# Richard M Durbin

## List of Publications by Citations

**Source:** https://exaly.com/author-pdf/6764460/richard-m-durbin-publications-by-citations.pdf
**Version:** 2024-04-10

This document has been generated based on the publications and citations recorded by exaly.com. For the latest version of this publication list, visit the link given above.
The third column is the impact factor (IF) of the journal, and the fourth column is the number of citations of the article.

| | | | |
|---|---|---|---|
| 249 papers | 167,841 citations | 108 h-index | 280 g-index |
| 280 ext. papers | 214,227 ext. citations | 17 avg, IF | 8.73 L-index |

| # | Paper | IF | Citations |
|---|---|---|---|
| 249 | The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **2009**, 25, 2078-9 | 7.2 | 30805 |
| 248 | Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **2009**, 25, 1754-60 | 7.2 | 26095 |
| 247 | Initial sequencing and analysis of the human genome. *Nature*, **2001**, 409, 860-921 | 50.4 | 17366 |
| 246 | A global reference for human genetic variation. *Nature*, **2015**, 526, 68-74 | 50.4 | 8599 |
| 245 | Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **2010**, 26, 589-95 | 7.2 | 6791 |
| 244 | The variant call format and VCFtools. *Bioinformatics*, **2011**, 27, 2156-8 | 7.2 | 6200 |
| 243 | A map of human genome variation from population-scale sequencing. *Nature*, **2010**, 467, 1061-73 | 50.4 | 6142 |
| 242 | An integrated map of genetic variation from 1,092 human genomes. *Nature*, **2012**, 491, 56-65 | 50.4 | 6049 |
| 241 | Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, **2003**, 421, 231-7 | 50.4 | 2758 |
| 240 | The Pfam protein families database. *Nucleic Acids Research*, **2004**, 32, D138-41 | 20.1 | 2720 |
| 239 | Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **2008**, 456, 53-9 | 50.4 | 2615 |
| 238 | Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids **1998**, | | 2271 |
| 237 | Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, **2008**, 18, 1851-8 | 9.7 | 2002 |
| 236 | The Pfam protein families database. *Nucleic Acids Research*, **2002**, 30, 276-80 | 20.1 | 1839 |
| 235 | Pfam: clans, web tools and services. *Nucleic Acids Research*, **2006**, 34, D247-51 | 20.1 | 1784 |
| 234 | Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **2004**, 428, 493-521 | 50.4 | 1689 |
| 233 | GeneWise and Genomewise. *Genome Research*, **2004**, 14, 988-95 | 9.7 | 1467 |

| | | | |
|---|---|---|---|
| 232 | A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, **2016**, 48, 1279-83 | 36.3 | 1447 |
| 231 | Inference of human population history from individual whole-genome sequences. *Nature*, **2011**, 475, 493-6 | 50.4 | 1299 |
| 230 | Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, **2011**, 477, 289-94 | 50.4 | 1087 |
| 229 | The Pfam protein families database. *Nucleic Acids Research*, **2000**, 28, 263-6 | 20.1 | 1074 |
| 228 | Population genomics of domestic and wild yeasts. *Nature*, **2009**, 458, 337-41 | 50.4 | 1073 |
| 227 | Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins: Structure, Function and Bioinformatics*, **1997**, 28, 405-20 | 4.2 | 841 |
| 226 | EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, **2009**, 19, 327-35 | 9.7 | 836 |
| 225 | The DNA sequence of the human X chromosome. *Nature*, **2005**, 434, 325-37 | 50.4 | 822 |
| 224 | Ensembl 2012. *Nucleic Acids Research*, **2012**, 40, D84-90 | 20.1 | 798 |
| 223 | The UK10K project identifies rare variants in health and disease. *Nature*, **2015**, 526, 82-90 | 50.4 | 776 |
| 222 | The diploid genome sequence of an Asian individual. *Nature*, **2008**, 456, 60-5 | 50.4 | 744 |
| 221 | Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics*, **2016**, 48, 1443-1448 | 36.3 | 699 |
| 220 | The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. *PLoS Biology*, **2003**, 1, E45 | 9.7 | 677 |
| 219 | Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, **2010**, 464, 721-7 | 50.4 | 668 |
| 218 | Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genetics*, **2008**, 40, 722-9 | 36.3 | 666 |
| 217 | An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, **1987**, 326, 689-91 | 50.4 | 602 |
| 216 | Ensembl 2011. *Nucleic Acids Research*, **2011**, 39, D800-6 | 20.1 | 590 |
| 215 | RNA sequence analysis using covariance models. *Nucleic Acids Research*, **1994**, 22, 2079-88 | 20.1 | 586 |

