We have analyzed the common ENTEx samples to validate and characterize our annotation. Specifically, in ongoing unpublished work, the ENTEx RNA-seq data sets were used to assess pseudogene expression in various human tissues. We also investigated personalized molecular profiling using MS data from ENtex samples and mapping peptides directly to exons rather than transcripts to reduce ambiguity when looking at protein isoforms and enable validation of individual exons and their differential expression.

Our pseudogene annotation pipeline11, including PseudoPipe27, 28, RCPedia29, and Retrofinder30, focused on refining the annotations of pseudogenes. We identified 271 mouse and 431 human putative unitary pseudogenes for manual review using a combination of multi-sequence alignments, manual curation, and a specialised workflow, including a new pipeline leveraging Ensembl protein-coding annotation of eukaryotic species. In addition, we created a population-based pseudogene annotation of 16 closely related mouse strains and used it to construct a pangenome pseudogene data set of 49,262 unique entries across strains31. We defined orthology relationship patterns of pseudogene gain and loss across the strains as a prototype for work annotating human pseudogenes leveraging variation across the human population31. We also extended our methods to quantify the transcription and activity of pseudogenes in human and mouse using RNA-seq data by minimizing the mis-mapping of RNA-seq reads caused by the sequence similarity between pseudogenes and parent genes, highlighting tissue and strain-specific transcribed pseudogenes31, 32.

### References

1. Stark Z, Dolman L, Manolio TA et al. Integrating Genomics into Healthcare: A Global Responsibility. *Am J Hum Genet*. 2019; 104(1) 13-20. doi:10.1016/j.ajhg.2018.11.014. PubMed PMID: 30609404. PMCID: PMC6323624.

2. Djebali S, Kapranov P, Foissac S et al. Efficient targeted transcript discovery via array-based normalization of RACE libraries. *Nat Methods*. 2008; 5(7) 629-635. doi:10.1038/nmeth.1216. PubMed PMID: 18500348. PMCID: PMC2713501.

3. Lagarde J, Uszczynska-Ratajczak B, Santoyo-Lopez J et al. Extension of human lncRNA transcripts by RACE coupled with long-read high-throughput sequencing (RACE-Seq). *Nat Commun*. 2016; 7 12339. doi:10.1038/ncomms12339. PubMed PMID: 27531712. PMCID: PMC4992054.

4. Howald C, Tanzer A, Chrast J et al. Combining RT-PCR-seq and RNA-seq to catalog all genic elements encoded in the human genome. *Genome Res*. 2012; 22(9) 1698-1710. doi:10.1101/gr.134478.111. PubMed PMID: 22955982. PMCID: PMC3431487.

5. Lagarde J, Uszczynska-Ratajczak B, Carbonell S et al. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet*. 2017; 49(12) 1731-1740. doi:10.1038/ng.3988. PubMed PMID: 29106417. PMCID: PMC5709232.

6. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Mouse GDG. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. 2019; 47(D1) D801-D806. doi:10.1093/nar/gky1056. PubMed PMID: 30407599. PMCID: PMC6323923.

7. Lilue J, Doran AG, Fiddes IT et al. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nat Genet*. 2018; 50(11) 1574-1583. doi:10.1038/s41588-018-0223-8. PubMed PMID: 30275530. PMCID: PMC6205630.

8. Harrow J, Denoeud F, Frankish A et al. GENCODE: producing a reference annotation for ENCODE. *Genome Biol*. 2006; 7 Suppl 1 S4.1-9. doi:10.1186/gb-2006-7-s1-s4. PubMed PMID: 16925838. PMCID: PMC1810553.

9. Zheng D, Frankish A, Baertsch R et al. Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res*. 2007; 17(6) 839-851. doi:10.1101/gr.5586307. PubMed PMID: 17568002. PMCID: PMC1891343.

10. Harrow J, Frankish A, Gonzalez JM et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012; 22(9) 1760-1774. doi:10.1101/gr.135350.111. PubMed PMID: 22955987. PMCID: PMC3431492.

11. Pei B, Sisu C, Frankish A et al. The GENCODE pseudogene resource. *Genome Biol*. 2012; 13(9) R51. doi:10.1186/gb-2012-13-9-r51. PubMed PMID: 22951037. PMCID: PMC3491395.

12. Derrien T, Johnson R, Bussotti G et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012; 22(9) 1775-1789. doi:10.1101/gr.132159.111. PubMed PMID: 22955988. PMCID: PMC3431493.

13. Mudge JM, Harrow J. Creating reference gene annotation for the mouse C57BL6/J genome assembly. *Mamm Genome*. 2015; 26(9-10) 366-378. doi:10.1007/s00335-015-9583-x. PubMed PMID: 26187010. PMCID: PMC4602055.

14. Frankish A, Diekhans M, Ferreira AM et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019; 47(D1) D766-D773. doi:10.1093/nar/gky955. PubMed PMID: 30357393. PMCID: PMC6323946.

15. Spooner W, McLaren W, Slidel T et al. Haplosaurus computes protein haplotypes for use in precision drug design. *Nat Commun*. 2018; 9(1) 4128. doi:10.1038/s41467-018-06542-1. PubMed PMID: 30297836. PMCID: PMC6175845.

16. Pujar S, O’Leary NA, Farrell CM et al. Consensus coding sequence (CCDS) database: a standardized set of human and mouse protein-coding regions supported by expert curation. *Nucleic Acids Res*. 2018; doi:10.1093/nar/gkx1031. PubMed PMID: 29126148. PMCID: PMC5753299.

