



Universiteit
Leiden
The Netherlands

Algorithms for structural variant detection

Lin, J.

Citation

Lin, J. (2022, June 24). *Algorithms for structural variant detection*. Retrieved from <https://hdl.handle.net/1887/3391016>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3391016>

Note: To cite this publication please use the final published version (if applicable).

Bibliography

- [1] Alkan C, Coe BP, Eichler EE: Genome structural variation discovery and genotyping. *Nat Rev Genet* 2011, 12:363–376.
- [2] Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25:1754–1760.
- [3] Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009, 10:R25.
- [4] Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT, et al.: A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* 2018, 36:983–987.
- [5] Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, et al.: An integrated map of structural variation in 2,504 human genomes. *Nature* 2015, 526:75–81.
- [6] Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, Khurana E, Waszak S, Korbelt JO, Haber JE, et al.: Patterns of somatic structural variation in human cancer genomes. *Nature* 2020, 578:112–121.
- [7] Collins RL, Brand H, Karczewski KJ, Zhao X, Alfoldi J, Francioli LC, Khera AV, Lowther C, Gauthier LD, Wang H, et al.: A structural variation reference for medical and population genetics. *Nature* 2020, 581:444–451.
- [8] Ho SS, Urban AE, Mills RE: Structural variation in the sequencing era. *Nat Rev Genet* 2020, 21:171–189.

BIBLIOGRAPHY

- [9] Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL, et al.: Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun* 2019, 10:1784.
- [10] Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, et al.: Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* 2021, 372.
- [11] Quinlan AR, Hall IM: Characterizing complex structural variation in germline and somatic genomes. *Trends Genet* 2012, 28:43–53.
- [12] Collins RL, Brand H, Redin CE, Hanscom C, Antolik C, Stone MR, Glessner JT, Mason T, Pregno G, Dorrani N, et al.: Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome Biol* 2017, 18:36.
- [13] Spielmann M, Lupianez DG, Mundlos S: Structural variation in the 3D genome. *Nat Rev Genet* 2018, 19:453–467.
- [14] Carvalho CM, Lupski JR: Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* 2016, 17:224–238.
- [15] Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, Szallasi Z, Barry WT, Winer EP, Tung NM, et al.: Homologous Recombination Deficiency (HRD) score predicts response to Platinum-containing neoadjuvant chemotherapy in patients with Triple-Negative Breast Cancer. *Clin Cancer Res* 2016, 22:3764–3773.
- [16] Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: Versatile and open software for comparing large genomes. *Genome Biol* 2004, 5:R12.
- [17] Li H: Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018, 34:3094–3100.
- [18] Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC: Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* 2018, 15:461–468.

- [19] Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, Brooks S, Howe E, Porubsky D, Logsdon GA, et al.: Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 2020, 585:79–84.
- [20] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25:2078–2079.
- [21] Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using De Bruijn graphs. *Genome Res* 2008, 18:821–829.
- [22] Ruan J, Li H: Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 2020, 17:155–158.
- [23] Cheng H, Concepcion GT, Feng X, Zhang H, Li H: Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* 2021, 18:170–175.
- [24] Simpson JT, Durbin R: Efficient de novo assembly of large genomes using compressed data structures. *Genome Res* 2012, 22:549–556.
- [25] Zepeda-Mendoza CJ, Morton CC: The iceberg under water: unexplored complexity of chromoanagenesis in congenital disorders. *Am J Hum Genet* 2019, 104:565–577.
- [26] Li Z, Chen Y, Mu D, Yuan J, Shi Y, Zhang H, Gan J, Li N, Hu X, Liu B, et al.: Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and De-Bruijn-graph. *Brief Funct Genomics* 2012, 11:25–37.
- [27] Marschall T, Costa IG, Canzar S, Bauer M, Klau GW, Schliep A, Schonhuth A: CLEVER: clique-enumerating variant finder. *Bioinformatics* 2012, 28:2875–2882.
- [28] Sherman RM, Salzberg SL: Pan-genomics in the human genome era. *Nat Rev Genet* 2020, 21:243–254.
- [29] Li H, Feng X, Chu C: The design and construction of reference pangenome graphs with minigraph. *Genome Biol* 2020, 21:265.
- [30] Hickey G, Heller D, Monlong J, Sibbesen JA, Siren J, Eizenga J, Dawson ET, Garrison E, Novak AM, Paten B: Genotyping structural