| | | | |
|---|---|---|---|
| 214 | A large genome center's improvements to the Illumina sequencing system. *Nature Methods*, **2008**, 5, 1005-10 | 21.6 | 575 |
| 213 | Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics*, **2012**, 44, 1084-9 | 36.3 | 572 |
| 212 | Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, **2014**, 46, 919-25 | 36.3 | 569 |
| 211 | A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **1995**, 167, GC1-10 | 3.8 | 563 |
| 210 | The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Research*, **2003**, 31, 315-8 | 20.1 | 556 |
| 209 | WormBase: a multi-species resource for nematode biology and genomics. *Nucleic Acids Research*, **2004**, 32, D411-7 | 20.1 | 543 |
| 208 | A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, **2008**, 26, 779-85 | 44.5 | 533 |
| 207 | Insights into hominid evolution from the gorilla genome sequence. *Nature*, **2012**, 483, 169-75 | 50.4 | 517 |
| 206 | Efficient de novo assembly of large genomes using compressed data structures. *Genome Research*, **2012**, 22, 549-56 | 9.7 | 501 |
| 205 | The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology*, **2005**, 6, R44 | 18.3 | 492 |
| 204 | BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature Methods*, **2008**, 5, 409-15 | 21.6 | 484 |
| 203 | Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, **2013**, 2, 10 | 7.6 | 461 |
| 202 | Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nature Protocols*, **2012**, 7, 500-7 | 18.8 | 460 |
| 201 | InterPro, progress and status in 2005. *Nucleic Acids Research*, **2005**, 33, D201-5 | 20.1 | 426 |
| 200 | Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science*, **2010**, 328, 593-9 | 33.3 | 419 |
| 199 | The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Research*, **2009**, 19, 1316-23 | 9.7 | 415 |
| 198 | TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Research*, **2006**, 34, D572-80 | 20.1 | 383 |
| 197 | Revising the human mutation rate: implications for understanding human evolution. *Nature Reviews Genetics*, **2012**, 13, 745-53 | 30.1 | 369 |

| | | | |
|---|---|---|---|
| 196 | Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Research*, **2017**, 27, 849-864 | 9.7 | 365 |
| 195 | Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Research*, **2011**, 21, 2224-41 | 9.7 | 364 |
| 194 | Patterns of cis regulatory variation in diverse human populations. *PLoS Genetics*, **2012**, 8, e1002639 | 6 | 361 |
| 193 | A dimension reduction framework for understanding cortical maps. *Nature*, **1990**, 343, 644-7 | 50.4 | 346 |
| 192 | Dindel: accurate indel calls from short-read data. *Genome Research*, **2011**, 21, 961-73 | 9.7 | 341 |
| 191 | The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genetics*, **2011**, 7, e1002003 | 6 | 336 |
| 190 | Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, **2018**, 115, 4325-4333 | 11.5 | 334 |
| 189 | POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*, **2015**, 349, aab3884 | 33.3 | 317 |
| 188 | An overview of Ensembl. *Genome Research*, **2004**, 14, 925-8 | 9.7 | 316 |
| 187 | Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature*, **2015**, 526, 112-7 | 50.4 | 308 |
| 186 | Product Units: A Computationally Powerful and Biologically Plausible Extension to Backpropagation Networks. *Neural Computation*, **1989**, 1, 133-142 | 2.9 | 305 |
| 185 | A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Computational Biology*, **2010**, 6, e1000770 | 5 | 295 |
| 184 | Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*, **2017**, 546, 370-375 | 50.4 | 294 |
| 183 | WormBase: a comprehensive resource for nematode research. *Nucleic Acids Research*, **2010**, 38, D463-7 | 20.1 | 289 |
| 182 | Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *American Journal of Human Genetics*, **2013**, 93, 876-90 | 11 | 269 |
| 181 | A genomic history of Aboriginal Australia. *Nature*, **2016**, 538, 207-214 | 50.4 | 268 |
| 180 | Using GeneWise in the Drosophila annotation experiment. *Genome Research*, **2000**, 10, 547-8 | 9.7 | 263 |
| 179 | BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, **2016**, 32, 1749-51 | 7.2 | 256 |

