# Ewan Birney

## List of Publications by Year in Descending Order

**Source:** https://exaly.com/author-pdf/1588031/ewan-birney-publications-by-year.pdf
**Version:** 2024-04-27

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.

| | | | |
|---|---|---|---|
| 235<br>papers | 106,543<br>citations | 99<br>h-index | 264<br>g-index |
| 264<br>ext. papers | 126,037<br>ext. citations | 20.4<br>avg, IF | 8.55<br>L-index |

| # | Paper | IF | Citations |
|---|---|---|---|
| 235 | The Medaka Inbred Kiyosu-Karlsruhe (MIKK) panel.. *Genome Biology*, **2022**, 23, 59 | 18.3 | 0 |
| 234 | Genomic variations and epigenomic landscape of the Medaka Inbred Kiyosu-Karlsruhe (MIKK) panel.. *Genome Biology*, **2022**, 23, 58 | 18.3 | 1 |
| 233 | Nanopore ReCappable sequencing maps SARS-CoV-2 5' capping sites and provides new insights into the structure of sgRNAs.. *Nucleic Acids Research*, **2022**, | 20.1 | 2 |
| 232 | Selective clonal persistence of human retroviruses in vivo: Radial chromatin organization, integration site, and host transcription.. *Science Advances*, **2022**, 8, eabm6210 | 14.3 | 1 |
| 231 | The Gene Curation Coalition: A global effort to harmonize gene-disease evidence resources.. *Genetics in Medicine*, **2022**, | 8.1 | 4 |
| 230 | RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nature Communications*, **2021**, 12, 7198 | 17.4 | 18 |
| 229 | GA4GH: International policies and standards for data sharing across genomic research and healthcare.. *Cell Genomics*, **2021**, 1, 100029-100029 | | 20 |
| 228 | The European Bioinformatics Institute (EMBL-EBI) in 2021. *Nucleic Acids Research*, **2021**, | 20.1 | 2 |
| 227 | AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, **2021**, | 20.1 | 285 |
| 226 | Genomic reconstruction of the SARS-CoV-2 epidemic in England. *Nature*, **2021**, | 50.4 | 16 |
| 225 | The European Bioinformatics Institute: empowering cooperation in response to a global health crisis. *Nucleic Acids Research*, **2021**, 49, D29-D37 | 20.1 | 12 |
| 224 | Genetic variation affects morphological retinal phenotypes extracted from UK Biobank optical coherence tomography images. *PLoS Genetics*, **2021**, 17, e1009497 | 6 | 5 |
| 223 | REMBI: Recommended Metadata for Biological Images-enabling reuse of microscopy data in biology. *Nature Methods*, **2021**, 18, 1418-1422 | 21.6 | 16 |
| 222 | Genome-wide meta-analysis identifies 127 open-angle glaucoma loci with consistent effect across ancestries. *Nature Communications*, **2021**, 12, 1258 | 17.4 | 47 |
| 221 | Highly accurate protein structure prediction for the human proteome. *Nature*, **2021**, 596, 590-596 | 50.4 | 399 |
| 220 | The International Human Genome Project. *Human Molecular Genetics*, **2021**, 30, R161-R163 | 5.6 | 3 |
| 219 | Personalized profiles for disease risk must capture all facets of health. *Nature*, **2021**, 597, 175-177 | 50.4 | 5 |