17. Mudge JM, Jungreis I, Hunt T et al. Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. *Genome Res*. 2019; 29(12) 2073-2087. doi:10.1101/gr.246462.118. PubMed PMID: 31537640. PMCID: PMC6886504.

18. Barnes IHA, Ibarra-Soria X, Fitzgerald S et al. Expert curation of the human and mouse olfactory receptor gene repertoires identifies conserved coding regions split across two exons. *BMC Genomics*. 2020; 21(1) 196. doi:10.1186/s12864-020-6583-3. PubMed PMID: 32126975. PMCID: PMC7055050.

19. Steward CA, Roovers J, Suner MM et al. Re-annotation of 191 developmental and epileptic encephalopathy-associated genes unmasks de novo variants in SCN1A. *NPJ Genom Med*. 2019; 4 31. doi:10.1038/s41525-019-0106-7. PubMed PMID: 31814998. PMCID: PMC6889285.

20. Pervouchine D, Popov Y, Berry A, Borsari B, Frankish A, Guigó R. Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. *Nucleic Acids Res*. 2019; 47(10) 5293-5306. doi:10.1093/nar/gkz193. PubMed PMID: 30916337. PMCID: PMC6547761.

21. Pertea M, Shumate A, Pertea G et al. CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol*. 2018; 19(1) 208. doi:10.1186/s13059-018-1590-2. PubMed PMID: 30486838. PMCID: PMC6260756.

22. Jungreis I, Tress ML, Mudge J et al. Nearly all new protein-coding predictions in the CHESS database are not protein-coding. *bioRxiv*. 2018; 360602.

23. van Heesch S, Witte F, Schneider-Lunitz V et al. The Translational Landscape of the Human Heart. *Cell*. 2019; 178(1) 242-260.e29. doi:10.1016/j.cell.2019.05.010. PubMed PMID: 31155234.

24. Irimia M, Weatheritt RJ, Ellis JD et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell*. 2014; 159(7) 1511-1523. doi:10.1016/j.cell.2014.11.035. PubMed PMID: 25525873. PMCID: PMC4390143.

25. Petryszak R, Fonseca NA, Füllgrabe A et al. The RNASeq-er API-a gateway to systematically updated analysis of public RNA-seq data. *Bioinformatics*. 2017; 33(14) 2218-2220. doi:10.1093/bioinformatics/btx143. PubMed PMID: 28369191. PMCID: PMC5870697.

26. Fiddes IT, Armstrong J, Diekhans M et al. Comparative Annotation Toolkit (CAT)-simultaneous clade and personal genome annotation. *Genome Res*. 2018; 28(7) 1029-1038. doi:10.1101/gr.233460.117. PubMed PMID: 29884752. PMCID: PMC6028123.

27. Zhang Z, Carriero N, Zheng D, Karro J, Harrison PM, Gerstein M. PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics*. 2006; 22(12) 1437-1439. doi:10.1093/bioinformatics/btl116. PubMed PMID: 16574694. PMCID: Open access at doi://10.1093/bioinformatics/btl116.

28. Zheng D, Gerstein MB. A computational approach for identifying pseudogenes in the ENCODE regions. *Genome Biol*. 2006; 7 Suppl 1 S13.1-10. doi:10.1186/gb-2006-7-s1-s13. PubMed PMID: 16925835. PMCID: PMC1810550.

29. Navarro FCP, Galante PAF. RCPedia: a database of retrocopied genes. *Bioinformatics*. 2013; 29(9) 1235-1237. doi:10.1093/bioinformatics/btt104. PubMed PMID: 23457042. PMCID: PMC3634192.

30. Baertsch R, Diekhans M, Kent WJ, Haussler D, Brosius J. Retrocopy contributions to the evolution of the human genome. *BMC Genomics*. 2008; 9 466. doi:10.1186/1471-2164-9-466. PubMed PMID: 18842134. PMCID: PMC2584115.

31. Sisu C, Muir P, Frankish A et al. Pseudogenes in the mouse lineage: transcriptional activity and strain-specific history. *bioRxiv*. 2018; 386656.

32. Navarro FC, Hoops J, Bellfy L et al. TeXP: Deconvolving the effects of pervasive and autonomous transcription of transposable elements. *PLoS Comput Biol*. 2019; 15(8) e1007293. doi:10.1371/journal.pcbi.1007293. PubMed PMID: 31425522. PMCID: PMC6715295.

33. Khan YA, Jungreis I, Wright JC et al. Evidence for a novel overlapping coding sequence in POLG initiated at a CUG start codon. *BMC Genet*. 2020; 21(1) 25. doi:10.1186/s12863-020-0828-7. PubMed PMID: 32138667. PMCID: PMC7059407.

34. Jungreis I, Chan CS, Waterhouse RM, Fields G, Lin MF, Kellis M. Evolutionary Dynamics of Abundant Stop Codon Readthrough. *Mol Biol Evol*. 2016; 33(12) 3108-3132. doi:10.1093/molbev/msw189. PubMed PMID: 27604222. PMCID: PMC5100048.

35. Loughran G, Jungreis I, Tzani I et al. Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response. *J Biol Chem*. 2018; 293(12) 4434-4444. doi:10.1074/jbc.M117.818526. PubMed PMID: 29386352. PMCID: PMC5868278.