| | | | |
|---|---|---|---|
| 178 | Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genetics*, **2014**, 10, e1004494 | 6 | 243 |
| 177 | Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics*, **2014**, 30, 1266-72 | 7.2 | 241 |
| 176 | Ensembl@ 10th year. *Nucleic Acids Research*, **2010**, 38, D557-62 | 20.1 | 240 |
| 175 | TreeFam: 2008 Update. *Nucleic Acids Research*, **2008**, 36, D735-40 | 20.1 | 234 |
| 174 | Trait variation in yeast is defined by population history. *PLoS Genetics*, **2011**, 7, e1002111 | 6 | 230 |
| 173 | Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nature Biotechnology*, **2018**, 36, 875-879 | 44.5 | 223 |
| 172 | Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, **2020**, 36, 2896-2898 | 7.2 | 222 |
| 171 | QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics*, **2002**, 18, 1546-7 | 7.2 | 214 |
| 170 | Structure and expression of the Huntington@ disease gene: evidence against simple inactivation due to an expanded CAG repeat. *Somatic Cell and Molecular Genetics*, **1994**, 20, 27-38 | | 210 |
| 169 | A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, **2014**, 31, 872-88 | 8.3 | 206 |
| 168 | Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, **2015**, 350, 1493-1498 | 33.3 | 204 |
| 167 | Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter?. *Trends in Ecology and Evolution*, **2018**, 33, 582-594 | 10.9 | 200 |
| 166 | Prepublication data sharing. *Nature*, **2009**, 461, 168-70 | 50.4 | 197 |
| 165 | Insights into human genetic variation and population history from 929 diverse genomes. *Science*, **2020**, 367, | 33.3 | 196 |
| 164 | Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature Communications*, **2015**, 6, 8111 | 17.4 | 186 |
| 163 | Health and population effects of rare gene knockouts in adult humans with related parents. *Science* , **2016**, 352, 474-7 | 33.3 | 185 |
| 162 | Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biology*, **2013**, 14, R75 | 18.3 | 185 |
| 161 | Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Research*, **2011**, 21, 1131-8 | 9.7 | 185 |

| | | | |
|---|---|---|---|
| 160 | Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, **2017**, 49, 913-924 | 36.3 | 178 |
| 159 | Maximum discrimination hidden Markov models of sequence consensus. *Journal of Computational Biology*, **1995**, 2, 9-23 | 1.7 | 167 |
| 158 | Efficient construction of an assembly string graph using the FM-index. *Bioinformatics*, **2010**, 26, i367-73 | 7.2 | 164 |
| 157 | The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science*, **2018**, 360, | 33.3 | 162 |
| 156 | Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, **2021**, 592, 737-746 | 36.4 | 161 |
| 155 | Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution*, **2018**, 2, 1940-1955 | 12.3 | 160 |
| 154 | WormBase 2012: more genomes, more data, new website. *Nucleic Acids Research*, **2012**, 40, D735-41 | 20.1 | 159 |
| 153 | Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Research*, **1999**, 9, 815-24 | 9.7 | 141 |
| 152 | Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nature Genetics*, **2015**, 47, 88-91 | 36.3 | 140 |
| 151 | WormBase: a comprehensive data resource for Caenorhabditis biology and genomics. *Nucleic Acids Research*, **2005**, 33, D383-9 | 20.1 | 140 |
| 150 | The population history of northeastern Siberia since the Pleistocene. *Nature*, **2019**, 570, 182-188 | 50.4 | 137 |
| 149 | InterPro: an integrated documentation resource for protein families, domains and functional sites. *Briefings in Bioinformatics*, **2002**, 3, 225-35 | 13.4 | 137 |
| 148 | An Analysis of the Elastic Net Approach to the Traveling Salesman Problem. *Neural Computation*, **1989**, 1, 348-358 | 2.9 | 137 |
| 147 | Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nature Genetics*, **1995**, 10, 67-76 | 36.3 | 135 |
| 146 | Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. *Nature Communications*, **2016**, 7, 12039 | 17.4 | 124 |
| 145 | High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biology*, **2012**, 13, 26 | 18.3 | 122 |
| 144 | SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. *Genome Research*, **2011**, 21, 952-60 | 9.7 | 117 |
| 143 | The genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nature Genetics*, **2015**, 47, 235-41 | 36.3 | 111 |