| | | | |
|---|---|---|---|
| 218 | The European Bioinformatics Institute in 2020: building a global infrastructure of interconnected data resources for the life sciences. *Nucleic Acids Research*, **2020**, 48, D17-D23 | 20.1 | 18 |
| 217 | Comparison of Associations with Different Macular Inner Retinal Thickness Parameters in a Large Cohort: The UK Biobank. *Ophthalmology*, **2020**, 127, 62-71 | 7.3 | 20 |
| 216 | Genetic and functional insights into the fractal structure of the heart. *Nature*, **2020**, 584, 589-594 | 50.4 | 26 |
| 215 | Leveraging European infrastructures to access 1 million human genomes by 2022. *Nature Reviews Genetics*, **2019**, 20, 693-701 | 30.1 | 36 |
| 214 | GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nature Genetics*, **2019**, 51, 343-353 | 36.3 | 71 |
| 213 | Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proceedings of the National Academy of Sciences of the United States of America*, **2019**, 116, 10883-10888 | 11.5 | 54 |
| 212 | Biomolecular Data Resources: Bioinformatics Infrastructure for Biomedical Data Science. *Annual Review of Biomedical Data Science*, **2019**, 2, 199-222 | 5.6 | 6 |
| 211 | The Convergence of Research and Clinical Genomics. *American Journal of Human Genetics*, **2019**, 104, 781-783 | 11 | 3 |
| 210 | 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 |
| 209 | The European Bioinformatics Institute in 2018: tools, infrastructure and training. *Nucleic Acids Research*, **2019**, 47, D15-D22 | 20.1 | 21 |
| 208 | Integrating Genomics into Healthcare: A Global Responsibility. *American Journal of Human Genetics*, **2019**, 104, 13-20 | 11 | 166 |
| 207 | The European Bioinformatics Institute in 2017: data coordination and integration. *Nucleic Acids Research*, **2018**, 46, D21-D29 | 20.1 | 36 |
| 206 | A roadmap for restoring trust in Big Data. *Lancet Oncology, The*, **2018**, 19, 1014-1015 | 21.7 | 10 |
| 205 | PhenotypeSimulator: A comprehensive framework for simulating multi-trait, multi-locus genotype to phenotype relationships. *Bioinformatics*, **2018**, 34, 2951-2956 | 7.2 | 9 |
| 204 | ChromoTrace: Computational reconstruction of 3D chromosome configurations for super-resolution microscopy. *PLoS Computational Biology*, **2018**, 14, e1006002 | 5 | 1 |
| 203 | A call for public archives for biological image data. *Nature Methods*, **2018**, 15, 849-854 | 21.6 | 61 |
| 202 | The human leukemia virus HTLV-1 alters the structure and transcription of host chromatin in cis. *ELife*, **2018**, 7, | 8.9 | 30 |
| 201 | Interactions between genetic variation and cellular environment in skeletal muscle gene expression. *PLoS ONE*, **2018**, 13, e0195788 | 3.7 | 9 |

| | | | |
|---|---|---|---|
| 200 | A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nature Genetics*, **2017**, 49, 341-348 | 36.3 | 54 |
| 199 | Promoter shape varies across populations and affects promoter evolution and expression noise. *Nature Genetics*, **2017**, 49, 550-558 | 36.3 | 51 |
| 198 | Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*, **2017**, 546, 370-375 | 50.4 | 294 |
| 197 | HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nature Medicine*, **2017**, 23, 517-525 | 50.5 | 444 |
| 196 | Genetic variants regulating expression levels and isoform diversity during embryogenesis. *Nature*, **2017**, 541, 402-406 | 50.4 | 39 |
| 195 | Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Research* , **2017**, 45, D985-D994 | 20.1 | 241 |
| 194 | MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Research*, **2017**, 6, 760 | 3.6 | 79 |
| 193 | eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell Reports*, **2016**, 17, 2137-2150 | 10.6 | 70 |
| 192 | The topography of mutational processes in breast cancer genomes. *Nature Communications*, **2016**, 7, 11383 | 17.4 | 172 |
| 191 | Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. *Nature Communications*, **2016**, 7, 12910 | 17.4 | 74 |
| 190 | Ensembl 2016. *Nucleic Acids Research*, **2016**, 44, D710-6 | 20.1 | 1094 |
| 189 | The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic Acids Research*, **2016**, 44, D20-6 | 20.1 | 91 |
| 188 | Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *PLoS Genetics*, **2016**, 12, e1006105 | 6 | 137 |
| 187 | The Mighty Fruit Fly Moves into Outbred Genetics. *PLoS Genetics*, **2016**, 12, e1006388 | 6 | 2 |
| 186 | The EMBL-EBI channel. *F1000Research*, **2016**, 5, 52 | 3.6 | |
| 185 | Ensembl regulation resources. *Database: the Journal of Biological Databases and Curation*, **2016**, 2016, | 5 | 33 |
| 184 | Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, **2016**, 534, 47-54 | 50.4 | 1193 |
| 183 | Human genomics: The end of the start for population sequencing. *Nature*, **2015**, 526, 52-3 | 50.4 | 48 |