36. Rodriguez JM, Rodriguez-Rivas J, Di Domenico T, Vázquez J, Valencia A, Tress ML. APPRIS 2017: principal isoforms for multiple gene sets. *Nucleic Acids Res*. 2018; 46(D1) D213-D217. doi:10.1093/nar/gkx997. PubMed PMID: 29069475. PMCID: PMC5753224.

37. Abascal F, Juan D, Jungreis I et al. Loose ends: almost one in five human genes still have unresolved coding status. *Nucleic Acids Res*. 2018; 46(14) 7070-7084. doi:10.1093/nar/gky587. PubMed PMID: 29982784. PMCID: PMC6101605.

38. Dalgleish R, Flicek P, Cunningham F et al. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. *Genome Med*. 2010; 2(4) 24. doi:10.1186/gm145. PubMed PMID: 20398331. PMCID: PMC3965024.

39. MacArthur JAL, Morales J, Tully RE et al. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. *Nucleic Acids Res*. 2014; 42(Database issue) D873-8. doi:10.1093/nar/gkt1198. PubMed PMID: 24285302. PMCID: PMC3965024.

40. The RNAcentral Consortium. RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res*. 2019; 47(D1) D221-D229. doi:10.1093/nar/gky1034. PubMed PMID: 30395267. PMCID: PMC6324050.

41. Aken BL, Achuthan P, Akanni W et al. Ensembl 2017. *Nucleic Acids Res*. 2017; 45(D1) D635-D642. doi:10.1093/nar/gkw1104. PubMed PMID: 27899575. PMCID: PMC5210575.

42. Cunningham F, Achuthan P, Akanni W et al. Ensembl 2019. *Nucleic Acids Res*. 2019; 47(D1) D745-D751. doi:10.1093/nar/gky1113. PubMed PMID: 30407521. PMCID: PMC6323964.

43. Yates AD, Achuthan P, Akanni W et al. Ensembl 2020. *Nucleic Acids Res*. 2020; 48(D1) D682-D688. doi:10.1093/nar/gkz966. PubMed PMID: 31691826.

44. Zerbino DR, Achuthan P, Akanni W et al. Ensembl 2018. *Nucleic Acids Res*. 2018; 46(D1) D754-D761. doi:10.1093/nar/gkx1098. PubMed PMID: 29155950. PMCID: PMC5753206.

45. Casper J, Zweig AS, Villarreal C et al. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res*. 2018; 46(D1) D762-D769. doi:10.1093/nar/gkx1020. PubMed PMID: 29106570. PMCID: PMC5753355.

46. Haeussler M, Zweig AS, Tyner C et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019; 47(D1) D853-D858. doi:10.1093/nar/gky1095. PubMed PMID: 30407534. PMCID: PMC6323953.

47. Lee CM, Barber GP, Casper J et al. UCSC Genome Browser enters 20th year. *Nucleic Acids Res*. 2020; 48(D1) D756-D761. doi:10.1093/nar/gkz1012. PubMed PMID: 31691824. PMCID: PMC7145642.

48. Gordon DE, Jang GM, Bouhaddou M et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020; doi:10.1038/s41586-020-2286-9. PubMed PMID: 32353859.

49. Martin AR, Williams E, Foulger RE et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet*. 2019; 51(11) 1560-1565. doi:10.1038/s41588-019-0528-2. PubMed PMID: 31676867.

50. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011; 27(13) i275-82. doi:10.1093/bioinformatics/btr209. PubMed PMID: 21685081. PMCID: PMC3117341.

51. She X, Horvath JE, Jiang Z et al. The structure and evolution of centromeric transition regions within the human genome. *Nature*. 2004; 430(7002) 857-864. doi:10.1038/nature02806. PubMed PMID: 15318213.

52. Genovese G, Handsaker RE, Li H et al. Using population admixture to help complete maps of the human genome. *Nat Genet*. 2013; 45(4) 406-14, 414e1. doi:10.1038/ng.2565. PubMed PMID: 23435088. PMCID: PMC3683849.

53. Wright JC, Mudge J, Weisser H et al. Improving GENCODE reference gene annotation using a high-stringency proteogenomics workflow. *Nat Commun*. 2016; 7 11778. doi:10.1038/ncomms11778. PubMed PMID: 27250503. PMCID: PMC4895710.

54. Avsec Ž, Kreuzhuber R, Israeli J et al. The Kipoi repository accelerates community exchange and reuse of predictive models for genomics. *Nat Biotechnol*. 2019; 37(6) 592-600. doi:10.1038/s41587-019-0140-0. PubMed PMID: 31138913. PMCID: PMC6777348.

55. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004; 11(2-3) 377-394. doi:10.1089/1066527041410418. PubMed PMID: 15285897.

56. Rosenberg AB, Patwardhan RP, Shendure J, Seelig G. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell*. 2015; 163(3) 698-711. doi:10.1016/j.cell.2015.09.054. PubMed PMID: 26496609.

57. Paggi JM, Bejerano G. A sequence-based, deep learning model accurately predicts RNA splicing branchpoints. *RNA*. 2018; 24(12) 1647-1658. doi:10.1261/rna.066290.118. PubMed PMID: 30224349. PMCID: PMC6239175.

58. Xiong HY, Alipanahi B, Lee LJ et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*. 2015; 347(6218) 1254806. doi:10.1126/science.1254806. PubMed PMID: 25525159. PMCID: PMC4362528.

59. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res*. 2014; 42(22) 13534-13544. doi:10.1093/nar/gku1206. PubMed PMID: 25416802. PMCID: PMC4267638.

60. Cheng J, Nguyen TYD, Cygan KJ et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol*. 2019; 20(1) 48. doi:10.1186/s13059-019-1653-z. PubMed PMID: 30823901. PMCID: PMC6396468.

61. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019; 176(3) 535-548.e24. doi:10.1016/j.cell.2018.12.015. PubMed PMID: 30661751.

62. Thomas JD, Polaski JT, Feng Q et al. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nat Genet*. 2020; 52(1) 84-94. doi:10.1038/s41588-019-0555-z. PubMed PMID: 31911676. PMCID: PMC6962552.

63. Edwards CR, Ritchie W, Wong JJ et al. A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. *Blood*. 2016; 127(17) e24-e34. doi:10.1182/blood-2016-01-692764. PubMed PMID: 26962124. PMCID: PMC4850870.

64. Monteuuis G, Wong JJL, Bailey CG, Schmitz U, Rasko JEJ. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res*. 2019; 47(22) 11497-11513. doi:10.1093/nar/gkz1068. PubMed PMID: 31724706. PMCID: PMC7145568.

65. Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol*. 2019; 20(7) 406-420. doi:10.1038/s41580-019-0126-2. PubMed PMID: 30992545. PMCID: PMC6855384.

66. Jensen TH, Jacquier A, Libri D. Dealing with pervasive transcription. *Mol Cell*. 2013; 52(4) 473-484. doi:10.1016/j.molcel.2013.10.032. PubMed PMID: 24267449.

67. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*. 2017; 18(1) 18-30. doi:10.1038/nrm.2016.116. PubMed PMID: 27677860. PMCID: PMC5483950.

68. Gruber AJ, Zavolan M. Alternative cleavage and polyadenylation in health and disease. *Nat Rev Genet*. 2019; 20(10) 599-614. doi:10.1038/s41576-019-0145-z. PubMed PMID: 31267064.

69. Chen J, Brunner AD, Cogan JZ et al. Pervasive functional translation of noncanonical human open reading frames. *Science*. 2020; 367(6482) 1140-1146. doi:10.1126/science.aay0262. PubMed PMID: 32139545.

70. Schulz J, Mah N, Neuenschwander M et al. Loss-of-function uORF mutations in human malignancies. *Sci Rep*. 2018; 8(1) 2395. doi:10.1038/s41598-018-19201-8. PubMed PMID: 29402903. PMCID: PMC5799362.

71. Wen Y, Liu Y, Xu Y et al. Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. *Nat Genet*. 2009; 41(2) 228-233. doi:10.1038/ng.276. PubMed PMID: 19122663.

72. Ma J, Ward CC, Jungreis I et al. Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J Proteome Res*. 2014; 13(3) 1757-1765. doi:10.1021/pr401280w. PubMed PMID: 24490786. PMCID: PMC3993966.

73. Ma J, Diedrich JK, Jungreis I et al. Improved Identification and Analysis of Small Open Reading Frame Encoded Polypeptides. *Anal Chem*. 2016; 88(7) 3967-3975. doi:10.1021/acs.analchem.6b00191. PubMed PMID: 27010111. PMCID: PMC4939623.

74. Sealfon RS, Lin MF, Jungreis I, Wolf MY, Kellis M, Sabeti PC. FRESCo: finding regions of excess synonymous constraint in diverse viruses. *Genome Biol*. 2015; 16 38. doi:10.1186/s13059-015-0603-7. PubMed PMID: 25853568. PMCID: PMC4376164.

75. Lindblad-Toh K, Garber M, Zuk O et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011; 478(7370) 476-482. doi:10.1038/nature10530. PubMed PMID: 21993624. PMCID: PMC3207357.

76. Rajput B, Pruitt KD, Murphy TD. RefSeq curation and annotation of stop codon recoding in vertebrates. *Nucleic Acids Res*. 2019; 47(2) 594-606. doi:10.1093/nar/gky1234. PubMed PMID: 30535227. PMCID: PMC6344875.

77. Alioto TS. U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res*. 2007; 35(Database issue) D110-5. doi:10.1093/nar/gkl796. PubMed PMID: 17082203. PMCID: PMC1635337.

78. Mariotti M, Guigó R. Selenoprofiles: profile-based scanning of eukaryotic genome sequences for selenoprotein genes. *Bioinformatics*. 2010; 26(21) 2656-2663. doi:10.1093/bioinformatics/btq516. PubMed PMID: 20861026. PMCID: PMC2958745.

79. Jungreis I, Lin MF, Spokony R et al. Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Res*. 2011; 21(12) 2096-2113. doi:10.1101/gr.119974.110. PubMed PMID: 21994247. PMCID: PMC3227100.

80. Loughran G, Chou MY, Ivanov IP et al. Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res*. 2014; 42(14) 8928-8938. doi:10.1093/nar/gku608. PubMed PMID: 25013167. PMCID: PMC4132726.

81. Gladyshev VN, Arnér ES, Berry MJ et al. Selenoprotein Gene Nomenclature. *J Biol Chem*. 2016; 291(46) 24036-24040. doi:10.1074/jbc.M116.756155. PubMed PMID: 27645994. PMCID: PMC5104929.

