# REFERENCES

1. Sboner A, Mu X, Greenbaum D, Auerbach RK, Gerstein MB: **The real cost of sequencing: higher than you think!** *Genome Biology* 2011:125.

2. Rodriguez LL, Brooks LD, Greenberg JH, Green ED: **The Complexities of Genomic Identifi ability**. *Science (80- )* 2013, **339**(January):275–276.

3. Sweeney L, Abu A, Winn J: **Identifying Participants in the Personal Genome Project by Name**. *SSRN Electron J* 2013:1–4.

4. **infographic-printable.pdf** [http://www.nih.gov/precisionmedicine/infographic-printable.pdf]

5. Collins FS: **A New Initiative on Precision Medicine**. *N Engl J Med* 2015, **372**:793–795.

6. **Plan for Increasing Access to Scientific Publications - NIH-Public-Access-Plan.pdf** [https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf]

7. **GENOMIC DATA SHARING (GDS) Home** [http://gds.nih.gov/index.html]

8. Sweeney L: *Uniqueness of Simple Demographics in the U.S. Population, LIDAP-WP4*. 2000.

9. Golle P: **Revisiting the uniqueness of simple demographics in the US population**. In *Proceedings of the 5th ACM workshop on Privacy in electronic society*; 2006:77–80.

10. Consortium TG: **The Genotype-Tissue Expression (GTEx) project.** *Nat Genet* 2013, **45**:580–5.

11. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57–74.

12. The 1000 Genomes Project Consortium: **An integrated map of genetic variation**. *Nature* 2012, **135**:0–9.

13. Collins FS: **The Cancer Genome Atlas ( TCGA )**. *Online* 2007:1–17.

14. Pakstis AJ, Speed WC, Fang R, Hyland FCL, Furtado MR, Kidd JR, Kidd KK: **SNPs for a universal individual identification panel**. *Hum Genet* 2010, **127**:315–324.

15. Wei YL, Li CX, Jia J, Hu L, Liu Y: **Forensic Identification Using a Multiplex Assay of 47 SNPs**. *J Forensic Sci* 2012, **57**:1448–1456.

16. Church G, Heeney C, Hawkins N, De Vries J, Boddington P, Kaye J, Bobrow M, Weir B: **Public access to genome-wide data: Five views on balancing research with privacy and protection**. *PLoS Genetics* 2009.

17. Lunshof JE, Chadwick R, Vorhaus DB, Church GM: **From genetic privacy to open consent.** *Nat Rev Genet* 2008, **9**:406–411.

18. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson J V., Stephan DA, Nelson SF, Craig DW: **Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays**. *PLoS Genet* 2008, **4**.

19. Im HK, Gamazon ER, Nicolae DL, Cox NJ: **On sharing quantitative trait GWAS results in an era of multiple-omics data and the limits of genomic privacy**. *Am J Hum Genet* 2012, **90**:591–598.

20. Narayanan A, Shmatikov V: **Robust de-anonymization of large sparse datasets**. In *Proceedings - IEEE Symposium on Security and Privacy*; 2008:111–125.

21. Lappalainen T, Sammeth M, Friedländer MR, ’t Hoen PAC, Monlong J, Rivas MA, Gonzàlez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlöf J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, MacArthur DG, Lek M, Lizano E, Buermans HPJ, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, et al.: **Transcriptome and genome sequencing uncovers functional variation in humans.** *Nature* 2013, **501**:506–11.

22. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57–63.

23. Habegger L, Sboner A, Gianoulis TA, Rozowsky J, Agarwal A, Snyder M, Gerstein M: **RSEQtools: A modular framework to analyze RNA-Seq data using compact, anonymized data summaries**. *Bioinformatics* 2011, **27**:281–283.

24. Du J, Leng J, Habegger L, Sboner A, McDermott D, Gerstein M: **IQSeq: Integrated isoform quantification analysis based on next-generation sequencing**. *PLoS One* 2012, **7**.

25. Rozowsky JS, Newburger D, Sayward F, Wu J, Jordan G, Korbel JO, Nagalakshmi U, Yang J, Zheng D, Guigó R, Gingeras TR, Weissman S, Miller P, Snyder M, Gerstein MB: **The DART classification of unannotated transcription within the ENCODE regions: Associating transcription with known and novel loci**. *Genome Res* 2007, **17**:732–745.