| | | | |
|---|---|---|---|
| 142 | A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Research*, **2001**, 29, 4006-13 | 20.1 | 110 |
| 141 | Extending reference assembly models. *Genome Biology*, **2015**, 16, 13 | 18.3 | 107 |
| 140 | Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *American Journal of Human Genetics*, **2015**, 96, 986-91 | 11 | 107 |
| 139 | The complete sequence of a human genome.. *Science*, **2022**, 376, 44-53 | 33.3 | 107 |
| 138 | Analysis of protein domain families in Caenorhabditis elegans. *Genomics*, **1997**, 46, 200-16 | 4.3 | 106 |
| 137 | De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Scientific Reports*, **2017**, 7, 3935 | 4.9 | 101 |
| 136 | Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nature Communications*, **2016**, 7, 10408 | 17.4 | 100 |
| 135 | Comparative ab initio prediction of gene structures using pair HMMs. *Bioinformatics*, **2002**, 18, 1309-18 | 7.2 | 93 |
| 134 | High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. *Genetics*, **2013**, 195, 1141-55 | 4 | 91 |
| 133 | WormBase 2007. *Nucleic Acids Research*, **2008**, 36, D612-7 | 20.1 | 91 |
| 132 | Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nature Genetics*, **2018**, 50, 1574-1583 | 36.3 | 91 |
| 131 | Immunofluorescence Analysis and Diagnosis of Primary Ciliary Dyskinesia with Radial Spoke Defects. *American Journal of Respiratory Cell and Molecular Biology*, **2015**, 53, 563-73 | 5.7 | 90 |
| 130 | WormBase: a cross-species database for comparative genomics. *Nucleic Acids Research*, **2003**, 31, 133-7 | 20.1 | 90 |
| 129 | Mapping trait loci by use of inferred ancestral recombination graphs. *American Journal of Human Genetics*, **2006**, 79, 910-22 | 11 | 87 |
| 128 | Genetic interactions affecting human gene expression identified by variance association mapping. *ELife*, **2014**, 3, e01381 | 8.9 | 86 |
| 127 | Estimating telomere length from whole genome sequence data. *Nucleic Acids Research*, **2014**, 42, e75 | 20.1 | 85 |
| 126 | Dynamic programming alignment accuracy. *Journal of Computational Biology*, **1998**, 5, 493-504 | 1.7 | 84 |
| 125 | Optimal Numberings of an $N times N$ Array. *SIAM Journal on Algebraic and Discrete Methods*, **1986**, 7, 571-582 | | 84 |

