Serafim Batzoglou

List of Publications by Year
in descending order

Source: https://exaly.com/author-pdf/10795203/publications.pdf

Version: 2024-02-01

87	48,843	36303	49909
papers	citations	51 h-index	87 g-index

96	96	96	58650
all docs	docs citations	times ranked	citing authors

#	Article	IF	Citations
1	Initial sequencing and analysis of the human genome. Nature, 2001, 409, 860-921.	27.8	21,074
2	Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 2007, 447, 799-816.	27.8	4,709
3	Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature, 2004, 428, 493-521.	27.8	1,943
4	Evolution of genes and genomes on the Drosophila phylogeny. Nature, 2007, 450, 203-218.	27.8	1,886
5	ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research, 2012, 22, 1813-1831.	5.5	1,708
6	Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLoS Computational Biology, 2010, 6, e1001025.	3.2	1,443
7	Architecture of the human regulatory network derived from ENCODE data. Nature, 2012, 489, 91-100.	27.8	1,384
8	Predicting Splicing from Primary Sequence with Deep Learning. Cell, 2019, 176, 535-548.e24.	28.9	1,305
9	Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature, 2005, 438, 1105-1115.	27.8	1,250
10	Distribution and intensity of constraint in mammalian genomic sequence. Genome Research, 2005, 15, 901-913.	5.5	1,230
11	ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Research, 2005, 15, 330-340.	5.5	982
12	LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Research, 2003, 13, 721-731.	5.5	960
13	Linking disease associations with regulatory information in the human genome. Genome Research, 2012, 22, 1748-1759.	5.5	657
14	Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nature Methods, 2008, 5, 829-834.	19.0	627
15	Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nature Methods, 2017, 14, 414-416.	19.0	577
16	ARACHNE: A Whole-Genome Shotgun Assembler. Genome Research, 2002, 12, 177-189.	5.5	508
17	Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell, 2018, 174, 758-769.e9.	28.9	459
18	CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics, 2006, 22, e90-e98.	4.1	458

#	Article	IF	Citations
19	Multiple sequence alignment. Current Opinion in Structural Biology, 2006, 16, 368-373.	5.7	353
20	Extensive Variation in Chromatin States Across Humans. Science, 2013, 342, 750-752.	12.6	338
21	Glocal alignment: finding rearrangements during alignment. Bioinformatics, 2003, 19, i54-i62.	4.1	333
22	Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112, E1116-25.	7.1	307
23	Predicting the clinical impact of human mutation with deep neural networks. Nature Genetics, 2018, 50, 1161-1170.	21.4	288
24	Graemlin: General and robust alignment of multiple large interaction networks. Genome Research, 2006, 16, 1169-1181.	5.5	274
25	The DNA methylation landscape of advanced prostate cancer. Nature Genetics, 2020, 52, 778-789.	21.4	198
26	Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. Genome Research, 2007, 17, 760-774.	5.5	184
27	Fast and scalable inference of multi-sample cancer lineages. Genome Biology, 2015, 16, 91.	8.8	180
28	Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. Genome Research, 2012, 22, 1735-1747.	5.5	168
29	Autoimmune Disease Classification by Inverse Association with SNP Alleles. PLoS Genetics, 2009, 5, e1000792.	3.5	151
30	Fast and sensitive multiple alignment of large genomic sequences. BMC Bioinformatics, 2003, 4, 66.	2.6	134
31	Eukaryotic Regulatory Element Conservation Analysis and Identification Using Comparative Genomics. Genome Research, 2004, 14, 451-458.	5.5	130
32	Characterization of Evolutionary Rates and Constraints in Three Mammalian Genomes. Genome Research, 2004, 14, 539-548.	5.5	125
33	Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. Nature Communications, 2018, 9, 4453.	12.8	123
34	Analysis of Cell Fate from Single-Cell Gene Expression Profiles in C. elegans. Cell, 2009, 139, 623-633.	28.9	122
35	MotifCut: regulatory motifs finding with maximum density subgraphs. Bioinformatics, 2006, 22, e150-e157.	4.1	119
36	Whole-Genome Sequencing and Assembly with High-Throughput, Short-Read Technologies. PLoS ONE, 2007, 2, e484.	2.5	112

#	Article	IF	Citations
37	The many faces of sequence alignment. Briefings in Bioinformatics, 2005, 6, 6-22.	6.5	107
38	Quantitative Estimates of Sequence Divergence for Comparative Analyses of Mammalian Genomes. Genome Research, 2003, 13, 813-820.	5.5	106
39	High-quality genome sequences of uncultured microbes by assembly of read clouds. Nature Biotechnology, 2018, 36, 1067-1075.	17.5	103
40	Genome evolution during progression to breast cancer. Genome Research, 2013, 23, 1097-1108.	5.5	98
41	Genome-wide reconstruction of complex structural variants using read clouds. Nature Methods, 2017, 14, 915-920.	19.0	96
42	Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome. Nature Biotechnology, 2016, 34, 64-69.	17.5	93
43	Network enhancement as a general method to denoise weighted biological networks. Nature Communications, 2018, 9, 3108.	12.8	82
44	SIMLR: A Tool for Largeâ€Scale Genomic Analyses by Multiâ€Kernel Learning. Proteomics, 2018, 18, 1700232.	2.2	81
45	Automated Whole-Genome Multiple Alignment of Rat, Mouse, and Human. Genome Research, 2004, 14, 685-692.	5.5	79
46	CONTRAST: a discriminative, phylogeny-free approach to multiple informant de novo gene prediction. Genome Biology, 2007, 8, R269.	9.6	79
47	Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. Genome Research, 2010, 20, 301-310.	5.5	77
48	A max-margin model for efficient simultaneous alignment and folding of RNA sequences. Bioinformatics, 2008, 24, i68-i76.	4.1	76
49	Read clouds uncover variation in complex regions of the human genome. Genome Research, 2015, 25, 1570-1580.	5.5	70
50	Effect of genetic divergence in identifying ancestral origin using HAPAA. Genome Research, 2008, 18, 676-682.	5.5	66
51	Automatic Parameter Learning for Multiple Local Network Alignment. Journal of Computational Biology, 2009, 16, 1001-1022.	1.6	66
52	Automatic Parameter Learning for Multiple Network Alignment. Lecture Notes in Computer Science, 2008, , 214-231.	1.3	58
53	Phylo-VISTA: interactive visualization of multiple DNA sequence alignments. Bioinformatics, 2004, 20, 636-643.	4.1	46
54	Current progress in network research: toward reference networks for key model organisms. Briefings in Bioinformatics, 2007, 8, 318-332.	6.5	46