82. Holford ME, Khurana E, Cheung K-H, Gerstein M. Using semantic web rules to reason on an ontology of pseudogenes. *Bioinformatics*. 2010; 26(12) i71-8. doi:10.1093/bioinformatics/btq173. PubMed PMID: 20529940. PMCID: PMC2881358.

83. Habegger L, Balasubramanian S, Chen DZ et al. VAT: a computational framework to functionally annotate variants in personal genomes within a cloud-computing environment. *Bioinformatics*. 2012; 28(17) 2267-2269. doi:10.1093/bioinformatics/bts368. PubMed PMID: 22743228. PMCID: PMC3426844.

84. Balasubramanian S, Fu Y, Pawashe M et al. Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nat Commun*. 2017; 8(1) 382. doi:10.1038/s41467-017-00443-5. PubMed PMID: 28851873. PMCID: PMC5575292.

85. Imada EL, Sanchez DF, Collado-Torres L et al. Recounting the FANTOM CAGEAssociated Transcriptome. *Genome Res*. 2020; doi:10.1101/gr.254656.119. PubMed PMID: 32079618.

86. Uszczynska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet*. 2018; 19(9) 535-548. doi:10.1038/s41576-018-0017-y. PubMed PMID: 29795125. PMCID: PMC6451964.

87. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet*. 2020; 21(2) 102-117. doi:10.1038/s41576-019-0184-5. PubMed PMID: 31729473.

88. Rheinbay E, Nielsen MM, Abascal F et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*. 2020; 578(7793) 102-111. doi:10.1038/s41586-020-1965-x. PubMed PMID: 32025015. PMCID: PMC7054214.

89. Karczewski KJ, Francioli LC, Tiao G et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv*. 2020; 531210. doi:10.1101/531210.

90. Ma L, Cao J, Liu L et al. LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res*. 2019; 47(D1) D128-D134. doi:10.1093/nar/gky960. PubMed PMID: 30329098. PMCID: PMC6323930.

91. Vancura A, Lanzós A, Bosch N, Torres M, Häfliger S, Johnson R. Semi-automated annotation of known and novel cancer long noncoding RNAs with the Cancer LncRNA Census 2 (CLC2). *bioRxiv*. 2020; 2020.04.28.066225. doi:10.1101/2020.04.28.066225.

92. Noviello TMR, Di Liddo A, Ventola GM et al. Detection of long non-coding RNA homology, a comparative study on alignment and alignment-free metrics. *BMC Bioinformatics*. 2018; 19(1) 407. doi:10.1186/s12859-018-2441-6. PubMed PMID: 30400819. PMCID: PMC6220562.

93. Siepel A, Bejerano G, Pedersen JS et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005; 15(8) 1034-1050. doi:10.1101/gr.3715005. PubMed PMID: 16024819. PMCID: PMC1182216.

94. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010; 20(1) 110-121. doi:10.1101/gr.097857.109. PubMed PMID: 19858363.

95. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2019; 47(D1) D155-D162. doi:10.1093/nar/gky1141. PubMed PMID: 30423142. PMCID: PMC6323917.

96. Djebali S, Davis CA, Merkel A et al. Landscape of transcription in human cells. *Nature*. 2012; 489(7414) 101-108. doi:10.1038/nature11233. PubMed PMID: 22955620. PMCID: PMC3684276.

97. Breschi A, Muñoz-Aguirre M, Wucher V et al. A limited set of transcriptional programs define major cell types. *bioRxiv*. 2020; 857169. doi:10.1101/857169.

98. Rahmanian S, Murad R, Breschi A et al. Dynamics of microRNA expression during mouse prenatal development. *Genome Res*. 2019; 29(11) 1900-1909. doi:10.1101/gr.248997.119. PubMed PMID: 31645363. PMCID: PMC6836743.

99. Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol*. 2011; 12(2) R16. doi:10.1186/gb-2011-12-2-r16. PubMed PMID: 21324177. PMCID: PMC3188798.

100. Kapranov P, St Laurent G. Dark Matter RNA: Existence, Function, and Controversy. *Front Genet*. 2012; 3 60. doi:10.3389/fgene.2012.00060. PubMed PMID: 22536205. PMCID: PMC3332219.

101. Dhir A, Dhir S, Proudfoot NJ, Jopling CL. Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat Struct Mol Biol*. 2015; 22(4) 319-327. doi:10.1038/nsmb.2982. PubMed PMID: 25730776. PMCID: PMC4492989.

102. Lorenzi L, Chiu H-S, Cobos FA et al. The RNA Atlas, a single nucleotide resolution map of the human transcriptome. *bioRxiv*. 2019; 807529. doi:10.1101/807529.

103. Tennyson CN, Klamut HJ, Worton RG. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat Genet*. 1995; 9(2) 184-190. doi:10.1038/ng0295-184. PubMed PMID: 7719347.

104. Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res*. 2009; 19(3) 347-359. doi:10.1101/gr.087775.108. PubMed PMID: 19106332. PMCID: PMC2661813.

105. Houseley J, Tollervey D. Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *PLoS One*. 2010; 5(8) e12271. doi:10.1371/journal.pone.0012271. PubMed PMID: 20805885. PMCID: PMC2923612.

106. Rosner A, Maslenin L, Spiegel S. The use of short and long PCR products for improved detection of prunus necrotic ringspot virus in woody plants. *J Virol Methods*. 1997; 67(2) 135-141. doi:10.1016/s0166-0934(97)00088-8. PubMed PMID: 9300378.