| | | | |
|---|---|---|---|
| 182 | The UK10K project identifies rare variants in health and disease. *Nature*, **2015**, 526, 82-90 | 50.4 | 776 |
| 181 | Using human genetics to make new medicines. *Nature Reviews Genetics*, **2015**, 16, 561-2 | 30.1 | 22 |
| 180 | Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nature Genetics*, **2015**, 47, 1415-25 | 36.3 | 292 |
| 179 | Ensembl 2015. *Nucleic Acids Research*, **2015**, 43, D662-9 | 20.1 | 1013 |
| 178 | MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research*, **2015**, 4, 1075 | 3.6 | 210 |
| 177 | Integrative knowledge management to enhance pharmaceutical R&D. *Nature Reviews Drug Discovery*, **2014**, 13, 239-40 | 64.1 | 9 |
| 176 | Reply to Brunet and Doolittle: Both selected effect and causal role elements can influence human biology and disease. *Proceedings of the National Academy of Sciences of the United States of America* , **2014**, 111, E3366 | 11.5 | 22 |
| 175 | The European Bioinformatics Institute's data resources 2014. *Nucleic Acids Research*, **2014**, 42, D18-25 | 20.1 | 54 |
| 174 | Ensembl 2014. *Nucleic Acids Research*, **2014**, 42, D749-55 | 20.1 | 1087 |
| 173 | Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, **2014**, 111, 6131-8 | 11.5 | 490 |
| 172 | The Reactome pathway knowledgebase. *Nucleic Acids Research*, **2014**, 42, D472-7 | 20.1 | 1188 |
| 171 | Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. *PLoS Genetics*, **2014**, 10, e1004798 | 6 | 38 |
| 170 | The EBI RDF platform: linked open data for the life sciences. *Bioinformatics*, **2014**, 30, 1338-9 | 7.2 | 160 |
| 169 | Genomic and phenotypic characterization of a wild medaka population: towards the establishment of an isogenic population genetic resource in fish. *G3: Genes, Genomes, Genetics*, **2014**, 4, 433-45 | 3.2 | 34 |
| 168 | Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Research*, **2013**, 41, 827-41 | 20.1 | 383 |
| 167 | Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, **2013**, 494, 77-80 | 50.4 | 501 |
| 166 | Ensembl 2013. *Nucleic Acids Research*, **2013**, 41, D48-55 | 20.1 | 797 |
| 165 | Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nature Genetics*, **2013**, 45, 912-917 | 36.3 | 276 |

| | | | |
|---|---|---|---|
| 164 | Policy challenges of clinical genome sequencing. *BMJ, The*, **2013**, 347, f6845 | 5.9 | 44 |
| 163 | Highly conserved elements discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers and can be transcribed during development. *Nucleic Acids Research*, **2013**, 41, 3600-18 | 20.1 | 20 |
| 162 | Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Research*, **2013**, 41, D171-6 | 20.1 | 212 |
| 161 | Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biology*, **2012**, 13, R48 | 18.3 | 194 |
| 160 | Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome Biology* , **2012**, 13, R49 | 18.3 | 71 |
| 159 | An integrated encyclopedia of DNA elements in the human genome. *Nature*, **2012**, 489, 57-74 | 50.4 | 11449 |
| 158 | Ensembl 2012. *Nucleic Acids Research*, **2012**, 40, D84-90 | 20.1 | 798 |
| 157 | Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, **2012**, 28, 1086-92 | 7.2 | 1129 |
| 156 | Modeling gene expression using chromatin features in various cellular contexts. *Genome Biology*, **2012**, 13, R53 | 18.3 | 182 |
| 155 | A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell*, **2012**, 148, 473-86 | 56.2 | 191 |
| 154 | The future of DNA sequence archiving. *GigaScience*, **2012**, 1, 2 | 7.6 | 19 |
| 153 | The making of ENCODE: Lessons for big-data projects. *Nature*, **2012**, 489, 49-51 | 50.4 | 77 |
| 152 | Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Research*, **2012**, 22, 1658-67 | 9.7 | 133 |
| 151 | The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **2012**, 484, 55-61 | 50.4 | 1187 |
| 150 | Paralogous annotation of disease-causing variants in long QT syndrome genes. *Human Mutation*, **2012**, 33, 1188-1191 | 4.7 | 26 |
| 149 | Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Research*, **2012**, 22, 9-24 | 9.7 | 86 |
| 148 | Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Research*, **2012**, 22, 1798-812 | 9.7 | 596 |
| 147 | Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic Acids Research*, **2012**, 40, D91-7 | 20.1 | 141 |