| | | | |
|---|---|---|---|
| 124 | DNAH11 Localization in the Proximal Region of Respiratory Cilia Defines Distinct Outer Dynein Arm Complexes. *American Journal of Respiratory Cell and Molecular Biology*, **2016**, 55, 213-24 | 5.7 | 79 |
| 123 | WormBase: new content and better access. *Nucleic Acids Research*, **2007**, 35, D506-10 | 20.1 | 76 |
| 122 | Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. *American Journal of Human Genetics*, **2017**, 100, 865-884 | 11 | 74 |
| 121 | A High-Quality Genome Assembly from a Single Mosquito Using PacBio Sequencing. *Genes*, **2019**, 10, | 4.2 | 72 |
| 120 | Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nature Methods*, **2020**, 17, 615-620 | 21.6 | 69 |
| 119 | Deficiency of ECHS1 causes mitochondrial encephalopathy with cardiac involvement. *Annals of Clinical and Translational Neurology*, **2015**, 2, 492-509 | 5.3 | 69 |
| 118 | WormBase: better software, richer content. *Nucleic Acids Research*, **2006**, 34, D475-8 | 20.1 | 68 |
| 117 | GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. *Genome Research*, **2002**, 12, 1418-27 | 9.7 | 65 |
| 116 | Gene structure conservation aids similarity based gene prediction. *Nucleic Acids Research*, **2004**, 32, 776-83 | 20.1 | 62 |
| 115 | Software for genome mapping by fingerprinting techniques. *Bioinformatics*, **1988**, 4, 125-32 | 7.2 | 61 |
| 114 | Clustering of phosphorylation site recognition motifs can be exploited to predict the targets of cyclin-dependent kinase. *Genome Biology*, **2007**, 8, R23 | 18.3 | 60 |
| 113 | The complete sequence of a human genome | | 58 |
| 112 | TTC25 Deficiency Results in Defects of the Outer Dynein Arm Docking Machinery and Primary Ciliary Dyskinesia with Left-Right Body Asymmetry Randomization. *American Journal of Human Genetics*, **2016**, 99, 460-9 | 11 | 58 |
| 111 | Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS Genetics*, **2011**, 7, e1001276 | 6 | 57 |
| 110 | Whole-genome sequence-based analysis of thyroid function. *Nature Communications*, **2015**, 6, 5681 | 17.4 | 56 |
| 109 | Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. *Proceedings of the National Academy of Sciences of the United States of America*, **2007**, 104, 17713-8 | 11.5 | 56 |
| 108 | A systematic comparative and structural analysis of protein phosphorylation sites based on the mtcPTM database. *Genome Biology*, **2007**, 8, R90 | 18.3 | 53 |
| 107 | Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. *Nature Communications*, **2017**, 8, 303 | 17.4 | 52 |

| | | | |
|---|---|---|---|
| 106 | Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. *Genome Biology*, **2014**, 15, R88 | 18.3 | 51 |
| 105 | Bi-allelic Truncating Mutations in TANGO2 Cause Infancy-Onset Recurrent Metabolic Crises with Encephalocardiomyopathy. *American Journal of Human Genetics*, **2016**, 98, 358-62 | 11 | 49 |
| 104 | No evidence for maintenance of a sympatric species barrier by chromosomal inversions. *Evolution Letters*, **2017**, 1, 138-154 | 5.3 | 49 |
| 103 | A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nature Communications*, **2014**, 5, 4871 | 17.4 | 46 |
| 102 | Image analysis of restriction enzyme fingerprint autoradiograms. *Bioinformatics*, **1989**, 5, 101-6 | 7.2 | 45 |
| 101 | A workbench for large-scale sequence homology analysis. *Bioinformatics*, **1994**, 10, 301-7 | 7.2 | 44 |
| 100 | Extent, causes, and consequences of small RNA expression variation in human adipose tissue. *PLoS Genetics*, **2012**, 8, e1002704 | 6 | 43 |
| 99 | Enhanced protein domain discovery by using language modeling techniques from speech recognition. *Proceedings of the National Academy of Sciences of the United States of America*, **2003**, 100, 4516-20 | 11.5 | 42 |
| 98 | Detecting archaic introgression using an unadmixed outgroup. *PLoS Genetics*, **2018**, 14, e1007641 | 6 | 42 |
| 97 | Whole-genome view of the consequences of a population bottleneck using 2926 genome sequences from Finland and United Kingdom. *European Journal of Human Genetics*, **2017**, 25, 477-484 | 5.3 | 41 |
| 96 | Ancestral Hybridization Facilitated Species Diversification in the Lake Malawi Cichlid Fish Adaptive Radiation. *Molecular Biology and Evolution*, **2020**, 37, 1100-1113 | 8.3 | 41 |
| 95 | Copy number variant detection in inbred strains from short read sequence data. *Bioinformatics*, **2010**, 26, 565-7 | 7.2 | 40 |
| 94 | A probabilistic model of 3Qend formation in Caenorhabditis elegans. *Nucleic Acids Research*, **2004**, 32, 3392-9 | 20.1 | 40 |
| 93 | Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. *PLoS Genetics*, **2014**, 10, e1004798 | 6 | 38 |
| 92 | Towards complete and error-free genome assemblies of all vertebrate species | | 38 |
| 91 | Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. *Nature Communications*, **2017**, 8, 15927 | 17.4 | 37 |
| 90 | A graph-based approach to diploid genome assembly. *Bioinformatics*, **2018**, 34, i105-i114 | 7.2 | 36 |
| 89 | Genome Graphs | | 34 |

