Kevin Y Yip

List of Publications by Year
in descending order

112 papers	24,828 citations	66343 42 h-index	31849 101 g-index
146 all docs	146 docs citations	146 times ranked	45079 citing authors

#	Article	IF	Citations
1	An integrated encyclopedia of DNA elements in the human genome. Nature, 2012, 489, 57-74.	27.8	15,516
2	Architecture of the human regulatory network derived from ENCODE data. Nature, 2012, 489, 91-100.	27.8	1,384
3	A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biology, 2011, 9, e1001046.	5.6	1,257
4	Integrative Analysis of the <i>Caenorhabditis elegans</i> Genome by the modENCODE Project. Science, 2010, 330, 1775-1787.	12.6	912
5	The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell, 2016, 167, 1145-1149.	28.9	404
6	FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. Genome Biology, 2014, 15, 480.	8.8	291
7	Comparative analysis of the transcriptome across distant species. Nature, 2014, 512, 445-448.	27.8	289
8	Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. Genome Biology, 2012, 13, R48.	9.6	233
9	Exome and genome sequencing of nasopharynx cancer identifies NF-κB pathway activating mutations. Nature Communications, 2017, 8, 14121.	12.8	227
10	Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. Cell, 2010, 143, 639-650.	28.9	200
11	Reconstruction of enhancer–target networks in 935 samples of human primary cells, tissues and cell lines. Nature Genetics, 2017, 49, 1428-1436.	21.4	194
12	Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. Genome Biology, 2014, 15, 408.	8.8	173
13	Decision Trees for Uncertain Data. IEEE Transactions on Knowledge and Data Engineering, 2011, 23, 64-78.	5.7	167
14	Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Research, 2012, 22, 1658-1667.	5.5	166
15	Efficient Clustering of Uncertain Data. IEEE International Conference on Data Mining, 2006, , .	0.0	135
16	Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. Genetics, 2016, 202, 351-362.	2.9	126
17	Genome maps across 26 human populations reveal population-specific patterns of structural variation. Nature Communications, 2019, 10, 1025.	12.8	123
18	HARP: a practical projected clustering algorithm. IEEE Transactions on Knowledge and Data Engineering, 2004, 16, 1387-1397.	5.7	121

#	Article	IF	Citations
19	A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. Genome Biology, 2011, 12, R15.	9.6	118
20	Establishment and characterization of new tumor xenografts and cancer cell lines from EBV-positive nasopharyngeal carcinoma. Nature Communications, 2018, 9, 4663.	12.8	106
21	Sustained Antidiabetic Effects of a Berberine-Containing Chinese Herbal Medicine Through Regulation of Hepatic Gene Expression. Diabetes, 2012, 61, 933-943.	0.6	103
22	Improved Reconstruction of In Silico Gene Regulatory Networks by Integrating Knockout and Perturbation Data. PLoS ONE, 2010, 5, e8121.	2.5	97
23	An integrative ENCODE resource for cancer genomics. Nature Communications, 2020, 11, 3696.	12.8	95
24	Identification of a Major Determinant for Serine-Threonine Kinase Phosphoacceptor Specificity. Molecular Cell, 2014, 53, 140-147.	9.7	91
25	YeastHub: a semantic web use case for integrating data in the life sciences domain. Bioinformatics, 2005, 21, i85-i96.	4.1	88
26	Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. Nucleic Acids Research, 2015, 43, 7247-7259.	14.5	87
27	HCLS 2.0/3.0: Health care and life sciences data mashup using Web 2.0/3.0. Journal of Biomedical Informatics, 2008, 41, 694-705.	4.3	78
28	CLUSTAG: hierarchical clustering and graph methods for selecting tag SNPs. Bioinformatics, 2005, 21, 1735-1736.	4.1	77
29	Aberrant enhancer hypomethylation contributes to hepatic carcinogenesis through global transcriptional reprogramming. Nature Communications, 2019, 10, 335.	12.8	77
30	Mining periodic patterns with gap requirement from sequences. ACM Transactions on Knowledge Discovery From Data, 2007, 1, 7.	3.5	73
31	Machine learning and genome annotation: a match meant to be?. Genome Biology, 2013, 14, 205.	9.6	72
32	Supervised enhancer prediction with epigenetic pattern recognition and targeted validation. Nature Methods, 2020, 17, 807-814.	19.0	71
33	An integrated system for studying residue coevolution in proteins. Bioinformatics, 2008, 24, 290-292.	4.1	70
34	The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. Bioinformatics, 2006, 22, 2968-2970.	4.1	63
35	A common set of distinct features that characterize noncoding RNAs across multiple species. Nucleic Acids Research, 2015, 43, 104-114.	14.5	63
36	Identification of Genes Associated With Hirschsprung Disease, Based on Whole-Genome Sequence Analysis, and Potential Effects on Enteric Nervous System Development. Gastroenterology, 2018, 155, 1908-1922.e5.	1.3	61