| | | | |
|---|---|---|---|
| 146 | Major submissions tool developments at the European Nucleotide Archive. *Nucleic Acids Research*, **2012**, 40, D43-7 | 20.1 | 27 |
| 145 | Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research*, **2011** , 39, D691-7 | 20.1 | 1090 |
| 144 | Considerations for the inclusion of 2x mammalian genomes in phylogenetic analyses. *Genome Biology*, **2011**, 12, 401 | 18.3 | 7 |
| 143 | Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, **2011**, 477, 289-94 | 50.4 | 1087 |
| 142 | A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biology*, **2011**, 9, e1001046 | 9.7 | 1060 |
| 141 | Strengths and Weaknesses of Selected Modeling Methods Used in Systems Biology **2011**, | | 2 |
| 140 | Assemblies: the good, the bad, the ugly. *Nature Methods*, **2011**, 8, 59-60 | 21.6 | 24 |
| 139 | Chromatin and heritability: how epigenetic studies can complement genetic approaches. *Trends in Genetics*, **2011**, 27, 172-6 | 8.5 | 18 |
| 138 | A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **2011**, 478, 476-82 | 50.4 | 802 |
| 137 | High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Research*, **2011**, 21, 456-64 | 9.7 | 250 |
| 136 | Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Research*, **2011**, 21, 734-40 | 9.7 | 233 |
| 135 | The European Nucleotide Archive. *Nucleic Acids Research*, **2011**, 39, D28-31 | 20.1 | 322 |
| 134 | Ensembl 2011. *Nucleic Acids Research*, **2011**, 39, D800-6 | 20.1 | 590 |
| 133 | RNAcentral: A vision for an international database of RNA sequences. *Rna*, **2011**, 17, 1941-6 | 5.8 | 54 |
| 132 | Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Research*, **2011**, 21, 1757-67 | 9.7 | 391 |
| 131 | A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*, **2010**, 463, 184-90 | 50.4 | 852 |
| 130 | International network of cancer genome projects. *Nature*, **2010**, 464, 993-8 | 50.4 | 1613 |
| 129 | The BioPAX community standard for pathway data sharing. *Nature Biotechnology*, **2010**, 28, 935-42 | 44.5 | 499 |