BIBLIOGRAPHY

- variants in pangenome graphs using the vg toolkit. *Genome Biol* 2020, 21:35.
- [31] Garrison E, Siren J, Novak AM, Hickey G, Eizenga JM, Dawson ET, Jones W, Garg S, Markello C, Lin MF, et al.: Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat Biotechnol* 2018, 36:875–879.
- [32] Sibbesen JA, Maretty L, Danish Pan-Genome C, Krogh A: Accurate genotyping across variant classes and lengths using variant graphs. *Nat Genet* 2018, 50:1054–1059.
- [33] Rautiainen M, Makinen V, Marschall T: Bit-parallel sequence-to-graph alignment. *Bioinformatics* 2019, 35:3599–3607.
- [34] Rautiainen M, Marschall T: GraphAligner: rapid and versatile sequence-to-graph alignment. *Genome Biol* 2020, 21:253.
- [35] Hadi K, Yao X, Behr JM, Deshpande A, Xanthopoulos C, Tian H, Kudman S, Rosiene J, Darmofal M, DeRose J, et al.: Distinct classes of complex structural variation uncovered across thousands of cancer genome graphs. *Cell* 2020, 183:197–210 e132.
- [36] Ye K, Schulz MH, Long Q, Apweiler R, Ning Z: Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009, 25:2865–2871.
- [37] Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO: DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 2012, 28:i333–i339.
- [38] Layer RM, Chiang C, Quinlan AR, Hall IM: LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* 2014, 15:R84.
- [39] Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, et al.: BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009, 6:677–681.
- [40] Cameron DL, Di Stefano L, Papenfuss AT: Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat Commun* 2019, 10:3240.

- [41] Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y: Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol* 2019, 20:117.
- [42] Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, Cox AJ, Kruglyak S, Saunders CT: Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 2016, 32:1220–1222.
- [43] Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, Tsai PC, Casasent A, Waters J, Zhang H, et al.: Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* 2016, 48:1119–1130.
- [44] Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, Alexandrov LB, Van Loo P, Haugland HK, Lilleng PK, et al.: Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell* 2017, 32:169–184 e167.
- [45] Nattestad M, Goodwin S, Ng K, Baslan T, Sedlazeck FJ, Rescheneder P, Garvin T, Fang H, Gurtowski J, Hutton E, et al.: Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Res* 2018, 28:1126–1135.
- [46] Sanchis-Juan A, Stephens J, French CE, Gleadall N, Megy K, Penkett C, Shamardina O, Stirrups K, Delon I, Dewhurst E, et al.: Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med* 2018, 10:95.
- [47] Greer SU, Nadauld LD, Lau BT, Chen J, Wood-Bouwens C, Ford JM, Kuo CJ, Ji HP: Linked read sequencing resolves complex genomic rearrangements in gastric cancer metastases. *Genome Med* 2017, 9:57.
- [48] Lee JJ, Park S, Park H, Kim S, Lee J, Lee J, Youk J, Yi K, An Y, Park IK, et al.: Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. *Cell* 2019, 177:1842–1857 e1821.
- [49] Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, et al.: Punctuated evolution of prostate cancer genomes. *Cell* 2013, 153:666–677.

BIBLIOGRAPHY

- [50] Korbelt JO, Campbell PJ: Criteria for inference of chromothripsis in cancer genomes. *Cell* 2013, 152:1226–1236.
- [51] Sanders AD, Meiers S, Ghareghani M, Porubsky D, Jeong H, van Vliet M, Rausch T, Richter-Pechanska P, Kunz JB, Jenni S, et al.: Single-cell analysis of structural variations and complex rearrangements with tri-channel processing. *Nat Biotechnol* 2019.
- [52] Carvalho CMB, Lupski JR: Mechanisms underlying structural variant formation in genomic disorders. *Nature Reviews Genetics* 2016, 17:224–238.
- [53] Malhotra A, Lindberg M, Faust GG, Leibowitz ML, Clark RA, Layer RM, Quinlan AR, Hall IM: Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res* 2013, 23:762–776.
- [54] Ye K, Wang J, Jayasinghe R, Lameijer EW, McMichael JF, Ning J, McLellan MD, Xie M, Cao S, Yellapantula V, et al.: Systematic discovery of complex insertions and deletions in human cancers. *Nat Med* 2016, 22:97–104.
- [55] Zhang CZ, Leibowitz ML, Pellman D: Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev* 2013, 27:2513–2530.
- [56] Soylev A, Le TM, Amini H, Alkan C, Hormozdiari F: Discovery of tandem and interspersed segmental duplications using high-throughput sequencing. *Bioinformatics* 2019, 35:3923–3930.
- [57] Zhao X, Emery SB, Myers B, Kidd JM, Mills RE: Resolving complex structural genomic rearrangements using a randomized approach. *Genome Biol* 2016, 17:126.
- [58] Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, Speed TP, Papenfuss AT: GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. *Genome Res* 2017, 27:2050–2060.
- [59] Arthur JG, Chen X, Zhou B, Urban AE, Wong WH: Detection of complex structural variation from paired-end sequencing data. *bioRxiv* 2017:200170.

