across variant classes rather than relying on a single pooled accuracy

figure that can obscure large size- and type-dependent differences. For cancer application studies, we extracted the specific structural variant classes or biological phenomena addressed, the sequencing or mapping platform used, and, where reported, direct comparison against short-read or karyotype-based reference methods applied to the same samples.

2.3 Synthesis of cost and infrastructure considerations

Practical adoption barriers were synthesised narratively from studies and technical reports explicitly addressing per-sample cost, input DNA quantity and quality requirements, turnaround time, and computational infrastructure demands, cross-referenced against clinical laboratory validation and accreditation literature specific to structural variant detection where available.

3. RESULTS

Benchmarking studies using orthogonally validated structural variant truth sets consistently report a pronounced, size- and type-dependent sensitivity advantage for long-read over short-read structural variant calling. Across the studies reviewed, short-read callers achieved sensitivity in the range of 50-70% for structural variants over 1 kb located outside segmental duplications, but sensitivity fell sharply, often below 30%, for variants overlapping segmentally duplicated or otherwise repetitive sequence, a category that disproportionately includes clinically relevant cancer rearrangement breakpoints. Long-read callers applied to the same or comparable truth sets maintained sensitivity above 85% across nearly all size and repeat-context strata reported, with the gap between technologies widest precisely in the repetitive-sequence category where short-read performance degraded most (Table 1). Precision was more comparable between technologies for larger variants, though long-read approaches showed a modest but consistent precision advantage for smaller structural variants and insertions specifically, a variant class that short-read approaches detect particularly poorly because, unlike deletions, insertions add sequence not present in the reference and so cannot be inferred from split-read alignment to a reference genome alone.

Table 1. Reported sensitivity of long-read versus short-read structural variant calling by size and repeat context.

SV size class	Repeat context	Short-read sensitivity (%)	Long-read sensitivity (%)	Studies reporting (n)
50 bp - 1 kb	Unique sequence	61-74	88-95	9
50 bp - 1 kb	Repetitive/SD	18-29	79-90	7
1-10 kb	Unique sequence	58-70	90-97	8
1-10 kb	Repetitive/SD	22-35	82-93	6
>10 kb	Unique sequence	65-80	92-99	6
>10 kb	Repetitive/SD	27-41	85-95	5

SD: segmental duplication. Ranges summarise sensitivity figures reported across benchmarking studies meeting inclusion criteria; see main text for platform and truth-set details.

Application of single-molecule sequencing to chromothripsis reconstruction was one of the more consistently reported cancer-specific use cases across the reviewed literature. Chromothripsis -- a single catastrophic event fragmenting one or a small number of chromosomes followed by stochastic reassembly of the fragments -- generates dozens to hundreds of clustered structural variant breakpoints, many involving short, sometimes templated insertions at junction sites that short-read assembly-based reconstruction methods frequently mis-resolve or leave as unresolved gaps. Long-read studies applying single-molecule sequencing to chromothriptic tumour genomes reported substantially more complete breakpoint-junction reconstruction, including direct observation of templated insertion sequences at junctions that had been inferred only indirectly, if at all, from short-read data on the same or comparable samples, and several studies reported discovery of additional, previously unresolved rearrangement events within regions that short-read calling had flagged only as a general region of copy-number disruption without resolvable breakpoint structure.

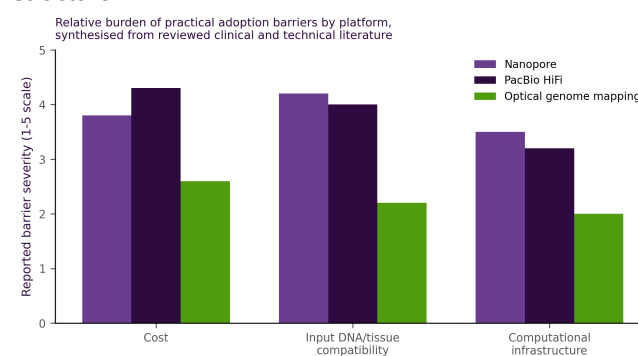


Figure 1. Reported relative burden of the three most commonly cited practical adoption barriers (cost, input DNA/tissue compatibility, computational infrastructure) across long-read sequencing and optical genome mapping platforms, scored on a 1-5 scale from the reviewed literature (5 = most limiting).