#	Article	IF	Citations
37	Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data. Genome Research, 2011, 21, 276-285.	5.5	60
38	Comparing classical pathways and modern networks: towards the development of an edge ontology. Trends in Biochemical Sciences, 2007, 32, 320-331.	7.5	59
39	A selective HDAC8 inhibitor potentiates antitumor immunity and efficacy of immune checkpoint blockade in hepatocellular carcinoma. Science Translational Medicine, 2021, 13, .	12.4	59
40	Whole-genome profiling of nasopharyngeal carcinoma reveals viral-host co-operation in inflammatory NF-Î°B activation and immune escape. Nature Communications, 2021, 12, 4193.	12.8	56
41	Decision Trees for Uncertain Data. Proceedings - International Conference on Data Engineering, 2009, , .	0.0	52
42	Jellyfish genomes reveal distinct homeobox gene clusters and conservation of small RNA processing. Nature Communications, 2020, 11, 3051.	12.8	47
43	Yin Yang 1â€mediated epigenetic silencing of tumourâ€suppressive microRNAs activates nuclear factorâ€Î°B in hepatocellular carcinoma. Journal of Pathology, 2016, 238, 651-664.	4.5	46
44	EBVâ€encoded miRNAs target ATMâ€mediated response in nasopharyngeal carcinoma. Journal of Pathology, 2018, 244, 394-407.	4.5	44
45	Progressive skylining over Web-accessible databases. Data and Knowledge Engineering, 2006, 57, 122-147.	3.4	43
46	Identification of a recurrent transforming UBR5â€"ZNF423 fusion gene in EBV â€associated nasopharyngeal carcinoma. Journal of Pathology, 2013, 231, 158-167.	4.5	43
47	G9a Plays Distinct Roles in Maintaining DNA Methylation, Retrotransposon Silencing, and Chromatin Looping. Cell Reports, 2020, 33, 108315.	6.4	43
48	Mining periodic patterns with gap requirement from sequences. , 2005, , .		40
49	Analysis of membrane proteins in metagenomics: Networks of correlated environmental features and protein families. Genome Research, 2010, 20, 960-971.	5.5	40
50	Loss of tumor suppressor IGFBP4 drives epigenetic reprogramming in hepatic carcinogenesis. Nucleic Acids Research, 2018, 46, 8832-8847.	14.5	40
51	OMBlast: alignment tool for optical mapping using a seed-and-extend approach. Bioinformatics, 2017, 33, 311-319.	4.1	39
52	On Discovery of Extremely Low-Dimensional Clusters Using Semi-Supervised Projected Clustering. , 0, , .		38
53	Complete genomic sequence of Epstein-Barr virus in nasopharyngeal carcinoma cell line C666-1. Infectious Agents and Cancer, 2013, 8, 29.	2.6	37
54	ACT: aggregation and correlation toolbox for analyses of genome tracks. Bioinformatics, 2011, 27, 1152-1154.	4.1	35