#	Article	IF	Citations
55	Inference of Tumor Phylogenies with Improved Somatic Mutation Discovery. Journal of Computational Biology, 2013, 20, 933-944.	1.6	45
56	Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. Genome Medicine, 2020, 12, 50.	8.2	43
57	Multiple alignment of protein sequences with repeats and rearrangements. Nucleic Acids Research, 2006, 34, 5932-5942.	14.5	40
58	Genome assembly from synthetic long read clouds. Bioinformatics, 2016, 32, i216-i224.	4.1	33
59	A suite of web-based programs to search for transcriptional regulatory motifs. Nucleic Acids Research, 2004, 32, W204-W207.	14.5	30
60	Genetic and Computational Identification of a Conserved Bacterial Metabolic Module. PLoS Genetics, 2008, 4, e1000310.	3.5	26
61	A hybrid cloud read aligner based on MinHash and kmer voting that preserves privacy. Nature Communications, 2017, 8, 15311.	12.8	24
62	Integrated Protein Interaction Networks for 11 Microbes. Lecture Notes in Computer Science, 2006, , 1-14.	1.3	24
63	A Dictionary-Based Approach for Gene Annotation. Journal of Computational Biology, 1999, 6, 419-430.	1.6	23
64	Identification of Promoter Regions in the Human Genome by Using a Retroviral Plasmid Library-Based Functional Reporter Gene Assay. Genome Research, 2003, 13, 1765-1774.	5.5	21
65	A machine-compiled database of genome-wide association studies. Nature Communications, 2019, 10, 3341.	12.8	21
66	A graph-based motif detection algorithm models complex nucleotide dependencies in transcription factor binding sites. Nucleic Acids Research, 2006, 34, 5730-5739.	14.5	19
67	Parente2: a fast and accurate method for detecting identity by descent. Genome Research, 2015, 25, 280-289.	5.5	18
68	A Classifier-based approach to identify genetic similarities between diseases. Bioinformatics, 2009, 25, i21-i29.	4.1	17
69	Cell-lineage heterogeneity and driver mutation recurrence in pre-invasive breast neoplasia. Genome Medicine, 2015, 7, 28.	8.2	17
70	Reconstruction of genealogical relationships with applications to Phase III of HapMap. Bioinformatics, 2011, 27, i333-i341.	4.1	16
71	Vicus: Exploiting local structures to improve network-based analysis of biological data. PLoS Computational Biology, 2017, 13, e1005621.	3.2	15
72	Ancestry Inference in Complex Admixtures via Variable-length Markov Chain Linkage Models. Journal of Computational Biology, 2013, 20, 199-211.	1.6	12

#	Article	IF	Citations
73	Automated cellular annotation for high-resolution images of adult Caenorhabditis elegans. Bioinformatics, 2013, 29, i18-i26.	4.1	11
74	HAPDeNovo: a haplotype-based approach for filtering and phasing de novo mutations in linked read sequencing data. BMC Genomics, 2018, 19, 467.	2.8	11
75	Discovery of widespread transcription initiation at microsatellites predictable by sequence-based deep neural network. Nature Communications, 2021, 12, 3297.	12.8	11
76	Meltos: multi-sample tumor phylogeny reconstruction for structural variants. Bioinformatics, 2020, 36, 1082-1090.	4.1	9
77	Algorithmic challenges in mammalian whole-genome assembly. , 2005, , .		8
78	Using multiple alignments to improve seeded local alignment algorithms. Nucleic Acids Research, 2005, 33, 4563-4577.	14.5	8
79	Reveel: large-scale population genotyping using low-coverage sequencing data. Bioinformatics, 2016, 32, 1686-1696.	4.1	7
80	An Effective Filter for IBD Detection in Large Data Sets. PLoS ONE, 2014, 9, e92713.	2.5	6
81	Chromosome-level de novo assembly of the pig-tailed macaque genome using linked-read sequencing and HiC proximity scaffolding. GigaScience, 2020, 9, .	6.4	6
82	Fast Metagenomic Binning via Hashing and Bayesian Clustering. Journal of Computational Biology, 2018, 25, 677-688.	1.6	5
83	Ancestry Inference in Complex Admixtures via Variable-Length Markov Chain Linkage Models. Lecture Notes in Computer Science, 2012, , 12-28.	1.3	5
84	Chaining Algorithms for Alignment of Draft Sequence. Lecture Notes in Computer Science, 2004, , 326-337.	1.3	4
85	Current Progress in Static and Dynamic Modeling of Biological Networks. Systems Biology, 2010, , 13-73.	0.1	4
86	Constraint and divergence of global gene expression in the mammalian embryo. ELife, 2015, 4, e05538.	6.0	3
87	Session Introduction. , 2002, , .		0