Serafim Batzoglou

List of Publications by Year
in descending order

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

Version: 2024-02-01

87 papers	48,843 citations	41627 51 h-index	56606 87 g-index
96 all docs	96 docs citations	96 times ranked	65632 citing authors

#	Article	IF	Citations
1	Discovery of widespread transcription initiation at microsatellites predictable by sequence-based deep neural network. Nature Communications, 2021, 12, 3297.	5.8	11
2	Meltos: multi-sample tumor phylogeny reconstruction for structural variants. Bioinformatics, 2020, 36, 1082-1090.	1.8	9
3	The DNA methylation landscape of advanced prostate cancer. Nature Genetics, 2020, 52, 778-789.	9.4	198
4	Chromosome-level de novo assembly of the pig-tailed macaque genome using linked-read sequencing and HiC proximity scaffolding. GigaScience, 2020, 9, .	3.3	6
5	Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. Genome Medicine, 2020, 12, 50.	3.6	43
6	A machine-compiled database of genome-wide association studies. Nature Communications, 2019, 10, 3341.	5.8	21
7	Predicting Splicing from Primary Sequence with Deep Learning. Cell, 2019, 176, 535-548.e24.	13.5	1,305
8	Fast Metagenomic Binning via Hashing and Bayesian Clustering. Journal of Computational Biology, 2018, 25, 677-688.	0.8	5
9	SIMLR: A Tool for Largeâ€Scale Genomic Analyses by Multiâ€Kernel Learning. Proteomics, 2018, 18, 1700232.	1.3	81
10	Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. Nature Communications, 2018, 9, 4453.	5.8	123
11	High-quality genome sequences of uncultured microbes by assembly of read clouds. Nature Biotechnology, 2018, 36, 1067-1075.	9.4	103
12	HAPDeNovo: a haplotype-based approach for filtering and phasing de novo mutations in linked read sequencing data. BMC Genomics, 2018, 19, 467.	1.2	11
13	Predicting the clinical impact of human mutation with deep neural networks. Nature Genetics, 2018, 50, 1161-1170.	9.4	288
14	Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell, 2018, 174, 758-769.e9.	13.5	459
15	Network enhancement as a general method to denoise weighted biological networks. Nature Communications, 2018, 9, 3108.	5.8	82
16	Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nature Methods, 2017, 14, 414-416.	9.0	577
17	Genome-wide reconstruction of complex structural variants using read clouds. Nature Methods, 2017, 14, 915-920.	9.0	96
18	A hybrid cloud read aligner based on MinHash and kmer voting that preserves privacy. Nature Communications, 2017, 8, 15311.	5.8	24

#	Article	IF	Citations
19	Vicus: Exploiting local structures to improve network-based analysis of biological data. PLoS Computational Biology, 2017, 13, e1005621.	1.5	15
20	Genome assembly from synthetic long read clouds. Bioinformatics, 2016, 32, i216-i224.	1.8	33
21	Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome. Nature Biotechnology, 2016, 34, 64-69.	9.4	93
22	Reveel: large-scale population genotyping using low-coverage sequencing data. Bioinformatics, 2016, 32, 1686-1696.	1.8	7
23	Fast and scalable inference of multi-sample cancer lineages. Genome Biology, 2015, 16, 91.	3.8	180
24	Cell-lineage heterogeneity and driver mutation recurrence in pre-invasive breast neoplasia. Genome Medicine, 2015, 7, 28.	3.6	17
25	Parente2: a fast and accurate method for detecting identity by descent. Genome Research, 2015, 25, 280-289.	2.4	18
26	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.	3.3	307
27	Read clouds uncover variation in complex regions of the human genome. Genome Research, 2015, 25, 1570-1580.	2.4	70
28	Constraint and divergence of global gene expression in the mammalian embryo. ELife, 2015, 4, e05538.	2.8	3
29	An Effective Filter for IBD Detection in Large Data Sets. PLoS ONE, 2014, 9, e92713.	1.1	6
30	Extensive Variation in Chromatin States Across Humans. Science, 2013, 342, 750-752.	6.0	338
31	Automated cellular annotation for high-resolution images of adult Caenorhabditis elegans. Bioinformatics, 2013, 29, i18-i26.	1.8	11
32	Ancestry Inference in Complex Admixtures via Variable-length Markov Chain Linkage Models. Journal of Computational Biology, 2013, 20, 199-211.	0.8	12
33	Inference of Tumor Phylogenies with Improved Somatic Mutation Discovery. Journal of Computational Biology, 2013, 20, 933-944.	0.8	45
34	Genome evolution during progression to breast cancer. Genome Research, 2013, 23, 1097-1108.	2.4	98
35	Linking disease associations with regulatory information in the human genome. Genome Research, 2012, 22, 1748-1759.	2.4	657
36	ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research, 2012, 22, 1813-1831.	2.4	1,708