| | | | |
|---|---|---|---|
| 128 | An international bioinformatics infrastructure to underpin the Arabidopsis community. *Plant Cell*, **2010**, 22, 2530-6 | 11.6 | 22 |
| 127 | Improvements to services at the European Nucleotide Archive. *Nucleic Acids Research*, **2010**, 38, D39-45 | 20.1 | 47 |
| 126 | Ensembl's 10th year. *Nucleic Acids Research*, **2010**, 38, D557-62 | 20.1 | 240 |
| 125 | Allele-specific and heritable chromatin signatures in humans. *Human Molecular Genetics*, **2010**, 19, R204-9 | 5.6 | 24 |
| 124 | A new strategy for genome assembly using short sequence reads and reduced representation libraries. *Genome Research*, **2010**, 20, 249-56 | 9.7 | 24 |
| 123 | EMMA--mouse mutant resources for the international scientific community. *Nucleic Acids Research*, **2010**, 38, D570-6 | 20.1 | 38 |
| 122 | Finding and sharing: new approaches to registries of databases and services for the biomedical sciences. *Database: the Journal of Biological Databases and Curation*, **2010**, 2010, baq014 | 5 | 8 |
| 121 | The genome sequence of the spontaneously hypertensive rat: Analysis and functional significance. *Genome Research*, **2010**, 20, 791-803 | 9.7 | 77 |
| 120 | Evolutionary constraints of phosphorylation in eukaryotes, prokaryotes, and mitochondria. *Molecular and Cellular Proteomics*, **2010**, 9, 2642-53 | 7.6 | 70 |
| 119 | An effective model for natural selection in promoters. *Genome Research*, **2010**, 20, 685-92 | 9.7 | 22 |
| 118 | Ensembl Genome Browser **2010**, 923-939 | | 3 |
| 117 | The consequence of natural selection on genetic variation in the mouse. *Genomics*, **2010**, 95, 196-202 | 4.3 | 9 |
| 116 | The systematic annotation of the three main GPCR families in Reactome. *Database: the Journal of Biological Databases and Curation*, **2010**, 2010, baq018 | 5 | 17 |
| 115 | Genomic information infrastructure after the deluge. *Genome Biology*, **2010**, 11, 402 | 18.3 | 16 |
| 114 | Locus Reference Genomic sequences: an improved basis for describing human DNA variants. *Genome Medicine*, **2010**, 2, 24 | 14.4 | 86 |
| 113 | A draft sequence of the Neandertal genome. *Science*, **2010**, 328, 710-722 | 33.3 | 2599 |
| 112 | Heritable individual-specific and allele-specific chromatin signatures in humans. *Science*, **2010**, 328, 235-9 | 33.3 | 260 |
| 111 | A database and API for variation, dense genotyping and resequencing data. *BMC Bioinformatics*, **2010**, 11, 238 | 3.6 | 26 |

| | | | |
|---|---|---|---|
| 110 | Ensembl variation resources. *BMC Genomics*, **2010**, 11, 293 | 4.5 | 104 |
| 109 | Pebble and rock band: heuristic resolution of repeats and scaffolding in the velvet short-read de novo assembler. *PLoS ONE*, **2009**, 4, e8407 | 3.7 | 151 |
| 108 | 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 |
| 107 | VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Research*, **2009**, 37, D583-7 | 20.1 | 200 |
| 106 | Petabyte-scale innovations at the European Nucleotide Archive. *Nucleic Acids Research*, **2009**, 37, D19-25 | 20.1 | 72 |
| 105 | MAPU 2.0: high-accuracy proteomes mapped to genomes. *Nucleic Acids Research*, **2009**, 37, D902-6 | 20.1 | 17 |
| 104 | Sequence progressive alignment, a framework for practical large-scale probabilistic consistency alignment. *Bioinformatics*, **2009**, 25, 295-301 | 7.2 | 42 |
| 103 | Prepublication data sharing. *Nature*, **2009**, 461, 168-70 | 50.4 | 197 |
| 102 | Sense from sequence reads: methods for alignment and assembly. *Nature Methods*, **2009**, 6, S6-S12 | 21.6 | 261 |
| 101 | Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, **2009**, 4, 1184-91 | 18.8 | 1626 |
| 100 | Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research*, **2009**, 37, D619-22 | 20.1 | 675 |
| 99 | ENFIN--A European network for integrative systems biology. *Comptes Rendus - Biologies*, **2009**, 332, 1050-8 | 1.4 | 6 |
| 98 | EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, **2009**, 19, 327-35 | 9.7 | 836 |
| 97 | Visualising the Epigenome **2009**, 55-66 | | |
| 96 | Genome analysis of the platypus reveals unique signatures of evolution. *Nature*, **2008**, 453, 175-83 | 50.4 | 545 |
| 95 | A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, **2008**, 26, 779-85 | 44.5 | 533 |
| 94 | SNP and haplotype mapping for genetic analysis in the rat. *Nature Genetics*, **2008**, 40, 560-6 | 36.3 | 150 |
| 93 | What everybody should know about the rat genome and its online resources. *Nature Genetics*, **2008** , 40, 523-7 | 36.3 | 37 |