26. Lu ZJ, Yip KY, Wang G, Shou C, Hillier LW, Khurana E, Agarwal A, Auerbach R, Rozowsky J, Cheng C, Kato M, Miller DM, Slack F, Snyder M, Waterston RH, Reinke V, Gerstein MB: **Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data**. *Genome Res* 2011, **21**:276–285.

27. Yan K-K, Wang D, Rozowsky J, Zheng H, Cheng C, Gerstein M: **OrthoClust: an orthology-based network framework for clustering data across multiple species.** *Genome Biol* 2014, **15**:R100.

28. Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB: **PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls.** *Nat Biotechnol* 2009, **27**:66–75.

29. 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* 2014, **15**:474.

30. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung M-S, Clawson H, Contrino S, et al.: **Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project.** *Science* 2010, **330**:1775–1787.

31. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE: **Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements.** *Nature* 2007, **446**:926–929.

32. Lareau LF, Brenner SE: **Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible.** *Mol Biol Evol* 2015, **32**:1072–9.

33. Washietl S, Pedersen JS, Korbel JO, Stocsits C, Gruber AR, Hackermüller J, Hertel J, Lindemeyer M, Reiche K, Tanzer A, Ucla C, Wyss C, Antonarakis SE, Denoeud F, Lagarde J, Drenkow J, Kapranov P, Gingeras TR, Guigó R, Snyder M, Gerstein MB, Reymond A, Hofacker IL, Stadler PF: **Structured RNAs in the ENCODE selected regions of the human genome**. *Genome Res* 2007, **17**:852–864.

34. Mark B. Gerstein\*, ... AOH: **Comparative Analysis of the Transcriptome across Distant Species**. *Nature* 2014.

35. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, et al.: **The developmental transcriptome of Drosophila melanogaster.** *Nature* 2011, **471**:473–479.

36. Hansen KD, Brenner SE, Dudoit S: **Biases in Illumina transcriptome sequencing caused by random hexamer priming**. *Nucleic Acids Res* 2010, **38**.

37. Brooks AN, Aspden JL, Podgornaia AI, Rio DC, Brenner SE: **Identification and experimental validation of splicing regulatory elements in Drosophila melanogaster reveals functionally conserved splicing enhancers in metazoans**. *RNA* 2011:1884–1894.

38. Brooks AN, Duff MO, May G, Yang L, Bolisetty M, Landolin J, Wan K, Sandler J, Celniker SE, Graveley BR, Brenner SE: **Regulation of alternative splicing in Drosophila by 56 RNA binding proteins.** *Genome Res* 2015:gr.192518.115–.

39. Li JJ, Huang H, Bickel PJ, Brenner SE: **Comparison of D. melanogaster and C. elegans developmental stages, tissues, and cells by modENCODE RNA-seq data**. *Genome Res* 2014, **24**:1086–1101.

40. Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, Gardner K, Hillier LW, Janette J, Jiang L, Kasper D, Kawli T, Kheradpour P, Kundaje A, Li JJ, Ma L, Niu W, Rehm EJ, Rozowsky J, Slattery M, Spokony R, Terrell R, Vafeados D, Wang D, Weisdepp P, Wu Y-C, Xie D, Yan K-K, Feingold EA, Good PJ, et al.: **Comparative analysis of regulatory information and circuits across distant species**. *Nature* 2014, **512**:453–456.

41. Miller J a, Ding S-L, Sunkin SM, Smith K a, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, Arnold JM, Bennet C, Bertagnolli D, Brouner K, Butler S, Caldejon S, Carey A, Cuhaciyan C, Dalley R a, Dee N, Dolbeare T a, Facer B a C, Feng D, Fliss TP, Gee G, Goldy J, Gourley L, Gregor BW, Gu G, Howard RE, et al.: **Transcriptional landscape of the prenatal human brain.** *Nature* 2014, **508**:199–206.

42. Greenbaum D, Harmanci A, Gerstein M: **Proposed social and technological solutions to issues of data privacy in personal genomics**. In *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*. IEEE; 2014:1–4.