107. Lorenz TC. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J Vis Exp*. 2012; 63) e3998. doi:10.3791/3998. PubMed PMID: 22664923. PMCID: PMC4846334.

108. Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol*. 2013; 31(11) 1009-1014. doi:10.1038/nbt.2705. PubMed PMID: 24108091. PMCID: PMC4075632.

109. St Laurent G, Shtokalo D, Dong B et al. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol*. 2013; 14(7) R73. doi:10.1186/gb-2013-14-7-r73. PubMed PMID: 23876380. PMCID: PMC4053963.

110. St Laurent G, Vyatkin Y, Antonets D et al. Functional annotation of the vlinc class of non-coding RNAs using systems biology approach. *Nucleic Acids Res*. 2016; 44(7) 3233-3252. doi:10.1093/nar/gkw162. PubMed PMID: 27001520. PMCID: PMC4838384.

111. Uemura M, Zheng Q, Koh CM, Nelson WG, Yegnasubramanian S, De Marzo AM. Overexpression of ribosomal RNA in prostate cancer is common but not linked to rDNA promoter hypomethylation. *Oncogene*. 2012; 31(10) 1254-1263. doi:10.1038/onc.2011.319. PubMed PMID: 21822302. PMCID: PMC3298623.

112. Bywater MJ, Pearson RB, McArthur GA, Hannan RD. Dysregulation of the basal RNA polymerase transcription apparatus in cancer. *Nat Rev Cancer*. 2013; 13(5) 299-314. doi:10.1038/nrc3496. PubMed PMID: 23612459.

113. Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT. Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol*. 1988; 203(2) 439-455. doi:10.1016/0022-2836(88)90011-3. PubMed PMID: 3199442.

114. Agrawal S, Ganley ARD. The conservation landscape of the human ribosomal RNA gene repeats. *PLoS One*. 2018; 13(12) e0207531. doi:10.1371/journal.pone.0207531. PubMed PMID: 30517151. PMCID: PMC6281188.

115. Gerstein MB, Bruce C, Rozowsky JS et al. What is a gene, post-ENCODE? History and updated definition. *Genome Res*. 2007; 17(6) 669-681. doi:10.1101/gr.6339607. PubMed PMID: 17567988.

116. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489(7414) 57-74. doi:10.1038/nature11247. PubMed PMID: 22955616. PMCID: PMC3439153.

117. Gerstein MB, Kundaje A, Hariharan M et al. Architecture of the human regulatory network derived from ENCODE data. *Nature*. 2012; 489(7414) 91-100. doi:10.1038/nature11245. PubMed PMID: 22955619. PMCID: PMC4154057.

118. Gerstein MB, Lu ZJ, Van Nostrand EL et al. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. *Science*. 2010; 330(6012) 1775-1787. doi:10.1126/science.1196914. PubMed PMID: 21177976.

119. Cheng C, Yan K-K, Yip KY et al. A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. *Genome Biol*. 2011; 12(2) R15. doi:10.1186/gb-2011-12-2-r15. PubMed PMID: 21324173.

120. Schmidt D, Wilson MD, Ballester B et al. Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. *Science*. 2010; 328(5981) 1036-1040. doi:10.1126/science.1186176. PubMed PMID: 20378774. PMCID: PMC3008766.

121. Villar D, Berthelot C, Aldridge S et al. Enhancer Evolution across 20 Mammalian Species. *Cell*. 2015; 160(3) 554-566. doi:10.1016/j.cell.2015.01.006. PubMed PMID: 25635462. PMCID: PMC4313353.

122. Wong ES, Thybert D, Schmitt BM, Stefflova K, Odom DT, Flicek P. Decoupling of evolutionary changes in transcription factor binding and gene expression in mammals. *Genome Res*. 2015; 25(2) 167-178. doi:10.1101/gr.177840.114. PubMed PMID: 25394363. PMCID: PMC4315291.

123. Claussnitzer M, Dankel SN, Kim K-H et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med*. 2015; 373(10) 895-907. doi:10.1056/NEJMoa1502214. PubMed PMID: 26287746.

124. Avsec Ž, Weilert M, Shrikumar A et al. Deep learning at base-resolution reveals cis-regulatory motif syntax. *bioRxiv*. 2020; 737981. doi:10.1101/737981.

125. Avanti S, Peyton G, Anshul K. Learning Important Features Through Propagating Activation Differences Proceedings of the 34th International Conference on Machine Learning. *Proceedings of Machine Learning Research*. 2017; 70 3145-3153.

126. Khurana E, Fu Y, Colonna V et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science*. 2013; 342(6154) 1235587. doi:10.1126/science.1235587. PubMed PMID: 24092746. PMCID: PMC3947637.

127. Fu Y, Liu Z, Lou S et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol*. 2014; 15(10) 480. doi:10.1186/s13059-014-0480-5. PubMed PMID: 25273974. PMCID: PMC4203974.

128. Ritchie GRS, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat Methods*. 2014; 11(3) 294-296. doi:10.1038/nmeth.2832. PubMed PMID: 24487584. PMCID: PMC5015703.

129. Lou S, Cotter KA, Li T et al. GRAM: A GeneRAlized Model to predict the molecular effect of a non-coding variant in a cell-type specific manner. *PLoS Genet*. 2019; 15(8) e1007860. doi:10.1371/journal.pgen.1007860. PubMed PMID: 31469829. PMCID: PMC6742416.