| | | | |
|---|---|---|---|
| 92 | Approaches to comparative sequence analysis: towards a functional view of vertebrate genomes. *Nature Reviews Genetics*, **2008**, 9, 303-13 | 30.1 | 42 |
| 91 | Integrating biological data--the Distributed Annotation System. *BMC Bioinformatics*, **2008**, 9 Suppl 8, S3 | 3.6 | 77 |
| 90 | Confounding between recombination and selection, and the Ped/Pop method for detecting selection. *Genome Research*, **2008**, 18, 1304-13 | 9.7 | 52 |
| 89 | Advanced genomic data mining. *PLoS Computational Biology*, **2008**, 4, e1000121 | 5 | 11 |
| 88 | Arabidopsis reactome: a foundation knowledgebase for plant systems biology. *Plant Cell*, **2008**, 20, 1426-36 | 11.6 | 44 |
| 87 | Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Research*, **2008**, 18, 1829-43 | 9.7 | 136 |
| 86 | Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Research*, **2008**, 18, 1814-28 | 9.7 | 218 |
| 85 | Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **2008**, 18, 821-9 | 9.7 | 7156 |
| 84 | An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Research*, **2008**, 18, 1518-29 | 9.7 | 304 |
| 83 | The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Research*, **2008**, 36, D445-8 | 20.1 | 163 |
| 82 | Priorities for nucleotide trace, sequence and annotation data capture at the Ensembl Trace Archive and the EMBL Nucleotide Sequence Database. *Nucleic Acids Research*, **2008**, 36, D5-12 | 20.1 | 39 |
| 81 | ENFIN - An Integrative Structure for Systems Biology. *Lecture Notes in Computer Science*, **2008**, 132-143 | 0.9 | |
| 80 | Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **2007**, 316, 222-34 | 33.3 | 1072 |
| 79 | Genome sequence of Aedes aegypti, a major arbovirus vector. *Science*, **2007**, 316, 1718-23 | 33.3 | 867 |
| 78 | Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Research*, **2007**, 17, 760-74 | 9.7 | 163 |
| 77 | Challenges and standards in integrating surveys of structural variation. *Nature Genetics*, **2007**, 39, S7-15 | 36.3 | 279 |
| 76 | Trawler: de novo regulatory motif discovery pipeline for chromatin immunoprecipitation. *Nature Methods*, **2007**, 4, 563-5 | 21.6 | 66 |
| 75 | Patterns of somatic mutation in human cancer genomes. *Nature*, **2007**, 446, 153-8 | 50.4 | 2400 |

| | | | |
|---|---|---|---|
| 74 | Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **2007**, 447, 799-816 | 50.4 | 4121 |
| 73 | Optimising oligonucleotide array design for ChIP-on-chip. *BMC Bioinformatics*, **2007**, 8, | 3.6 | 1 |
| 72 | ENFIN a network to enhance integrative systems biology. *Annals of the New York Academy of Sciences*, **2007**, 1115, 23-31 | 6.5 | 8 |
| 71 | Estimating the neutral rate of nucleotide substitution using introns. *Molecular Biology and Evolution*, **2007**, 24, 522-31 | 8.3 | 25 |
| 70 | The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Research*, **2007**, 17, 691-707 | 9.7 | 315 |
| 69 | Genome browsing with Ensembl: a practical overview. *Briefings in Functional Genomics & Proteomics* , **2007**, 6, 202-19 | | 27 |
| 68 | In vivo validation of a computationally predicted conserved Ath5 target gene set. *PLoS Genetics*, **2007**, 3, 1661-71 | 6 | 38 |
| 67 | VectorBase: a home for invertebrate vectors of human pathogens. *Nucleic Acids Research*, **2007**, 35, D503-5 | 10.1 | 97 |
| 66 | Identification of novel peptide hormones in the human proteome by hidden Markov model screening. *Genome Research*, **2007**, 17, 320-7 | 9.7 | 170 |
| 65 | The implications of alternative splicing in the ENCODE protein complement. *Proceedings of the National Academy of Sciences of the United States of America*, **2007**, 104, 5495-500 | 11.5 | 177 |
| 64 | Optimized design and assessment of whole genome tiling arrays. *Bioinformatics*, **2007**, 23, i195-204 | 7.2 | 46 |
| 63 | Reactome: An integrated expert model of human molecular processes and access toolkit. *Journal of Integrative Bioinformatics*, **2007**, 4, 286-296 | 3.8 | |
| 62 | Reactome: a knowledge base of biologic pathways and processes. *Genome Biology*, **2007**, 8, R39 | 18.3 | 454 |
| 61 | Update of the Anopheles gambiae PEST genome assembly. *Genome Biology*, **2007**, 8, R5 | 18.3 | 82 |
| 60 | Dry work in a wet world: computation in systems biology. *Molecular Systems Biology*, **2006**, 2, 40 | 12.2 | 14 |
| 59 | EGASP: the human ENCODE Genome Annotation Assessment Project. *Genome Biology*, **2006**, 7 Suppl 1, S2.1-31 | 18.3 | 187 |
| 58 | Picking pyknons out of the human genome. *Cell*, **2006**, 125, 836-8 | 56.2 | 9 |
| 57 | TranscriptSNPView: a genome-wide catalog of mouse coding variation. *Nature Genetics*, **2006**, 38, 853 | 36.3 | 14 |