- [60] Liao VCC, Chen MS: DFSP: a Depth-First SPelling algorithm for sequential pattern mining of biological sequences. *Knowledge and Information Systems* 2014, 38:623–639.
- [61] Tsai HP, Yang DN, Chen MS: Mining group movement patterns for tracking moving objects efficiently. *Ieee Transactions on Knowledge and Data Engineering* 2011, 23:266–281.
- [62] Huang Y, Zhang LQ, Zhang PS: A framework for mining sequential patterns from spatio-temporal event data sets. *Ieee Transactions on Knowledge and Data Engineering* 2008, 20:433–448.
- [63] Ye K, Kosters WA, Ijzerman AP: An efficient, versatile and scalable pattern growth approach to mine frequent patterns in unaligned protein sequences. *Bioinformatics* 2007, 23:687–693.
- [64] Pei J, Han J, Wang W: Constraint-based sequential pattern mining: the pattern-growth methods. *Journal of Intelligent Information Systems* 2007, 28:133–160.
- [65] Pei J, Han JW, Mortazavi-Asl B, Wang JY, Pinto H, Chen QM, Dayal U, Hsu MC: Mining sequential patterns by pattern-growth: the PrefixSpan approach. *IEEE Transactions on Knowledge and Data Engineering* 2004, 16:1424–1440.
- [66] Li H, Homer N: A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform* 2010, 11:473–483.
- [67] Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25:1754–1760.
- [68] Bolognini D, Sanders A, Korbel JO, Magi A, Benes V, Rausch T: VISOR: a versatile haplotype-aware structural variant simulator for short and long read sequencing. *Bioinformatics* 2019.
- [69] McPherson A, Wu C, Wyatt AW, Shah S, Collins C, Sahinalp SC: nFuse: discovery of complex genomic rearrangements in cancer using high-throughput sequencing. *Genome Res* 2012, 22:2250–2261.
- [70] Dzamba M, Ramani AK, Buczkowicz P, Jiang Y, Yu M, Hawkins C, Brudno M: Identification of complex genomic rearrangements in cancers using CouGaR. *Genome Res* 2017, 27:107–117.

BIBLIOGRAPHY

- [71] Delcher AL, Phillippy A, Carlton J, Salzberg SL: Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* 2002, 30:2478–2483.
- [72] Zhao X, Weber AM, Mills RE: A recurrence-based approach for validating structural variation using long-read sequencing technology. *Gigascience* 2017, 6:1–9.
- [73] Ottaviani D, LeCain M, Sheer D: The role of microhomology in genomic structural variation. *Trends Genet* 2014, 30:85–94.
- [74] Kramara J, Osia B, Malkova A: Break-induced replication: the where, the why, and the how. *Trends Genet* 2018, 34:518–531.
- [75] Hartlerode AJ, Willis NA, Rajendran A, Manis JP, Scully R: Complex breakpoints and template switching associated with non-canonical termination of homologous recombination in mammalian cells. *PLoS Genet* 2016, 12:e1006410.
- [76] Zhou W, Zhang F, Chen X, Shen Y, Lupski JR, Jin L: Increased genome instability in human DNA segments with self-chains: homology-induced structural variations via replicative mechanisms. *Hum Mol Genet* 2013, 22:2642–2651.
- [77] Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, Zhang C, Ren X, Protopopov A, Chin L, et al.: Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 2013, 153:919–929.
- [78] Chen W, McKenna A, Schreiber J, Haeussler M, Yin Y, Agarwal V, Noble WS, Shendure J: Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res* 2019, 47:7989–8003.
- [79] Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, De Angeli P, Palenikova P, Khodak A, Kiselev V, Kosicki M, et al.: Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol* 2018.
- [80] Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, Foye A, Kothari V, Perry MD, Bailey AM, et al.: Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* 2018, 175:889.