130. Sisu C, Pei B, Leng J et al. Comparative analysis of pseudogenes across three phyla. *Proc Natl Acad Sci U S A*. 2014; 111(37) 13361-13366. doi:10.1073/pnas.1407293111. PubMed PMID: 25157146. PMCID: PMC4169933.

131. Moore JE, Pratt HE, Purcaro MJ, Weng Z. A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. *Genome Biol*. 2020; 21(1) 17. doi:10.1186/s13059-019-1924-8. PubMed PMID: 31969180. PMCID: PMC6977301.

132. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*. 2017; 550(7675) 204-213. doi:10.1038/nature24277. PubMed PMID: 29022597. PMCID: PMC5776756.

133. Kerimov N, Hayhurst JD, Manning JR et al. eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. *bioRxiv*. 2020; 2020.01.29.924266. doi:10.1101/2020.01.29.924266.

134. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012; 489(7416) 391-399. doi:10.1038/nature11405. PubMed PMID: 22996553. PMCID: PMC4243026.

135. Yao P, Lin P, Gokoolparsadh A, Assareh A, Thang MW, Voineagu I. Coexpression networks identify brain region-specific enhancer RNAs in the human brain. *Nat Neurosci*. 2015; 18(8) 1168-1174. doi:10.1038/nn.4063. PubMed PMID: 26167905.

136. Kundaje A, Meuleman W, Ernst J et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518(7539) 317-330. doi:10.1038/nature14248. PubMed PMID: 25693563. PMCID: PMC4530010.

137. Adsera CB, Park YP, Meuleman W, Kellis M. Integrative analysis of 10,000 epigenomic maps across 800 samples for regulatory genomics and disease dissection. *bioRxiv*. 2019; 810291. doi:10.1101/810291.

138. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The Ensembl Regulatory Build. *Genome Biology*. 2015; 16(1) 56. doi:10.1186/s13059-015-0621-5. PubMed PMID: 25887522. PMCID: PMC4407537.

139. Consortium ENCODEP. A User’s Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol*. 2011; 9(4) e1001046. doi:10.1371/journal.pbio.1001046. PubMed PMID: 21526222. PMCID: PMC3079585.

140. Wang D, Liu S, Warrell J et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018; 362(6420) doi:10.1126/science.aat8464. PubMed PMID: 30545857. PMCID: PMC6413328.

141. Zhang J, Lee D, Dhiman V et al. An integrative ENCODE resource for cancer genomics. *bioRxiv*. 2019; 706424. doi:10.1101/706424.

142. Lee D, Shi M, Moran J et al. STARRPeaker: Uniform processing and accurate identification of STARR-seq active regions. *bioRxiv*. 2019; 694869. doi:10.1101/694869.

143. Sethi A, Gu M, Gumusgoz E et al. A cross-organism framework for supervised enhancer prediction with epigenetic pattern recognition and targeted validation. *bioRxiv*. 2018; 385237. doi:10.1101/385237.

144. Greenside P, Shimko T, Fordyce P, Kundaje A. Discovering epistatic feature interactions from neural network models of regulatory DNA sequences. *Bioinformatics*. 2018; 34(17) i629-i637. doi:10.1093/bioinformatics/bty575. PubMed PMID: 30423062. PMCID: PMC6129272.

145. Sherman RM, Forman J, Antonescu V et al. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat Genet*. 2019; 51(1) 30-35. doi:10.1038/s41588-018-0273-y. PubMed PMID: 30455414. PMCID: PMC6309586.

146. Seo JS, Rhie A, Kim J et al. De novo assembly and phasing of a Korean human genome. *Nature*. 2016; 538(7624) 243-247. doi:10.1038/nature20098. PubMed PMID: 27706134.

147. Mallick S, Li H, Lipson M et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. 2016; 538(7624) 201-206. doi:10.1038/nature18964. PubMed PMID: 27654912. PMCID: PMC5161557.

148. Shi L, Guo Y, Dong C et al. Long-read sequencing and de novo assembly of a Chinese genome. *Nat Commun*. 2016; 7 12065. doi:10.1038/ncomms12065. PubMed PMID: 27356984. PMCID: PMC4931320.

149. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015; 33(3) 290-295. doi:10.1038/nbt.3122. PubMed PMID: 25690850. PMCID: PMC4643835.

150. Kang YJ, Yang DC, Kong L et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res*. 2017; 45(W1) W12-W16. doi:10.1093/nar/gkx428. PubMed PMID: 28521017. PMCID: PMC5793834.

151. Morales J, Welter D, Bowler EH et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol*. 2018; 19(1) 489. doi:10.1186/s13059-018-1396-2. PMCID: PMC5815218.

152. Hooker SE, Woods-Burnham L, Bathina M et al. Genetic Ancestry Analysis Reveals Misclassification of Commonly Used Cancer Cell Lines. *Cancer Epidemiol Biomarkers Prev*. 2019; 28(6) 1003-1009. doi:10.1158/1055-9965.EPI-18-1132. PubMed PMID: 30787054. PMCID: PMC6548687.

153. Jiang S, Cheng SJ, Ren LC et al. An expanded landscape of human long noncoding RNA. *Nucleic Acids Res*. 2019; 47(15) 7842-7856. doi:10.1093/nar/gkz621. PubMed PMID: 31350901.