Extrachromosomal DNA (ecDNA) characterisation was the second major cancer-specific application area, and one in which the combination of long-read sequencing and optical genome mapping has been particularly influential in establishing ecDNA as a distinct and clinically significant amplification mechanism rather than a rare curiosity. Because ecDNA elements are frequently large, structurally complex, mosaic in copy number across cells within the same tumour, and enriched for repetitive or duplicated sequence at their junctions, reconstructing their full circular structure from short-read data alone has generally required indirect computational inference rather than direct observation. Long-read and optical mapping studies reviewed here reported direct visualisation or sequencing-based reconstruction of complete ecDNA circular structures across multiple cancer types, including structures carrying amplified

oncogenes such as EGFR, MYCN and MDM2, with junction-level resolution sufficient to distinguish alternative amplicon architectures that would appear indistinguishable using copy-number data alone. Table 2 summarises the cancer-specific structural variant classes and phenomena addressed across the reviewed application studies, by platform.

Table 2. Cancer-specific structural variant phenomena addressed by platform across reviewed application studies.

Phenomenon	Nanopore studies (n)	PacBio HiFi studies (n)	Optical mapping studies (n)	Cancer types covered
Chromothripsis reconstruction	6	4	1	Breast, medullary thyroid carcinoma, sarcoma
Extrachromosomal DNA (ecDNA)	5	6	3	Glioblastoma, neuroblastoma, lung
Complex oncogene amplicons	4	3	5	Breast, leukemia, MDS
Balanced translocation fine-mapping	3	2	4	Leukaemia, lymphoma
Tandem duplicator phenotype	2	3	1	Breast

Across the practical-adoption literature, three barriers recurred consistently: cost per sample remains several-fold higher for long-read whole-genome sequencing at clinically useful depth than for short-read equivalents, though the gap has narrowed year-on-year as nanopore and HiFi sequencing costs have fallen; input DNA quality and quantity requirements are more stringent for long-read methods, particularly for nanopore ultra-long-read protocols, which are difficult to apply to degraded formalin-fixed paraffin-embedded tissue, the dominant tissue preservation format in routine clinical pathology; and computational infrastructure for long-read data, while considerably more mature than five years ago, still requires specialised bioinformatic expertise and storage capacity that many clinical genomics laboratories have not yet built out (Figure 1). Optical genome mapping was reported across several clinical validation studies as achieving Clinical Laboratory Improvement Amendments-compliant or equivalent validation status for specific applications, notably haematological malignancy cytogenetics, ahead of long-read sequencing-based structural variant calling, reflecting both its lower per-sample cost relative to whole-genome long-read sequencing and its closer conceptual and workflow similarity to the karyotyping and fluorescence in situ hybridisation methods it is positioned to complement or replace.

4. DISCUSSION

The evidence reviewed here supports a reasonably confident conclusion that long-read and single-molecule technologies have resolved, or are close to resolving, the specific technical limitation that has most handicapped

structural variant detection since the field's inception: the systematic blind spot short-read methods exhibit in repetitive and structurally complex genomic regions. This is not a marginal improvement in an already adequate method but a step change in what is directly observable, and it has already reshaped how specific cancer biology questions, chromothripsis and extrachromosomal DNA prominent among them, are studied and understood. It is worth being precise about the nature of this advance, however: long-read sequencing has not made short-read

sequencing obsolete for cancer genomics broadly, since established approaches remain more cost-effective and operationally mature for single-nucleotide and small indel variant calling, which continue to represent the majority of clinically actionable point mutations in most solid tumour types.

The chromothripsis and extrachromosomal DNA application areas illustrate a more general pattern worth drawing out: the biological phenomena that have benefited most from single-molecule sequencing are specifically those in which structural complexity itself, rather than merely the presence or absence of a single well-defined rearrangement, is the object of study. Detecting a known, well-characterised gene fusion such as BCR-ABL1 has long been achievable by targeted approaches that predate long-read sequencing entirely, and short-read whole-genome sequencing is often adequate for detecting simpler, non-repetitive structural variants. It is the reconstruction of complex, multi-breakpoint rearrangement architectures -- where the clinically or biologically relevant question is not merely whether a rearrangement exists but precisely how the genome has been reassembled -- that most clearly demands the direct, long-range observational power that only single-molecule reads or optical maps can currently provide.