- [81] Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, Shiah YJ, Yousif F, Lin X, Masella AP, et al.: Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* 2017, 541:359–364.
- [82] Fujimoto A, Wong JH, Yoshii Y, Akiyama S, Tanaka A, Yagi H, Shigemizu D, Nakagawa H, Mizokami M, Shimada M: Whole-genome sequencing with long reads reveals complex structure and origin of structural variation in human genetic variations and somatic mutations in cancer. *Genome Med* 2021, 13:65.
- [83] Bolognini D, Sanders A, Korbel JO, Magi A, Benes V, Rausch T: VISOR: a versatile haplotype-aware structural variant simulator for short- and long-read sequencing. *Bioinformatics* 2020, 36:1267–1269.
- [84] Krumsiek J, Arnold R, Rattei T: Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 2007, 23:1026–1028.
- [85] Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010, 26:841–842.
- [86] Guennewig B, Lim J, Marshall L, McCorkindale AN, Paasila PJ, Patrick E, Kril JJ, Halliday GM, Cooper AA, Sutherland GT: Defining early changes in Alzheimer’s disease from RNA sequencing of brain regions differentially affected by pathology. *Sci Rep* 2021, 11:4865.
- [87] Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, Dougherty ML, Nelson BJ, Shah A, Dutcher SK, et al.: Characterizing the major structural variant alleles of the human genome. *Cell* 2019, 176:663–675 e619.
- [88] Belyeu JR, Chowdhury M, Brown J, Pedersen BS, Cormier MJ, Quinlan AR, Layer RM: Samplot: a platform for structural variant visual validation and automated filtering. *Genome Biol* 2021, 22:161.
- [89] Nattestad M, Aboukhalil R, Chin CS, Schatz MC: Ribbon: intuitive visualization for complex genomic variation. *Bioinformatics* 2021, 37:413–415.
- [90] Aganezov S, Goodwin S, Sherman RM, Sedlazeck FJ, Arun G, Bhatia S, Lee I, Kirsche M, Wappel R, Kramer M, et al.: Comprehensive analysis of structural variants in breast cancer genomes using single-molecule sequencing. *Genome Res* 2020, 30:1258–1273.

BIBLIOGRAPHY

- [91] Jiang T, Liu S, Cao S, Liu Y, Cui Z, Wang Y, Guo H: Long-read sequencing settings for efficient structural variation detection based on comprehensive evaluation. *BMC Bioinformatics* 2021, 22:552.
- [92] Zook JM, Hansen NF, Olson ND, Chapman L, Mullikin JC, Xiao C, Sherry S, Koren S, Phillippy AM, Boutros PC, et al.: A robust benchmark for detection of germline large deletions and insertions. *Nat Biotechnol* 2020, 38:1347–1355.
- [93] Sone J, Mitsuhashi S, Fujita A, Mizuguchi T, Hamanaka K, Mori K, Koike H, Hashiguchi A, Takashima H, Sugiyama H, et al.: Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat Genet* 2019, 51:1215–1221.
- [94] Hiatt SM, Lawlor JMJ, Handley LH, Ramaker RC, Rogers BB, Partridge EC, Boston LB, Williams M, Plott CB, Jenkins J, et al.: Long-read genome sequencing for the molecular diagnosis of neurodevelopmental disorders. *HGG Adv* 2021, 2.
- [95] Pauper M, Kucuk E, Wenger AM, Chakraborty S, Baybayan P, Kwint M, van der Sanden B, Nelen MR, Derks R, Brunner HG, et al.: Long-read trio sequencing of individuals with unsolved intellectual disability. *Eur J Hum Genet* 2021, 29:637–648.
- [96] Gong L, Wong CH, Cheng WC, Tjong H, Menghi F, Ngan CY, Liu ET, Wei CL: Picky comprehensively detects high-resolution structural variants in nanopore long reads. *Nat Methods* 2018, 15:455–460.
- [97] Zhou B, Ho SS, Greer SU, Zhu X, Bell JM, Arthur JG, Spies N, Zhang X, Byeon S, Pattni R, et al.: Comprehensive, integrated, and phased whole-genome analysis of the primary ENCODE cell line K562. *Genome Res* 2019, 29:472–484.
- [98] Sakamoto Y, Xu L, Seki M, Yokoyama TT, Kasahara M, Kashima Y, Ohashi A, Shimada Y, Motoi N, Tsuchihara K, et al.: Long-read sequencing for non-small-cell lung cancer genomes. *Genome Res* 2020, 30:1243–1257.
- [99] Zhou B, Ho SS, Greer SU, Spies N, Bell JM, Zhang X, Zhu X, Arthur JG, Byeon S, Pattni R, et al.: Haplotype-resolved and integrated genome analysis of the cancer cell line HepG2. *Nucleic Acids Res* 2019, 47:3846–3861.