43. Greenbaum D: **Genomic Data Disclosure: Time to Reassess the Realities**. *Am J Bioeth* 2013, **13**:47–50.

44. Greenbaum D, Sboner A, Mu XJ, Gerstein M: **Genomics and privacy: Implications of the new reality of closed data for the field**. *PLoS Computational Biology* 2011.

45. Greenbaum D, Gerstein M: **The Role of Cloud Computing in Managing the Deluge of Potentially Private Genetic Data**. *The American Journal of Bioethics* 2011:39–41.

46. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M: **L -diversity**. *ACM Trans Knowl Discov Data* 2007, **1**:3–es.

47. Ninghui L, Tiancheng L, Venkatasubramanian S: **t-Closeness: Privacy beyond k-anonymity and ℓ-diversity**. In *Proceedings - International Conference on Data Engineering*; 2007:106–115.

48. Wong RC-WW, Fu AW-CC, Wang K, Pei J: **Minimality attack in privacy preserving data publishing**. In *Proceedings of the 33rd international conference on Very large data bases*; 2007:543–554.

49. Fredrikson M, Lantz E, Jha S, Lin S: **Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing**. In *23rd USENIX Security Symposium*; 2014.

50. Rozowsky J, Abyzov A, Wang J, Alves P, Raha D, Harmanci A, Leng J, Bjornson R, Kong Y, Kitabayashi N, Bhardwaj N, Rubin M, Snyder M, Gerstein M: **AlleleSeq: analysis of allele-specific expression and binding in a network framework.** *Mol Syst Biol* 2011, **7**:522.

51. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nature Methods* 2012:357–359.

52. Howie B, Marchini J, Stephens M, Chakravarti A: **Genotype Imputation with Thousands of Genomes**. *G3&#58; Genes|Genomes|Genetics* 2011:457–470.

53. O’Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I, McQuillan R, Fraser RM, Campbell H, Polasek O, Asiki G, Ekoru K, Hayward C, Wright AF, Vitart V, Navarro P, Zagury JF, Wilson JF, Toniolo D, Gasparini P, Soranzo N, Sandhu MS, Marchini J: **A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness**. *PLoS Genet* 2014, **10**.

54. Fritz MHY, Leinonen R, Cochrane G, Birney E: **Efficient storage of high throughput DNA sequencing data using reference-based compression**. *Genome Res* 2011, **21**:734–740.

55. The International HapMap 3 Consortium: **Integrating common and rare genetic variation in diverse human populations.** *Nature* 2010, **467**:52–8.

56. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET: **Transcriptome genetics using second generation sequencing in a Caucasian population.** *Nature* 2010, **464**:773–777.

57. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras J-B, Stephens M, Gilad Y, Pritchard JK: **Understanding mechanisms underlying human gene expression variation with RNA sequencing.** *Nature* 2010, **464**:768–772.

58. Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis NI, Yurovsky A, Lappalainen T, Romano-Palumbo L, Planchon A, Bielser D, Bryois J, Padioleau I, Udin G, Thurnheer S, Hacker D, Core LJ, Lis JT, Hernandez N, Reymond A, Deplancke B, Dermitzakis ET: **Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription.** *Science* 2013, **342**:744–7.

59. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, Boyle AP, Zhang QC, Zakharia F, Spacek D V, Li J, Xie D, Olarerin-George A, Steinmetz LM, Hogenesch JB, Kellis M, Batzoglou S, Snyder M: **Extensive variation in chromatin states across humans.** *Science (New York, NY)* 2013:750–752.

60. Toung JM, Morley M, Li M, Cheung VG: **RNA-sequence analysis of human B-cells**. *Genome Res* 2011, **21**:991–998.

61. Wang IX, Core LJ, Kwak H, Brady L, Bruzel A, McDaniel L, Richards AL, Wu M, Grunseich C, Lis JT, Cheung VG: **RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II**. *Cell Rep* 2014, **6**:906–915.

62. Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG: **Widespread RNA and DNA sequence differences in the human transcriptome.** *Science* 2011, **333**:53–58.

63. Abyzov A, Urban AE, Snyder M, Gerstein M: **CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing**. *Genome Res* 2011, **21**:974–984.