| | | | |
|---|---|---|---|
| 88 | Managing clinically significant findings in research: the UK10K example. *European Journal of Human Genetics*, **2014**, 22, 1100-4 | 5.3 | 33 |
| 87 | Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association*, **2020**, 115, 1472-1487 | 2.8 | 33 |
| 86 | Deep short-read sequencing of chromosome 17 from the mouse strains A/J and CAST/Ei identifies significant germline variation and candidate genes that regulate liver triglyceride levels. *Genome Biology*, **2009**, 10, R112 | 18.3 | 32 |
| 85 | A high-content platform to characterise human induced pluripotent stem cell lines. *Methods*, **2016**, 96, 85-96 | 4.6 | 28 |
| 84 | Alfresco--a workbench for comparative genomic sequence analysis. *Genome Research*, **2000**, 10, 1148-57 | 9.7 | 27 |
| 83 | Identity-by-descent-based phasing and imputation in founder populations using graphical models. *Genetic Epidemiology*, **2011**, 35, 853-60 | 2.6 | 26 |
| 82 | Haplotype-aware graph indexes. *Bioinformatics*, **2020**, 36, 400-407 | 7.2 | 23 |
| 81 | Whole-exome sequencing of 228 patients with sporadic Parkinson© disease. *Scientific Reports*, **2017**, 7, 41188 | 4.9 | 21 |
| 80 | Estimation of epistatic variance components and heritability in founder populations and crosses. *Genetics*, **2014**, 198, 1405-16 | 4 | 19 |
| 79 | Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biology*, **2021**, 22, 120 | 18.3 | 19 |
| 78 | The ACEDB Genome Database **1994**, 45-55 | | 19 |
| 77 | The anatomy of successful computational biology software. *Nature Biotechnology*, **2013**, 31, 894-7 | 44.5 | 18 |
| 76 | Homozygous loss-of-function variants in European cosmopolitan and isolate populations. *Human Molecular Genetics*, **2015**, 24, 5464-74 | 5.6 | 18 |
| 75 | Gene expression and development databases for C. elegans. *Seminars in Cell and Developmental Biology*, **1997**, 8, 459-67 | 7.5 | 17 |
| 74 | Whole genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow | | 17 |
| 73 | A genome-wide survey of genetic variation in gorillas using reduced representation sequencing. *PLoS ONE*, **2013**, 8, e65066 | 3.7 | 16 |
| 72 | A Method for Checking Genomic Integrity in Cultured Cell Lines from SNP Genotyping Data. *PLoS ONE*, **2016**, 11, e0155014 | 3.7 | 16 |
| 71 | Population-scale proteome variation in human induced pluripotent stem cells. *ELife*, **2020**, 9, | 8.9 | 16 |

| | | | |
|---|---|---|---|
| 70 | Identifying Extrinsic versus Intrinsic Drivers of Variation in Cell Behavior in Human iPSC Lines from Healthy Donors. *Cell Reports*, **2019**, 26, 2078-2087.e3 | 10.6 | 16 |
| 69 | Using reference-free compressed data structures to analyze sequencing reads from thousands of human genomes. *Genome Research*, **2017**, 27, 300-309 | 9.7 | 15 |
| 68 | The Earth BioGenome Project 2020: Starting the clock.. *Proceedings of the National Academy of Sciences of the United States of America*, **2022**, 119, | 11.5 | 15 |
| 67 | A reference panel of 64,976 haplotypes for genotype imputation | | 15 |
| 66 | Enhanced protein domain discovery using taxonomy. *BMC Bioinformatics*, **2004**, 5, 56 | 3.6 | 14 |
| 65 | The C. elegans expression pattern database: a beginning. *Trends in Genetics*, **1996**, 12, 370-371 | 8.5 | 14 |
| 64 | Diffraction methods for biological macromolecules. Oscillation method with large unit cells. *Methods in Enzymology*, **1985**, 114, 211-37 | 1.7 | 14 |
| 63 | htsget: a protocol for securely streaming genomic data. *Bioinformatics*, **2019**, 35, 119-121 | 7.2 | 13 |
| 62 | Accounting for Non-genetic Factors Improves the Power of eQTL Studies **2008**, 411-422 | | 13 |
| 61 | Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly | | 13 |
| 60 | Sequence assembly with CAFTOOLS. *Genome Research*, **1998**, 8, 260-7 | 9.7 | 12 |
| 59 | Method for calculation of probability of matching a bounded regular expression in a random data string. *Journal of Computational Biology*, **1995**, 2, 25-31 | 1.7 | 12 |
| 58 | Late Quaternary dynamics of Arctic biota from ancient environmental genomics. *Nature*, **2021**, 600, 86-92 | 30.4 | 12 |
| 57 | Insights into human genetic variation and population history from 929 diverse genomes | | 12 |
| 56 | Crumble: reference free lossy compression of sequence quality values. *Bioinformatics*, **2019**, 35, 337-339 | 7.2 | 11 |
| 55 | WormBase: Annotating many nematode genomes. *Worm*, **2012**, 1, 15-21 | | 11 |
| 54 | Vertebrate gene finding from multiple-species alignments using a two-level strategy. *Genome Biology*, **2006**, 7 Suppl 1, S6.1-12 | 18.3 | 11 |
| 53 | A direct multi-generational estimate of the human mutation rate from autozygous segments seen in thousands of parentally related individuals | | 11 |