- [100] Peneau C, Imbeaud S, La Bella T, Hirsch TZ, Caruso S, Calderaro J, Paradis V, Blanc JF, Letouze E, Nault JC, et al.: Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma. *Gut* 2021.
- [101] De Roeck A, De Coster W, Bossaerts L, Cacace R, De Pooter T, Van Dongen J, D’Hert S, De Rijk P, Strazisar M, Van Broeckhoven C, Slegers K: NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol* 2019, 20:239.
- [102] Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, Liu Y, Liu B, Wang Y: Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol* 2020, 21:189.
- [103] Heller D, Vingron M: SVIM: structural variant identification using mapped long reads. *Bioinformatics* 2019, 35:2907–2915.
- [104] Tham CY, Tirado-Magallanes R, Goh Y, Fullwood MJ, Koh BTH, Wang W, Ng CH, Chng WJ, Thiery A, Tenen DG, Benoukraf T: NanoVar: accurate characterization of patients’ genomic structural variants using low-depth nanopore sequencing. *Genome Biol* 2020, 21:56.
- [105] Cretu Stancu M, van Roosmalen MJ, Renkens I, Nieboer MM, Middeldkamp S, de Ligt J, Pregno G, Giachino D, Mandrile G, Espejo Valle-Inclan J, et al.: Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* 2017, 8:1326.
- [106] Hiltmann S, Jenster G, Trapman J, van der Spek P, Stubbs A: Discriminating somatic and germline mutations in tumor DNA samples without matching normals. *Genome Res* 2015, 25:1382–1390.
- [107] Franco I, Helgadottir HT, Moggio A, Larsson M, Vrtacnik P, Johansson A, Norgren N, Lundin P, Mas-Ponte D, Nordstrom J, et al.: Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome Biol* 2019, 20:285.
- [108] Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, Yardimci GG, Chakraborty A, Bann DV, Wang Y, et al.: Integrative detection and

BIBLIOGRAPHY

- analysis of structural variation in cancer genomes. *Nat Genet* 2018, 50:1388–1398.
- [109] Thibodeau ML, O’Neill K, Dixon K, Reisle C, Mungall KL, Krzywinski M, Shen Y, Lim HJ, Cheng D, Tse K, et al.: Improved structural variant interpretation for hereditary cancer susceptibility using long-read sequencing. *Genet Med* 2020, 22:1892–1897.
- [110] Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, Paczkowska M, Reynolds S, Wyczalkowski MA, Oak N, et al.: Pathogenic germline variants in 10,389 adult cancers. *Cell* 2018, 173:355–370 e314.
- [111] Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, Johann PD, Balasubramanian GP, Segura-Wang M, Brabetz S, et al.: The landscape of genomic alterations across childhood cancers. *Nature* 2018, 555:321–327.
- [112] Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bahler J, Sedlazeck FJ: Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 2017, 8:14061.
- [113] Beyter D, Ingimundardottir H, Oddsson A, Eggertsson HP, Bjornsson E, Jonsson H, Atlason BA, Kristmundsdottir S, Mehringer S, Hardarson MT, et al.: Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. *Nat Genet* 2021, 53:779–786.
- [114] Alvarez EG, Demeulemeester J, Otero P, Jolly C, Garcia-Souto D, Pequeno-Valtierra A, Zamora J, Tojo M, Temes J, Baez-Ortega A, et al.: Aberrant integration of Hepatitis B virus DNA promotes major restructuring of human hepatocellular carcinoma genome architecture. *Nat Commun* 2021, 12:6910.
- [115] Zhou A, Lin T, Xing J: Evaluating nanopore sequencing data processing pipelines for structural variation identification. *Genome Biol* 2019, 20:237.
- [116] Chiu R, Rajan-Babu IS, Friedman JM, Birol I: Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biol* 2021, 22:224.