REFRENCES.

1 Quinlan, A. R. & Hall, I. M. Characterizing complex structural variation in germline and somatic genomes. *Trends Genet* **28**, 43-53, doi:10.1016/j.tig.2011.10.002 (2012).

2 Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome research* **21**, 974-984, doi:10.1101/gr.114876.110 (2011).

3 Yang, L. *et al.* Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* **153**, 919-929, doi:10.1016/j.cell.2013.04.010 (2013).

4 Lindberg, M. R., Hall, I. M. & Quinlan, A. R. Population-based structural variation discovery with Hydra-Multi. *Bioinformatics* **31**, 1286-1289, doi:10.1093/bioinformatics/btu771 (2015).

5 Korbel, J. O. *et al.* PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol* **10**, R23, doi:10.1186/gb-2009-10-2-r23 (2009).

6 Fan, X., Abbott, T. E., Larson, D. & Chen, K. BreakDancer - Identification of Genomic Structural Variation from Paired-End Read Mapping. *Curr Protoc Bioinformatics* **2014**, doi:10.1002/0471250953.bi1506s45 (2014).

7 Ye, K., Schulz, M. H., 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* **25**, 2865-2871, doi:10.1093/bioinformatics/btp394 (2009).

8 Malhotra, A. *et al.* Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome research* **23**, 762-776, doi:10.1101/gr.143677.112 (2013).

9 Malhotra, A. *et al.* Ploidy-Seq: inferring mutational chronology by sequencing polyploid tumor subpopulations. *Genome Med* **7**, 6, doi:10.1186/s13073-015-0127-5 (2015).

10 Zhang, Z. D. *et al.* Identification of genomic indels and structural variations using split reads. *BMC genomics* **12**, 375, doi:10.1186/1471-2164-12-375 (2011).

11 Simpson, J. T. & Durbin, R. Efficient de novo assembly of large genomes using compressed data structures. *Genome research* **22**, 549-556, doi:10.1101/gr.126953.111 (2012).

12 Chen, K. *et al.* TIGRA: a targeted iterative graph routing assembler for breakpoint assembly. *Genome research* **24**, 310-317, doi:10.1101/gr.162883.113 (2014).

13 Abyzov, A. & Gerstein, M. AGE: defining breakpoints of genomic structural variants at single-nucleotide resolution, through optimal alignments with gap excision. *Bioinformatics* **27**, 595-603, doi:10.1093/bioinformatics/btq713 (2011).

14 Weckselblatt, B. & Rudd, M. K. Human structural variation: mechanisms of chromosome rearrangements. *Trends Genet*, doi:10.1016/j.tig.2015.05.010 (2015).

15 Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nat Rev Genet* **10**, 551-564, doi:10.1038/nrg2593 (2009).

16 Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS genetics* **5**, e1000327, doi:10.1371/journal.pgen.1000327 (2009).

17 Lam, H. Y. *et al.* Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nature biotechnology* **28**, 47-55, doi:10.1038/nbt.1600 (2010).

18 Klambauer, G. *et al.* cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res* **40**, e69, doi:10.1093/nar/gks003 (2012).

19 Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxford, England)* **28**, i333-i339, doi:10.1093/bioinformatics/bts378 (2012).

20 Handsaker, R. E. *et al.* Large multiallelic copy number variations in humans. *Nature genetics* **47**, 296-303, doi:10.1038/ng.3200 (2015).

21 Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* **15**, R84, doi:10.1186/gb-2014-15-6-r84 (2014).

22 Levandowsky, M. & Winter, D. Distance between Sets. *Nature* **234**, 34-35 (1971).

23 Sudmant, P. H. An integrated map of structural variation in 2,504 human genomes. *Nature* **Accepted, in print** (2015).

24 Wong, K. K. *et al.* A comprehensive analysis of common copy-number variations in the human genome. *American journal of human genetics* **80**, 91-104, doi:10.1086/510560 (2007).

25 Bailey, J. A., Kidd, J. M. & Eichler, E. E. Human copy number polymorphic genes. *Cytogenet Genome Res* **123**, 234-243, doi:10.1159/000184713 (2008).

26 Habegger, L. *et al.* VAT: a computational framework to functionally annotate variants in personal genomes within a cloud-computing environment. *Bioinformatics* **28**, 2267-2269, doi:10.1093/bioinformatics/bts368 (2012).

27 Lu, Z. J. *et al.* Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome research* **21**, 276-285, doi:10.1101/gr.110189.110 (2011).

28 Mu, X. J., Lu, Z. J., Kong, Y., Lam, H. Y. & Gerstein, M. B. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. *Nucleic Acids Res* **39**, 7058-7076, doi:10.1093/nar/gkr342 (2011).

29 Rozowsky, J. *et al.* PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature biotechnology* **27**, 66-75, doi:10.1038/nbt.1518 (2009).

30 Harmanci, A., Rozowsky, J. & Gerstein, M. MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. *Genome Biol* **15**, 474, doi:10.1186/s13059-014-0474-3 (2014).

31 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

32 Cheng, C. *et al.* A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. *Genome Biol* **12**, R15, doi:10.1186/gb-2011-12-2-r15 (2011).

33 Cheng, C. *et al.* Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome research* **22**, 1658-1667, doi:10.1101/gr.136838.111 (2012).

34 Gerstein, M. B. *et al.* Comparative analysis of the transcriptome across distant species. *Nature* **512**, 445-448, doi:10.1038/nature13424 (2014).