| | | | |
|---|---|---|---|
| 56 | The discovery, positioning and verification of a set of transcription-associated motifs in vertebrates. *Genome Biology*, **2005**, 6, R104 | 18.3 | 38 |
| 55 | Gene finding in the chicken genome. *BMC Bioinformatics*, **2005**, 6, 131 | 3.6 | 26 |
| 54 | Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **2005**, 6, 31 | 3.6 | 1485 |
| 53 | A survey of homozygous deletions in human cancer genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **2005**, 102, 4542-7 | 11.5 | 70 |
| 52 | Transcriptome analysis for the chicken based on 19,626 finished cDNA sequences and 485,337 expressed sequence tags. *Genome Research*, **2005**, 15, 174-83 | 9.7 | 71 |
| 51 | Sockeye: a 3D environment for comparative genomics. *Genome Research*, **2004**, 14, 956-62 | 9.7 | 22 |
| 50 | Biological database design and implementation. *Briefings in Bioinformatics*, **2004**, 5, 31-8 | 13.4 | 11 |
| 49 | The Ensembl core software libraries. *Genome Research*, **2004**, 14, 929-33 | 9.7 | 100 |
| 48 | A SNP map of the rat genome generated from cDNA sequences. *Science*, **2004**, 303, 807 | 33.3 | 44 |
| 47 | An overview of Ensembl. *Genome Research*, **2004**, 14, 925-8 | 9.7 | 316 |
| 46 | Comparison of human chromosome 21 conserved nongenic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment. *Genome Research*, **2004**, 14, 852-9 | 9.7 | 55 |
| 45 | GeneWise and Genomewise. *Genome Research*, **2004**, 14, 988-95 | 9.7 | 1467 |
| 44 | Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **2004** , 428, 493-521 | 50.4 | 1689 |
| 43 | Genome information resources - developments at Ensembl. *Trends in Genetics*, **2004**, 20, 268-72 | 8.5 | 55 |
| 42 | The Anopheles gambiae genome: an update. *Trends in Parasitology*, **2004**, 20, 49-52 | 6.4 | 56 |
| 41 | The International Protein Index: an integrated database for proteomics experiments. *Proteomics*, **2004**, 4, 1985-8 | 4.8 | 650 |
| 40 | EnsMart: a generic system for fast and flexible access to biological data. *Genome Research*, **2004**, 14, 160-9 | 9.7 | 306 |
| 39 | Unrestricted free access works and must continue. *Nature*, **2003**, 422, 801 | 50.4 | 5 |