| | | | |
|---|---|---|---|
| 52 | Sequence variation aware genome references and read mapping with the variation graph toolkit | | 11 |
| 51 | Genomix: a method for combining gene-findersQpredictions, which uses evolutionary conservation of sequence and intron-exon structure. *Bioinformatics*, **2007**, 23, 1468-75 | 7.2 | 10 |
| 50 | A table-driven, full-sensitivity similarity search algorithm. *Journal of Computational Biology*, **2003**, 10, 103-17 | 1.7 | 9 |
| 49 | Sequence locally, think globally: The Darwin Tree of Life Project.. *Proceedings of the National Academy of Sciences of the United States of America*, **2022**, 119, | 11.5 | 9 |
| 48 | trio-sga: facilitating de novo assembly of highly heterozygous genomes with parent-child trios | | 9 |
| 47 | Haplotype-aware graph indexes | | 9 |
| 46 | souporcell: Robust clustering of single cell RNAseq by genotype and ambient RNA inference without reference genotypes | | 9 |
| 45 | A haplotype-aware de novo assembly of related individuals using pedigree sequence graph. *Bioinformatics*, **2020**, 36, 2385-2392 | 7.2 | 9 |
| 44 | Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. *Genome Biology*, **2020**, 21, 250 | 18.3 | 9 |
| 43 | A high-quality, chromosome-level genome assembly of the Black Soldier Fly (Hermetia illucens L.). *G3: Genes, Genomes, Genetics*, **2021**, 11, | 3.2 | 9 |
| 42 | Inferring selection on amino acid preference in protein domains. *Molecular Biology and Evolution*, **2009**, 26, 527-36 | 8.3 | 7 |
| 41 | A conserved sequence motif in 3Quntranslated regions of ribosomal protein mRNAs in nematodes. *Rna*, **2006**, 12, 1786-9 | 5.8 | 7 |
| 40 | Multiple laboratory mouse reference genomes define strain specific haplotypes and novel functional loci | | 7 |
| 39 | Base qualities help sequencing software. *Genome Research*, **1998**, 8, 161-2 | 9.7 | 6 |
| 38 | Why sequence all eukaryotes?. *Proceedings of the National Academy of Sciences of the United States of America*, **2022**, 119, | 11.5 | 6 |
| 37 | The rate of false polymorphisms introduced when imputing genotypes from global imputation panels | | 6 |
| 36 | A haplotype-resolved, de novo genome assembly for the wood tiger moth (Arctia plantaginis) through trio binning. *GigaScience*, **2020**, 9, | 7.6 | 6 |
| 35 | Environmental genomics of Late Pleistocene black bears and giant short-faced bears. *Current Biology*, **2021**, 31, 2728-2736.e8 | 6.3 | 6 |