35 Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol* **13**, R48, doi:10.1186/gb-2012-13-9-r48 (2012).

36 Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100, doi:10.1038/nature11245 (2012).

37 Khurana, E., Fu, Y., Chen, J. & Gerstein, M. Interpretation of genomic variants using a unified biological network approach. *PLoS Comput Biol* **9**, e1002886, doi:10.1371/journal.pcbi.1002886 (2013).

38 Kim, P. M., Korbel, J. O. & Gerstein, M. B. Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. *Proc Natl Acad Sci U S A* **104**, 20274-20279, doi:10.1073/pnas.0710183104 (2007).

39 Cheng, C. *et al.* An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. *Genome Biol* **16**, 63, doi:10.1186/s13059-015-0624-2 (2015).

40 Khurana, E. *et al.* Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* **342**, 1235587, doi:10.1126/science.1235587 (2013).

41 Fu, Y. *et al.* FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol* **15**, 480, doi:10.1186/s13059-014-0480-5 (2014).

42 Abyzov, A. *et al.* Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms. *Nat Commun* **6**, 7256, doi:10.1038/ncomms8256 (2015).

43 Psych, E. C. *et al.* The PsychENCODE project. *Nat Neurosci* **18**, 1707-1712, doi:10.1038/nn.4156 (2015).

44 Habegger, L. *et al.* RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. *Bioinformatics* **27**, 281-283, doi:10.1093/bioinformatics/btq643 (2011).

45 Du, J. *et al.* IQSeq: integrated isoform quantification analysis based on next-generation sequencing. *PLoS One* **7**, e29175, doi:10.1371/journal.pone.0029175 (2012).

46 Sboner, A. *et al.* FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol* **11**, R104, doi:10.1186/gb-2010-11-10-r104 (2010).

47 Rozowsky, J. *et al.* AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol* **7**, 522, doi:10.1038/msb.2011.54 (2011).

48 Chen, J. *et al.* A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun* **7**, 11101, doi:10.1038/ncomms11101 (2016).

49 Harmanci, A. & Gerstein, M. Quantification of private information leakage from phenotype-genotype data: linking attacks. *Nat Methods* **13**, 251-256, doi:10.1038/nmeth.3746 (2016).

50 Blin, K. *et al.* DoRiNA 2.0--upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res* **43**, D160-167, doi:10.1093/nar/gku1180 (2015).

51 Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92-97, doi:10.1093/nar/gkt1248 (2014).

52 Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141, doi:10.1016/j.cell.2010.03.009 (2010).

53 Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654-665, doi:10.1016/j.cell.2013.03.043 (2013).

54 Garcia, D. M. *et al.* Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol* **18**, 1139-1146, doi:10.1038/nsmb.2115 (2011).

55 Ouyang, Z., Snyder, M. P. & Chang, H. Y. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome research* **23**, 377-387, doi:10.1101/gr.138545.112 (2013).

56 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

57 Ziller, M. J. *et al.* Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature* **518**, 355-359, doi:10.1038/nature13990 (2015).

58 Leung, D. *et al.* Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518**, 350-354, doi:10.1038/nature14217 (2015).

59 Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358, doi:10.1093/bioinformatics/bts163 (2012).

60 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

61 Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321-1325, doi:10.1126/science.1098119 (2004).

62 Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome research* **15**, 901-913, doi:10.1101/gr.3577405 (2005).

63 Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology* **28**, 1045-1048, doi:10.1038/nbt1010-1045 (2010).

64 Wendl, M. C. & Wilson, R. K. Statistical aspects of discerning indel-type structural variation via DNA sequence alignment. *BMC genomics* **10**, 359, doi:10.1186/1471-2164-10-359 (2009).

65 Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American journal of human genetics* **83**, 311-321, doi:10.1016/j.ajhg.2008.06.024 (2008).

66 Cohen, J. C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-872, doi:10.1126/science.1099870 (2004).

67 Morgenthaler, S. & Thilly, W. G. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutation research* **615**, 28-56, doi:10.1016/j.mrfmmm.2006.09.003 (2007).

68 Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *American journal of human genetics* **89**, 82-93, doi:10.1016/j.ajhg.2011.05.029 (2011).

69 Lu, C. *et al.* Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun* **6**, 10086, doi:10.1038/ncomms10086 (2015).

70 Handsaker, R. E., Korn, J. M., Nemesh, J. & McCarroll, S. A. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nature genetics* **43**, 269-276, doi:10.1038/ng.768 (2011).

71 Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034, doi:10.1093/bioinformatics/btv098 (2015).

72 Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American journal of human genetics* **91**, 224-237, doi:10.1016/j.ajhg.2012.06.007 (2012).

73 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316**, 445-449, doi:10.1126/science.1138659 (2007).
74 Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539-543, doi:10.1126/science.1155174 (2008).
75 Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nature genetics* **43**, 838-846, doi:10.1038/ng.909 (2011).
76 Madsen, B. E. & Browning, S. R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS genetics* **5**, e1000384, doi:10.1371/journal.pgen.1000384 (2009).
77 Pan, W. & Shen, X. Adaptive tests for association analysis of rare variants. *Genetic epidemiology* **35**, 381-388, doi:10.1002/gepi.20586 (2011).
78 Liu, R. *et al.* Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res* **43**, e97, doi:10.1093/nar/gkv412 (2015).
79 Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *American journal of human genetics* **95**, 5-23, doi:10.1016/j.ajhg.2014.06.009 (2014).