#	Article	IF	Citations
37	Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. Genome Research, 2012, 22, 1735-1747.	2.4	168
38	Architecture of the human regulatory network derived from ENCODE data. Nature, 2012, 489, 91-100.	13.7	1,384
39	Ancestry Inference in Complex Admixtures via Variable-Length Markov Chain Linkage Models. Lecture Notes in Computer Science, 2012, , 12-28.	1.0	5
40	Reconstruction of genealogical relationships with applications to Phase III of HapMap. Bioinformatics, 2011, 27, i333-i341.	1.8	16
41	Current Progress in Static and Dynamic Modeling of Biological Networks. Systems Biology, 2010, , 13-73.	0.1	4
42	Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. Genome Research, 2010, 20, 301-310.	2.4	77
43	Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLoS Computational Biology, 2010, 6, e1001025.	1.5	1,443
44	Automatic Parameter Learning for Multiple Local Network Alignment. Journal of Computational Biology, 2009, 16, 1001-1022.	0.8	66
45	Autoimmune Disease Classification by Inverse Association with SNP Alleles. PLoS Genetics, 2009, 5, e1000792.	1.5	151
46	A Classifier-based approach to identify genetic similarities between diseases. Bioinformatics, 2009, 25, i21-i29.	1.8	17
47	Analysis of Cell Fate from Single-Cell Gene Expression Profiles in C. elegans. Cell, 2009, 139, 623-633.	13.5	122
48	Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nature Methods, 2008, 5, 829-834.	9.0	627
49	Effect of genetic divergence in identifying ancestral origin using HAPAA. Genome Research, 2008, 18, 676-682.	2.4	66
50	Genetic and Computational Identification of a Conserved Bacterial Metabolic Module. PLoS Genetics, 2008, 4, e1000310.	1.5	26
51	A max-margin model for efficient simultaneous alignment and folding of RNA sequences. Bioinformatics, 2008, 24, i68-i76.	1.8	76
52	Automatic Parameter Learning for Multiple Network Alignment. Lecture Notes in Computer Science, 2008, , 214-231.	1.0	58
53	Current progress in network research: toward reference networks for key model organisms. Briefings in Bioinformatics, 2007, 8, 318-332.	3.2	46
54	CONTRAST: a discriminative, phylogeny-free approach to multiple informant de novo gene prediction. Genome Biology, 2007, 8, R269.	13.9	79

#	Article	IF	Citations
55	Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. Genome Research, 2007, 17, 760-774.	2.4	184
56	Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 2007, 447, 799-816.	13.7	4,709
57	Evolution of genes and genomes on the Drosophila phylogeny. Nature, 2007, 450, 203-218.	13.7	1,886
58	Whole-Genome Sequencing and Assembly with High-Throughput, Short-Read Technologies. PLoS ONE, 2007, 2, e484.	1.1	112
59	Graemlin: General and robust alignment of multiple large interaction networks. Genome Research, 2006, 16, 1169-1181.	2.4	274
60	Multiple sequence alignment. Current Opinion in Structural Biology, 2006, 16, 368-373.	2.6	353
61	Multiple alignment of protein sequences with repeats and rearrangements. Nucleic Acids Research, 2006, 34, 5932-5942.	6.5	40
62	A graph-based motif detection algorithm models complex nucleotide dependencies in transcription factor binding sites. Nucleic Acids Research, 2006, 34, 5730-5739.	6.5	19
63	MotifCut: regulatory motifs finding with maximum density subgraphs. Bioinformatics, 2006, 22, e150-e157.	1.8	119
64	CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics, 2006, 22, e90-e98.	1.8	458
65	Integrated Protein Interaction Networks for 11 Microbes. Lecture Notes in Computer Science, 2006, , 1-14.	1.0	24
66	Algorithmic challenges in mammalian whole-genome assembly. , 2005, , .		8
67	Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature, 2005, 438, 1105-1115.	13.7	1,250
68	Using multiple alignments to improve seeded local alignment algorithms. Nucleic Acids Research, 2005, 33, 4563-4577.	6.5	8
69	Distribution and intensity of constraint in mammalian genomic sequence. Genome Research, 2005, 15, 901-913.	2.4	1,230
70	The many faces of sequence alignment. Briefings in Bioinformatics, 2005, 6, 6-22.	3.2	107
71	ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Research, 2005, 15, 330-340.	2.4	982
72	Automated Whole-Genome Multiple Alignment of Rat, Mouse, and Human. Genome Research, 2004, 14, 685-692.	2.4	79

#	Article	IF	Citations
73	Eukaryotic Regulatory Element Conservation Analysis and Identification Using Comparative Genomics. Genome Research, 2004, 14, 451-458.	2.4	130
74	Characterization of Evolutionary Rates and Constraints in Three Mammalian Genomes. Genome Research, 2004, 14, 539-548.	2.4	125
75	A suite of web-based programs to search for transcriptional regulatory motifs. Nucleic Acids Research, 2004, 32, W204-W207.	6.5	30
76	Phylo-VISTA: interactive visualization of multiple DNA sequence alignments. Bioinformatics, 2004, 20, 636-643.	1.8	46
77	Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature, 2004, 428, 493-521.	13.7	1,943
78	Chaining Algorithms for Alignment of Draft Sequence. Lecture Notes in Computer Science, 2004, , 326-337.	1.0	4
79	Fast and sensitive multiple alignment of large genomic sequences. BMC Bioinformatics, 2003, 4, 66.	1.2	134
80	Glocal alignment: finding rearrangements during alignment. Bioinformatics, 2003, 19, i54-i62.	1.8	333
81	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.	2.4	21
82	Quantitative Estimates of Sequence Divergence for Comparative Analyses of Mammalian Genomes. Genome Research, 2003, 13, 813-820.	2.4	106
83	LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Research, 2003, 13, 721-731.	2.4	960
84	ARACHNE: A Whole-Genome Shotgun Assembler. Genome Research, 2002, 12, 177-189.	2.4	508
85	Session Introduction. , 2002, , .		0
86	Initial sequencing and analysis of the human genome. Nature, 2001, 409, 860-921.	13.7	21,074
87	A Dictionary-Based Approach for Gene Annotation. Journal of Computational Biology, 1999, 6, 419-430.	0.8	23