#	Article	IF	Citations
55	A comprehensive web tool for toehold switch design. Bioinformatics, 2018, 34, 2862-2864.	4.1	31
56	Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication. Communications Biology, 2021, 4, 83.	4.4	31
57	Are special read alignment strategies necessary and cost-effective when handling sequencing reads from patient-derived tumor xenografts?. BMC Genomics, 2014, 15, 1172.	2.8	28
58	OMTools: a software package for visualizing and processing optical mapping data. Bioinformatics, 2017, 33, 2933-2935.	4.1	28
59	OMSV enables accurate and comprehensive identification of large structural variations from nanochannel-based single-molecule optical maps. Genome Biology, 2017, 18, 230.	8.8	28
60	LinkHub: a Semantic Web system that facilitates cross-database queries and information retrieval in proteomics. BMC Bioinformatics, 2007, 8, S5.	2.6	27
61	Training set expansion: an approach to improving the reconstruction of biological networks from limited and uneven reliable interactions. Bioinformatics, 2009, 25, 243-250.	4.1	25
62	Development of Grid-like Applications for Public Health Using Web 2.0 Mashup Techniques. Journal of the American Medical Informatics Association: JAMIA, 2008, 15, 783-786.	4.4	23
63	New guidelines for DNA methylome studies regarding 5-hydroxymethylcytosine for understanding transcriptional regulation. Genome Research, 2019, 29, 543-553.	5.5	21
64	A web services choreography scenario for interoperating bioinformatics applications. BMC Bioinformatics, 2004, 5, 25.	2.6	20
65	On mining micro-array data by Order-Preserving Submatrix. International Journal of Bioinformatics Research and Applications, 2007, 3, 42.	0.2	20
66	Interactome-transcriptome analysis discovers signatures complementary to GWAS Loci of Type 2 Diabetes. Scientific Reports, 2016, 6, 35228.	3.3	20
67	Sirtuin 7 super-enhancer drives epigenomic reprogramming in hepatocarcinogenesis. Cancer Letters, 2022, 525, 115-130.	7.2	19
68	The Essential Component in DNA-Based Information Storage System: Robust Error-Tolerating Module. Frontiers in Bioengineering and Biotechnology, 2014, 2, 49.	4.1	18
69	Identification of specificity determining residues in peptide recognition domains using an information theoretic approach applied to large-scale binding maps. BMC Biology, 2011, 9, 53.	3.8	16
70	Genome-wide analysis of chromatin features identifies histone modification sensitive and insensitive yeast transcription factors. Genome Biology, 2011, 12, R111.	9.6	16
71	Metric and trigonometric pruning for clustering of uncertain data in 2D geometric space. Information Systems, 2011, 36, 476-497.	3.6	13
72	Whole-genome analysis of noncoding genetic variations identifies multiscale regulatory element perturbations associated with Hirschsprung disease. Genome Research, 2020, 30, 1618-1632.	5.5	13

#	Article	IF	Citations
73	A network approach to exploring the functional basis of gene–gene epistatic interactions in disease susceptibility. Bioinformatics, 2018, 34, 1741-1749.	4.1	11
74	Identifying projected clusters from gene expression profiles. Journal of Biomedical Informatics, 2004, 37, 345-357.	4.3	10
75	Systematic exploration of autonomous modules in noisy microRNA-target networks for testing the generality of the ceRNA hypothesis. BMC Genomics, 2014, 15, 1178.	2.8	10
76	Identification and characterization of a novel Epstein-Barr Virus-encoded circular RNA from LMP-2 Gene. Scientific Reports, 2021, 11, 14392.	3.3	10
77	Biosphere. Applied Bioinformatics, 2004, 3, 253-256.	1.6	9
78	Multi-level learning: improving the prediction of protein, domain and residue interactions by allowing information flow between levels. BMC Bioinformatics, 2009, 10, 241.	2.6	9
79	Comparative analysis of single-cell parallel sequencing approaches in oocyte application. International Journal of Biochemistry and Cell Biology, 2019, 107, 1-5.	2.8	9
80	Case Report: Exome sequencing reveals recurrent RETSAT mutations and a loss-of-function POLDIP2 mutation in a rare undifferentiated tongue sarcoma. F1000Research, 2018, 7, 499.	1.6	9
81	Mining Order-Preserving Submatrices from Data with Repeated Measurements. , 2008, , .		8
82	Towards a More Accurate Error Model for BioNano Optical Maps. Lecture Notes in Computer Science, 2016, , 67-79.	1.3	8
83	Quantifying full-length circular RNAs in cancer. Genome Research, 2021, 31, 2340-2353.	5.5	8
84	Mining Order-Preserving Submatrices from Data with Repeated Measurements. IEEE Transactions on Knowledge and Data Engineering, 2013, 25, 1587-1600.	5.7	7
85	A novel method for discovering local spatial clusters of genomic regions with functional relationships from DNA contact maps. Bioinformatics, 2016, 32, i111-i120.	4.1	7
86	Reusability report: Capturing properties of biological objects and their relationships using graph neural networks. Nature Machine Intelligence, 2022, 4, 222-226.	16.0	7
87	A semi-supervised approach to projected clustering with applications to microarray data. International Journal of Data Mining and Bioinformatics, 2009, 3, 229.	0.1	6
88	Analysis of sequencing data for probing RNA secondary structures and protein–RNA binding in studying posttranscriptional regulations. Briefings in Bioinformatics, 2015, 17, bbv106.	6.5	6
89	START: a system for flexible analysis of hundreds of genomic signal tracks in few lines of SQL-like queries. BMC Genomics, 2017, 18, 749.	2.8	6
90	A unified framework for integrative study of heterogeneous gene regulatory mechanisms. Nature Machine Intelligence, 2020, 2, 447-456.	16.0	6