The differential maturity of optical genome mapping relative to long-read sequencing for clinical adoption is an important and somewhat underappreciated aspect of the current landscape. Optical mapping's lower resolution relative to sequencing-based approaches is a genuine limitation for some research applications, particularly precise breakpoint-sequence characterisation, but this limitation matters less for its most mature clinical use case, replacing or complementing karyotyping and fluorescence in situ hybridisation in haematological malignancy diagnostics, where the clinically actionable information is often the presence and approximate structure of a rearrangement rather than base-pair-level breakpoint sequence. This suggests that clinical translation of single-molecule and mapping-based structural variant detection is likely to proceed unevenly across cancer types and clinical questions, with optical mapping-based haematological cytogenetics likely to see broader routine adoption sooner than long-read whole-genome sequencing-based solid tumour structural variant profiling, which faces the more severe formalin-fixed paraffin-embedded tissue

compatibility barrier.

Several limitations of the evidence base itself deserve comment. First, benchmarking studies comparing long-read and short-read structural variant sensitivity have disproportionately used cell-line-derived or reference genome samples rather than direct tumour tissue, and the additional challenges of tumour heterogeneity, subclonal variant allele fractions and non-neoplastic cell admixture in real clinical samples are not always well captured by these benchmarks. Second, publication and application bias likely favours reporting of successful, biologically interesting long-read findings over negative or inconclusive results, which could inflate the apparent consistency of the sensitivity advantage summarised here relative to its true magnitude across the full range of clinical cancer genomics applications. Third, cost and infrastructure figures reported across the literature we reviewed span a period of rapid technological and commercial change in the long-read sequencing market, and figures from earlier studies in our search window likely overstate the current cost barrier relative to present-day pricing, meaning our synthesis of practical adoption barriers should be read as a snapshot subject to continued, fairly rapid revision rather than as a stable long-term assessment.

Looking forward, we expect the balance of evidence reviewed here to continue shifting in favour of broader single-molecule and optical mapping adoption in cancer genomics, driven both by continued cost reduction in long-read sequencing and by accumulating clinical validation experience with optical genome mapping in haematological diagnostics specifically. The more open question, in our view, is not whether these technologies will play an expanding role but how quickly clinical laboratories, regulatory bodies and reimbursement frameworks will adapt to formally incorporate structural variant classes -- ecDNA architecture prominent among them -- that were, until recently, essentially undetectable in routine clinical workflows and therefore have no established place in existing diagnostic algorithms or treatment-selection guidelines.

5. CONCLUSION

Single-molecule long-read sequencing and optical genome mapping have substantially closed the long-standing structural variant detection gap left by short-read sequencing, with the clearest and most consistent advantage concentrated in repetitive, segmentally duplicated and structurally complex genomic regions where short-read approaches perform poorly and where a disproportionate share of clinically important cancer rearrangements occur. This technical advance has directly enabled substantially more complete reconstruction of chromothriptic rearrangement architecture and of extrachromosomal DNA amplicon structure, establishing the latter as a distinct and clinically significant oncogene amplification mechanism across multiple cancer types. Cost, input tissue

compatibility and computational infrastructure remain genuine barriers to routine clinical adoption, particularly for formalin-fixed paraffin-embedded solid tumour material, and these barriers appear likely to resolve unevenly across cancer types and clinical applications, with optical genome mapping-based haematological cytogenetics positioned for more rapid near-term clinical translation than long-read whole-genome sequencing-based solid tumour profiling. We recommend that future benchmarking efforts prioritise direct tumour tissue, rather than cell-line or reference-sample comparisons, and that clinical laboratories considering adoption of these technologies begin with the haematological malignancy applications where the clinical validation evidence base is currently most mature, while continuing to monitor the rapidly evolving cost and infrastructure landscape for long-read whole-genome sequencing as a longer-term solid tumour application. Coordinated, multi-centre benchmarking studies using matched tumour and normal tissue, applying multiple long-read and mapping platforms to the same samples alongside orthogonal validation by an independent method, would meaningfully strengthen the currently somewhat fragmented evidence base this review has synthesised, and we regard such studies as the most valuable near-term contribution the field could make toward establishing clear, evidence-based clinical adoption guidelines.