154. Carninci P, Kvam C, Kitamura A et al. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*. 1996; 37(3) 327-336. doi:10.1006/geno.1996.0567. PubMed PMID: 8938445.

155. Carninci P, Hayashizaki Y. High-efficiency full-length cDNA cloning. *Methods Enzymol*. 1999; 303 19-44. doi:10.1016/s0076-6879(99)03004-9. PubMed PMID: 10349636.

156. Loveland JE, Gilbert JG, Griffiths E, Harrow JL. Community gene annotation in practice. *Database (Oxford)*. 2012; 2012 bas009. doi:10.1093/database/bas009. PubMed PMID: 22434843. PMCID: PMC3308165.

157. Searle SMJ, Gilbert J, Iyer V, Clamp M. The otter annotation system. *Genome Res*. 2004; 14(5) 963-970. doi:10.1101/gr.1864804. PubMed PMID: 15123593. PMCID: PMC479127.

158. Kokocinski F, Harrow J, Hubbard T. AnnoTrack--a tracking system for genome annotation. *BMC Genomics*. 2010; 11 538. doi:10.1186/1471-2164-11-538. PubMed PMID: 20923551. PMCID: PMC3091687.

159. Barson G, Griffiths E. SeqTools: visual tools for manual analysis of sequence alignments. *BMC Res Notes*. 2016; 9 39. doi:10.1186/s13104-016-1847-3. PubMed PMID: 26801397. PMCID: PMC4724122.

160. Taliun D, Harris DN, Kessler MD et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*. 2019; 563866. doi:10.1101/563866.

161. Boccaletto P, Machnicka MA, Purta E et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res*. 2018; 46(D1) D303-D307. doi:10.1093/nar/gkx1030. PubMed PMID: 29106616. PMCID: PMC5753262.

162. Workman RE, Tang AD, Tang PS et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*. 2019; 16(12) 1297-1305. doi:10.1038/s41592-019-0617-2. PubMed PMID: 31740818.

163. Rodriguez JM, Maietta P, Ezkurdia I et al. APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res*. 2013; 41(Database issue) D110-117. doi:10.1093/nar/gks1058. PubMed PMID: 23161672. PMCID: PMC3531113.

164. Tress ML, Wesselink J-J, Frankish A et al. Determination and validation of principal gene products. *Bioinformatics*. 2008; 24(1) 11-17. doi:10.1093/bioinformatics/btm547. PubMed PMID: 18006548. PMCID: PMC2734078.

165. Ezkurdia I, Rodriguez JM, Carrillo-de Santa Pau E, Vázquez J, Valencia A, Tress ML. Most highly expressed protein-coding genes have a single dominant isoform. *J Proteome Res*. 2015; 14(4) 1880-1887. doi:10.1021/pr501286b. PubMed PMID: 25732134. PMCID: PMC4768900.

166. Tress ML, Abascal F, Valencia A. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends Biochem Sci*. 2017; 42(2) 98-110. doi:10.1016/j.tibs.2016.08.008. PubMed PMID: 27712956. PMCID: PMC6526280.

167. Martinez-Gomez L, Abascal F, Jungreis I et al. Few SINEs of life: Alu elements have little evidence for biological relevance despite elevated translation. *NAR Genom Bioinform*. 2020; 2(1) lqz023. doi:10.1093/nargab/lqz023. PubMed PMID: 31886458. PMCID: PMC6924539.

168. Kim M-S, Pinto SM, Getnet D et al. A draft map of the human proteome. *Nature*. 2014; 509(7502) 575-581. doi:10.1038/nature13302. PubMed PMID: 24870542. PMCID: PMC4403737.

169. Wang D, Eraslan B, Wieland T et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol*. 2019; 15(2) e8503. doi:10.15252/msb.20188503. PubMed PMID: 30777892. PMCID: PMC6379049.

170. Rozowsky J, Abyzov A, Wang J et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol*. 2011; 7 522. doi:10.1038/msb.2011.54. PubMed PMID: 21811232. PMCID: PMC3208341.

171. Chen J, Rozowsky J, Galeev TR et al. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun*. 2016; 7 11101. doi:10.1038/ncomms11101. PubMed PMID: 27089393. PMCID: PMC4837449.

172. Guigó R, Flicek P, Abril JF et al. EGASP: the human ENCODE Genome Annotation Assessment Project. *Genome Biol*. 2006; 7 Suppl 1 S2.1-31. doi:10.1186/gb-2006-7-s1-s2. PubMed PMID: 16925836. PMCID: PMC1810551.

173. Engström PG, Steijger T, Sipos B et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods*. 2013; 10(12) 1185-1191. doi:10.1038/nmeth.2722. PubMed PMID: 24185836. PMCID: PMC4018468.

174. Kinsella RJ, Kähäri A, Haider S et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)*. 2011; 2011 bar030. doi:10.1093/database/bar030. PubMed PMID: 21785142. PMCID: PMC3170168.

175. Karolchik D, Hinrichs AS, Furey TS et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004; 32(Database issue) D493-6. doi:10.1093/nar/gkh103. PubMed PMID: 14681465.

176. McLaren W, Gil L, Hunt SE et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016; 17(1) 122. doi:10.1186/s13059-016-0974-4. PubMed PMID: 27268795. PMCID: PMC4893825.