#	Article	IF	Citations
91	Computational identification of protein binding sites on RNAs using high-throughput RNA structure-probing data. Bioinformatics, 2014, 30, 1049-1055.	4.1	5
92	Integrating Information in Biological Ontologies and Molecular Networks to Infer Novel Terms. Scientific Reports, 2016, 6, 39237.	3.3	5
93	OMMA enables population-scale analysis of complex genomic features and phylogenomic relationships from nanochannel-based optical maps. GigaScience, 2019, 8, .	6.4	5
94	Shaping the nebulous enhancer in the era of high-throughput assays and genome editing. Briefings in Bioinformatics, 2020, 21, 836-850.	6.5	4
95	Deep learning identifies and quantifies recombination hotspot determinants. Bioinformatics, 2022, 38, 2683-2691.	4.1	4
96	On Mining Micro-array data by Order-Preserving Submatrix. , 2005, , .		3
97	A general near-exact k-mer counting method with low memory consumption enables <i>de novo</i> assembly of 106Ã— human sequence data in 2.7 hours. Bioinformatics, 2020, 36, i625-i633.	4.1	3
98	Input Validation for Semi-supervised Clustering. , 2006, , .		2
99	A proof-of-concept study for the pathogenetic role of enhancer hypomethylation of MYBPHL in multiple myeloma. Scientific Reports, 2021, 11, 7009.	3.3	2
100	Identifying projected clusters from gene expression profiles. , 0, , .		1
101	The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. Bioinformatics, 2007, 23, 1048-1048.	4.1	1
102	A Two-Stage Audio Retrieval Method for Searching Unannotated Audio Clips. , 2008, , .		1
103	VAS: a convenient web portal for efficient integration of genomic features with millions of genetic variants. BMC Genomics, 2014, 15, 886.	2.8	1
104	ECplot: an online tool for making standardized plots from large datasets for bioinformatics publications. Bioinformatics, 2014, 30, 1467-1468.	4.1	1
105	A Survey of the Computational Methods for Enhancers and Enhancer-target Predictions. , 2016, , 3-27.		1
106	Reply to â€˜Inflated performance measures in enhancerâ€"promoter interaction-prediction methodsâ€™. Nature Genetics, 2019, 51, 1201-1202.	21.4	1
107	Flexible k-mers with variable-length indels for identifying binding sequences of protein dimers. Briefings in Bioinformatics, 2020, 21, 1787-1797.	6.5	1
108	Abstract 4854: A computational framework for prioritizing noncoding regulatory variants in cancer. Cancer Research, 2015, 75, 4854-4854.	0.9	1

#	Article	IF	Citations
109	Accurate reconstruction of viral genomes in human cells from short reads using iterative refinement. BMC Genomics, 2022, 23, .	2.8	1
110	Abstract 4064: The identification of UBR5-ZNF423 recurrent fusion gene in EBV-associated nasopharyngeal carcinoma. , 2016, , .		0
111	Abstract 482: Epstein-Barr virus-encoded miRNAs target ATM-mediated response in nasopharyngeal carcinoma. , 2018, , .		0
112	IDDF2020-ABS-0201â€...Targeting hepatoma-intrinsic pparÎ³ signaling overcomes immune checkpoint therapy resistance by inflaming the tumor microenvironment. , 2020, , .		0