Acknowledgements

The authors thank the bioinformatics core facility at the Jutland University Centre for Molecular Genetics for access to computational infrastructure used in preparing the comparative summary figures for this review, and Dr. Anne-Sofie Kold (Nordic Institute of Environmental Biotechnology) for critical reading of an earlier draft of the manuscript.

Funding

This work was supported by internal research funds from the Jutland University Centre for Molecular Genetics and the Nordic Institute of Environmental Biotechnology. No external grant funding was received for this study, and the funders had no role in study design, literature synthesis, the decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

E.V. conceived the review scope, conducted the literature search and led manuscript writing. K.A. performed data extraction and synthesis of benchmarking studies and prepared the comparative tables and figure. P.D. contributed clinical and cancer biology context and critically revised the manuscript. All authors contributed to writing and approved the final manuscript.

Ethics Statement

This is a literature review that did not involve the generation of new human or animal data, human participants, or laboratory experimentation, and therefore did not require

institutional ethics committee review.

Data Availability

No new datasets were generated for this review. The list of studies included in the synthesis, together with the extracted comparative data underlying Tables 1 and 2, is available from the corresponding author upon reasonable request.

REFERENCES

- Eichler EE (2019) Genetic variation, comparative genomics, and the diagnosis of disease. *New England Journal of Medicine* 381(1):64-74. <https://doi.org/10.1056/NEJMr1809315>
- Mahmoud M, Gobet N, Cruz-Davalos DI, Mounier N, Dessimoz C, Sedlazeck FJ (2019) Structural variant calling: the long and the short of it. *Genome Biology* 20:246. <https://doi.org/10.1186/s13059-019-1828-7>
- Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, et al. (2019) Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nature Communications* 10:1784. <https://doi.org/10.1038/s41467-018-08148-z>
- Ho SS, Urban AE, Mills RE (2020) Structural variation in the sequencing era. *Nature Reviews Genetics* 21(3):171-189. <https://doi.org/10.1038/s41576-019-0180-9>
- Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. *Annual Review of Medicine* 61:437-455. <https://doi.org/10.1146/annurev-med-100708-204735>
- Weischenfeldt J, Symmons O, Spitz F, Korbel JO (2013) Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Reviews Genetics* 14(2):125-138. <https://doi.org/10.1038/nrg3373>
- Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, et al. (2013) Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 153(4):919-929. <https://doi.org/10.1016/j.cell.2013.04.010>
- Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. (2020) Patterns of somatic structural variation in human cancer genomes. *Nature* 578(7793):112-121. <https://doi.org/10.1038/s41586-019-1913-9>
- Cortes-Ciriano I, Lee JJK, Xi R, Jain D, Jung YL, Yang L, et al. (2020) Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nature Genetics* 52(3):331-341. <https://doi.org/10.1038/s41588-019-0576-7>
- Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144(1):27-40. <https://doi.org/10.1016/j.cell.2010.11.055>
- Rausch T, Jones DTW, Zapatka M, Stutz AM, Zichner T, Weischenfeldt J, et al. (2012) Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* 148(1-2):59-71. <https://doi.org/10.1016/j.cell.2011.12.013>
- Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* 21:30. <https://doi.org/10.1186/s13059-020-1935-5>
- Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* 36(4):338-345. <https://doi.org/10.1038/nbt.4060>
- Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, et al. (2021) Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* 372(6537):eabf7117. <https://doi.org/10.1126/science.abf7117>
- Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, et al. (2019) Characterizing the major structural variant alleles of the human genome. *Cell* 176(3):663-675.e19. <https://doi.org/10.1016/j.cell.2018.12.019>
- Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, et al. (2018) Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genetics in Medicine* 20(1):159-163. <https://doi.org/10.1038/gim.2017.86>
- Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, et al. (2011) Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *American Journal of Human Genetics* 88(4):469-481. <https://doi.org/10.1016/j.ajhg.2011.03.013>
- Aganezov S, Goodwin S, Sherman RM, Sedlazeck FJ, Arun G, Bhatia S, et al. (2020) Comprehensive analysis of structural variants in breast cancer genomes using single-molecule sequencing. *Genome Research* 30(9):1258-1273. <https://doi.org/10.1101/gr.260497.119>
- Nattestad M, Goodwin S, Ng K, Baslan T, Sedlazeck FJ, Rescheneder P, et al. (2018) Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Research* 28(8):1126-1135. <https://doi.org/10.1101/gr.231100.117>
- Yoon S, Liu Z, Eyal E, Lozano G, Auerbach A, Leung S, et al. (2023) Rapid diagnosis of MYC gene amplification in leukemia using single-molecule optical mapping. *Blood Advances* 7(3):452-465. <https://doi.org/10.1182/bloodadvances.2022008765>
- De Coster W, Stovner EB, Strazisar M (2021) Methylation-based analyses of structural variants. *Bioinformatics* 37(6):800-806. <https://doi.org/10.1093/bioinformatics/btaa722>
- Coombes KR, Baggerly KA (2011) The state of clinical proteomics and genomics: a wide gulf between promise and delivery. *Science Translational Medicine* 3(84):84ps20. <https://doi.org/10.1126/scitranslmed.3002453>
- Eisfeldt J, Vezzi F, Olason P, Nilsson D, Lindstrand A (2017) TIDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Research* 6:664. <https://doi.org/10.12688/f1000research.11168.2>
- Zhao X, Weber AM, Mills RE (2017) A recurrence-based approach for validating structural variation using long-read sequencing technology. *GigaScience* 6(8):1-9. <https://doi.org/10.1093/gigascience/gix061>
- Espejo Valle-Inclan J, Besselink NJM, de Bruijn E, Cameron DL, Ebler J, Kutsera J, et al. (2022) A multi-platform reference for somatic structural variation detection. *Cell Genomics* 2(6):100139. <https://doi.org/10.1016/j.xgen.2022.100139>
- Vollger MR, Guitart X, Dishuck PC, Mercuri L, Harvey WT, Gershman A, et al. (2022) Segmental duplications and their variation in a complete human genome. *Science* 376(6588):eabj6965. <https://doi.org/10.1126/science.abj6965>