| | | | |
|---|---|---|---|
| 34 | Pathway-based factor analysis of gene expression data produces highly heritable phenotypes that associate with age. *G3: Genes, Genomes, Genetics*, **2015**, 5, 839-47 | 3.2 | 5 |
| 33 | Whole-exome sequencing in an isolated population from the Dalmatian island of Vis. *European Journal of Human Genetics*, **2016**, 24, 1479-87 | 5.3 | 5 |
| 32 | Complete vertebrate mitogenomes reveal widespread gene duplications and repeats | | 5 |
| 31 | Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph | | 5 |
| 30 | The genome sequence of the brown trout, Linnaeus 1758. *Wellcome Open Research*, **2021**, 6, 108 | 4.8 | 5 |
| 29 | Standards recommendations for the Earth BioGenome Project.. *Proceedings of the National Academy of Sciences of the United States of America*, **2022**, 119, | 11.5 | 4 |
| 28 | Inferring human population size and separation history from multiple genome sequences | | 4 |
| 27 | Health and population effects of rare gene knockouts in adult humans with related parents | | 4 |
| 26 | Ethical, legal, and social issues in the Earth BioGenome Project.. *Proceedings of the National Academy of Sciences of the United States of America*, **2022**, 119, | 11.5 | 3 |
| 25 | Placing ancient DNA sequences into reference phylogenies.. *Molecular Biology and Evolution*, **2022**, | 8.3 | 3 |
| 24 | The C. elegans expression pattern database: a beginning **1996**, 12, 370-370 | | 3 |
| 23 | Reference-based phasing using the Haplotype Reference Consortium panel | | 3 |
| 22 | Identifying and removing haplotypic duplication in primary genome assemblies | | 3 |
| 21 | False gene and chromosome losses affected by assembly and sequence errors | | 3 |
| 20 | Genomic consequences of domestication of the Siamese fighting fish.. *Science Advances*, **2022**, 8, eabm4950 | 11.5 | 3 |
| 19 | Viral coinfection analysis using a MinHash toolkit. *BMC Bioinformatics*, **2019**, 20, 389 | 3.6 | 2 |
| 18 | The genome sequence of the Eurasian red squirrel, Linnaeus 1758. *Wellcome Open Research*, **2020**, 5, 18 | 4.8 | 2 |
| 17 | GFAKluge: A C++ library and command line utilities for the Graphical Fragment Assembly formats. *Journal of Open Source Software*, **2019**, 4, | 5.2 | 2 |

| | | | |
|---|---|---|---|
| 16 | Mapping epigenetic divergence in the massive radiation of Lake Malawi cichlid fishes. *Nature Communications*, **2021**, 12, 5870 | 17.4 | 2 |
| 15 | A haplotype-aware de novo assembly of related individuals using pedigree graph | | 2 |
| 14 | Placing ancient DNA sequences into reference phylogenies | | 2 |
| 13 | Ancestral hybridisation facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation | | 2 |
| 12 | The genome sequence of the ringlet, Aphantopus hyperantus Linnaeus 1758. *Wellcome Open Research*,6, 165 | 4.8 | 2 |
| 11 | Epigenetic Divergence during Early Stages of Speciation in an African Crater Lake Cichlid Fish | | 2 |
| 10 | The genome sequence of the eastern grey squirrel, Gmelin, 1788. *Wellcome Open Research*, **2020**, 5, 27 | 4.8 | 1 |
| 9 | The genome sequence of the Eurasian river otter, Lutra lutra Linnaeus 1758. *Wellcome Open Research*, **2020**, 5, 33 | 4.8 | 1 |
| 8 | Mapping epigenetic divergence in the massive radiation of Lake Malawi cichlid fishes | | 1 |
| 7 | The population history of northeastern Siberia since the Pleistocene | | 1 |
| 6 | Genomic consequences of domestication of the Siamese fighting fish | | 1 |
| 5 | Substantial somatic genomic variation and selection for BCOR mutations in human induced pluripotent stem cells | | 1 |
| 4 | Differential use of multiple genetic sex determination systems in divergent ecomorphs of an African crater lake cichlid | | 1 |
| 3 | [X]uniqMAP: unique gene sequence regions in the human and mouse genomes. *BMC Genomics*, **2006**, 7, 249 | 4.5 | 0 |
| 2 | Efficient iterative Hi-C scaffolder based on N-best neighbors. *BMC Bioinformatics*, **2021**, 22, 569 | 3.6 | 0 |
| 1 | The genome sequence of the European golden eagle, Linnaeus 1758. *Wellcome Open Research*, **2021**, 6, 112 | 4.8 | 0 |