## (2000-2003)

| | | | |
|---|---|---|---|
| 38 | Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nature Reviews Genetics*, **2003**, 4, 251-62 | 30.1 | 172 |
| 37 | Discovering novel cis-regulatory motifs using functional networks. *Genome Research*, **2003**, 13, 883-95 | 9.7 | 18 |
| 36 | The European Bioinformatics Institute's data resources. *Nucleic Acids Research*, **2003**, 31, 43-50 | 20.1 | 38 |
| 35 | Genome annotation techniques: new approaches and challenges. *Drug Discovery Today*, **2002**, 7, S70-6 | 8.8 | 21 |
| 34 | A physical map of the mouse genome. *Nature*, **2002**, 418, 743-50 | 50.4 | 282 |
| 33 | Initial sequencing and comparative analysis of the mouse genome. *Nature*, **2002**, 420, 520-62 | 50.4 | 5376 |
| 32 | Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. *Science*, **2002**, 298, 149-59 | 33.3 | 455 |
| 31 | Databases and tools for browsing genomes. *Annual Review of Genomics and Human Genetics*, **2002**, 3, 293-310 | 9.7 | 17 |
| 30 | The Bioperl toolkit: Perl modules for the life sciences. *Genome Research*, **2002**, 12, 1611-8 | 9.7 | 1234 |
| 29 | The genome sequence of the malaria mosquito Anopheles gambiae. *Science*, **2002**, 298, 129-49 | 33.3 | 1622 |
| 28 | The Pfam protein families database. *Nucleic Acids Research*, **2002**, 30, 276-80 | 20.1 | 1839 |
| 27 | Immunity-related genes and gene families in Anopheles gambiae. *Science*, **2002**, 298, 159-65 | 33.3 | 743 |
| 26 | Progress in sequencing the mouse genome. *Genesis*, **2001**, 31, 137-41 | 1.9 | 17 |
| 25 | Mining the draft human genome. *Nature*, **2001**, 409, 827-8 | 50.4 | 51 |
| 24 | Cancer and genomics. *Nature*, **2001**, 409, 850-2 | 50.4 | 108 |
| 23 | Initial sequencing and analysis of the human genome. *Nature*, **2001**, 409, 860-921 | 50.4 | 17366 |
| 22 | Searching databases to find protein domain organization. *Advances in Protein Chemistry*, **2000**, 54, 137-57 | | 15 |
| 21 | Open annotation offers a democratic solution to genome sequencing. *Nature*, **2000**, 403, 825 | 50.4 | 26 |

| | | | |
|---|---|---|---|
| 20 | Comparative genomics of the eukaryotes. *Science*, **2000**, 287, 2204-15 | 33.3 | 1364 |
| 19 | Using GeneWise in the Drosophila annotation experiment. *Genome Research*, **2000**, 10, 547-8 | 9.7 | 263 |
| 18 | The Pfam protein families database. *Nucleic Acids Research*, **2000**, 28, 263-6 | 20.1 | 1074 |
| 17 | The Bioperl project. *ACM SIGBIO Newsletter*, **2000**, 20, 13-14 | | 1 |
| 16 | Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Research*, **1999**, 9, 815-24 | 9.7 | 141 |
| 15 | PH domain: the first anniversary. *Trends in Biochemical Sciences*, **1994**, 19, 349-53 | 10.3 | 306 |
| 14 | Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Research*, **1993**, 21, 5803-16 | 20.1 | 610 |
| 13 | A putative homolog of U2AF65 in S. cerevisiae. *Nucleic Acids Research*, **1992**, 20, 4663 | 20.1 | 10 |
| 12 | FORGE: A tool to discover cell specific enrichments of GWAS associated SNPs in regulatory regions. *F1000Research*,4, 18 | 3.6 | 13 |
| 11 | FORGE : A tool to discover cell specific enrichments of GWAS associated SNPs in regulatory regions. | | 5 |
| 10 | GARFIELD - GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction | | 16 |
| 9 | A large cross-ancestry meta-analysis of genome-wide association studies identifies 69 novel risk loci for primary open-angle glaucoma and includes a genetic link with Alzheimer’s disease | | 4 |
| 8 | Genomics in healthcare: GA4GH looks to 2022 | | 38 |
| 7 | LiMMBo: a simple, scalable approach for linear mixed models in high-dimensional genetic association studies | | 2 |
| 6 | Genomic analysis reveals a functional role for myocardial trabeculae in adults | | 1 |
| 5 | RNA modifications detection by comparative Nanopore direct RNA sequencing | | 33 |
| 4 | Visualization of loop extrusion by DNA nanoscale tracing in single human cells | | 4 |
| 3 | Genomic variations and epigenomic landscape of the Medaka Inbred Kiyosu-Karlsruhe (MIKK) panel | | 3 |

# LIST OF PUBLICATIONS