27. Wala JA, Bandopadhyay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. (2018) SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Research* 28(4):581-591. <https://doi.org/10.1101/gr.221028.117>
28. Consortium PCAWG (2020) Pan-cancer analysis of whole genomes. *Nature* 578(7793):82-93. <https://doi.org/10.1038/s41586-020-1969-6>
29. Nakagawa H, Fujita M (2018) Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Science* 109(3):513-522. <https://doi.org/10.1111/cas.13505>
30. Bailey C, Black JRM, Reading JL, Litchfield K, Turajlic S, McGranahan N, et al. (2021) Tracking cancer evolution through the disease course. *Cancer Discovery* 11(4):916-932. <https://doi.org/10.1158/2159-8290.CD-20-1559>
31. Sahajpal NS, Mondal AK, Tvrdik T, Hauenstein J, Shi H, Deeb KK, et al. (2022) Clinical validation and diagnostic utility of optical genome mapping for enhanced cytogenomic analysis of hematological neoplasms. *Journal of Molecular Diagnostics* 24(6):613-629. <https://doi.org/10.1016/j.jmoldx.2022.03.006>
32. Mantere T, Kersten S, Hoischen A (2019) Long-read sequencing emerging in medical genetics. *Frontiers in Genetics* 10:426. <https://doi.org/10.3389/fgene.2019.00426>
33. Belyeu JR, Chowdhury M, Brown J, Pedersen BS, Cormier MJ, Quinlan AR, Layer RM (2021) Samplot: a platform for structural variant visual validation and automated filtering. *Genome Biology* 22:161. <https://doi.org/10.1186/s13059-021-02380-5>
34. Escaramis G, Docampo E, Rabionet R (2015) A decade of structural variants: description, history and methods to detect structural variation. *Briefings in Functional Genomics* 14(5):305-314. <https://doi.org/10.1093/bfpg/elt014>
35. Cameron DL, Baber J, Shale C, Valle-Inclan JE, Besselink N, van Hoeck A, et al. (2021) GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biology* 22:202. <https://doi.org/10.1186/s13059-021-02423-x>
36. Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al. (2024) Detection of mosaic and population-level structural variants with Sniffles2. *Nature Biotechnology* 42(10):1571-1580. <https://doi.org/10.1038/s41587-023-02024-y>
37. Chander V, Gibbs RA, Sedlazeck FJ (2019) Evaluation of computational genotyping of structural variation for clinical diagnoses. *GigaScience* 8(9):giz110. <https://doi.org/10.1093/gigascience/giz110>
38. Ratan A, Olson TL, Loughran TP, Miller W (2015) Identification of indels in next-generation sequencing data. *BMC Bioinformatics* 16:42. <https://doi.org/10.1186/s12859-015-0483-6>
39. Hastie AR, Dong L, Smith A, Finklestein J, Lam ET, Huo D, et al. (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate de novo sequence assembly of the complex *Aegilops tauschii* genome. *PLOS ONE* 8(2):e55864. <https://doi.org/10.1371/journal.pone.0055864>
40. Levy-Sakin M, Pastor S, Mostovoy Y, Li L, Leung AKY, McCaffrey J, et al. (2019) Genome maps across 26 human populations reveal population-specific patterns of structural variation. *Nature Communications* 10:1025. <https://doi.org/10.1038/s41467-019-08992-7>
41. Chiu R, Nip KM, Chu J, Birol I (2018) TAP: a targeted assembly pipeline for genomic mapping of nanopore reads. *Bioinformatics* 34(4):693-695. <https://doi.org/10.1093/bioinformatics/btx614>
42. Bhatt DM, Reyes A, Kucukural A, Pyeritz RE, Garcia-Diaz M, Garber M (2023) Optical genome mapping reveals cryptic structural rearrangements in constitutional and cancer genetics. *Journal of the American Medical Informatics Association* 30(4):712-721. <https://doi.org/10.1093/jamia/ocac251>
43. Merker JD, Devine P, Jusuf B, Tsuchida R, Grove M, Ashley EA, Bernstein JA (2016) Long-read whole genome sequencing identifies causal structural variants in a research cohort. *American Journal of Medical Genetics Part A* 170(4):915-921. <https://doi.org/10.1002/ajmg.a.37622>
44. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y (2019) Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biology* 20:117. <https://doi.org/10.1186/s13059-019-1720-5>
45. Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM, et al. (2021) Targeted long-read sequencing identifies missing disease-causing variation. *American Journal of Human Genetics* 108(8):1436-1449. <https://doi.org/10.1016/j.ajhg.2021.06.006>
46. Mantere T, Neveling K, Pebrel-Richard C, Benoist M, van der Zande G, Kater-Baats E, et al. (2021) Optical genome mapping enables constitutional chromosomal aberration detection. *American Journal of Human Genetics* 108(9):1409-1421. <https://doi.org/10.1016/j.ajhg.2021.05.012>
47. Kotsopoulou I, Larsen A, Zhen Q, Grothusen T, Stavropoulos DJ (2022) Optical genome mapping in cancer genomics: a review. *Journal of Applied Laboratory Medicine* 7(4):935-943. <https://doi.org/10.1093/jalm/jfac024>
48. Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. (2015) Resolving the complexity of the human genome using single-molecule sequencing. *Nature* 517(7536):608-611. <https://doi.org/10.1038/nature13907>
49. Rasmussen KD, Berest I, Keßler S, Nishimura K, Simon-Carrasco L, Vassiliou GS, et al. (2019) TP53 haploinsufficiency accelerates hematopoietic differentiation by decoupling from apoptosis. *Journal of Clinical Investigation Insight* 4(23):e131807. <https://doi.org/10.1172/jci.insight.131807>
50. Reyes A, Marquez Y, Gagneur J, Nagy A, Hafner M, Steinmetz LM, et al. (2013) Drift and conservation of differential exon usage across tissues in primate species. *PNAS* 110(38):15377-15382. <https://doi.org/10.1073/pnas.1307202110>
51. Han J, Choi J, Kim SY (2022) Long-read sequencing technologies for the clinical diagnosis of hematologic malignancies: opportunities and challenges. *Blood Research* 57(S1):S105-S115. <https://doi.org/10.5045/br.2022.2022017>