Xuegong Zhang

List of Publications by Year
in descending order

120 papers	10,424 citations	117625 34 h-index	36028 97 g-index
139 all docs	139 docs citations	139 times ranked	18869 citing authors

#	Article	IF	Citations
1	Toward a unified information framework for cell atlas assembly. National Science Review, 2022, 9, nwab179.	9.5	9
2	ARIC: accurate and robust inference of cell type proportions from bulk gene expression or DNA methylation data. Briefings in Bioinformatics, 2022, 23, .	6.5	6
3	Multifaceted Spatial and Functional Zonation of Cardiac Cells in Adult Human Heart. Circulation, 2022, 145, 315-318.	1.6	8
4	Cell type annotation of single-cell chromatin accessibility data via supervised Bayesian embedding. Nature Machine Intelligence, 2022, 4, 116-126.	16.0	42
5	DeepCAGE: Incorporating Transcription Factors in Genome-Wide Prediction of Chromatin Accessibility. Genomics, Proteomics and Bioinformatics, 2022, 20, 496-507.	6.9	7
6	scGraph: a graph neural network-based approach to automatically identify cell types. Bioinformatics, 2022, 38, 2996-3003.	4.1	9
7	Discovering single-cell eQTLs from scRNA-seq data only. Gene, 2022, 829, 146520.	2.2	9
8	DualGCN: a dual graph convolutional network model to predict cancer drug response. BMC Bioinformatics, 2022, 23, 129.	2.6	7
9	hECA: The cell-centric assembly of a cell atlas. IScience, 2022, 25, 104318.	4.1	21
10	Single-cell genomic profile-based analysis of tissue differentiation in colorectal cancer. Science China Life Sciences, 2021, 64, 1311-1325.	4.9	4
11	Prediction and analysis of metagenomic operons via MetaRon: a pipeline for prediction of Metagenome and whole-genome opeRons. BMC Genomics, 2021, 22, 60.	2.8	2
12	Responses of cyanobacterial aggregate microbial communities to algal blooms. Water Research, 2021, 196, 117014.	11.3	31
13	OpenAnnotate: a web server to annotate the chromatin accessibility of genomic regions. Nucleic Acids Research, 2021, 49, W483-W490.	14.5	17
14	SOMDE: a scalable method for identifying spatially variable genes with self-organizing map. Bioinformatics, 2021, 37, 4392-4398.	4.1	32
15	HGC: fast hierarchical clustering for large-scale single-cell data. Bioinformatics, 2021, 37, 3964-3965.	4.1	13
16	Single-cell Transcriptomes Reveal Characteristics of MicroRNAs in Gene Expression Noise Reduction. Genomics, Proteomics and Bioinformatics, 2021, 19, 394-407.	6.9	5
17	CellTracker: an automated toolbox for single-cell segmentation and tracking of time-lapse microscopy images. Bioinformatics, 2021, 37, 285-287.	4.1	9
18	An Experiment on Ab Initio Discovery of Biological Knowledge from scRNA-Seq Data Using Machine Learning. Patterns, 2020, 1, 100071.	5.9	3

#	Article	IF	Citations
19	SCeQTL: an R package for identifying eQTL from single-cell parallel sequencing data. BMC Bioinformatics, 2020, 21, 184.	2.6	6
20	Single-cell alternative splicing analysis reveals dominance of single transcript variant. Genomics, 2020, 112, 2418-2425.	2.9	19
21	Using DenseFly algorithm for cell searching on massive scRNA-seq datasets. BMC Genomics, 2020, 21, 222.	2.8	5
22	QB: Embracing the future of quantitative understanding and engineering of life. Quantitative Biology, 2019, 7, 1-2.	0.5	2
23	A case study on the detailed reproducibility of a Human Cell Atlas project. Quantitative Biology, 2019, 7, 162-169.	0.5	6
24	A new statistic for efficient detection of repetitive sequences. Bioinformatics, 2019, 35, 4596-4606.	4.1	4
25	Estimating the total genome length of a metagenomic sample using k-mers. BMC Genomics, 2019, 20, 183.	2.8	2
26	Integrative modeling reveals key chromatin and sequence signatures predicting super-enhancers. Scientific Reports, 2019, 9, 2877.	3.3	11
27	Pixel-Level Clustering Reveals Intra-Tumor Heterogeneity in Non-Small Cell Lung Cancer. , 2019, , .		3
28	A Method for Generating Synthetic Electronic Medical Record Text. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019, 18, 1-1.	3.0	14
29	Tongue coating microbiome as a potential biomarker for gastritis including precancerous cascade. Protein and Cell, 2019, 10, 496-509.	11.0	71
30	DEsingle for detecting three types of differential expression in single-cell RNA-seq data. Bioinformatics, 2018, 34, 3223-3224.	4.1	193
31	SEASTAR: systematic evaluation of alternative transcription start sites in RNA. Nucleic Acids Research, 2018, 46, e45-e45.	14.5	17
32	Comprehensive simulation of metagenomic sequencing data with non€uniform sampling distribution. Quantitative Biology, 2018, 6, 175-185.	0.5	2
33	Characterization of kinase gene expression and splicing profile in prostate cancer with RNA-Seq data. BMC Genomics, 2018, 19, 564.	2.8	6
34	Super-enhancers are transcriptionally more active and cell type-specific than stretch enhancers. Epigenetics, 2018, 13, 910-922.	2.7	37
35	Making genome browsers portable and personal. Genome Biology, 2018, 19, 93.	8.8	3
36	Multiple domains of bacterial and human Lon proteases define substrate selectivity. Emerging Microbes and Infections, 2018, 7, 1-18.	6.5	21

#	Article	IF	Citations
37	Precise temporal regulation of alternative splicing during neural development. Nature Communications, 2018, 9, 2189.	12.8	155
38	ulfasQTL: an ultra-fast method of composite splicing QTL analysis. BMC Genomics, 2017, 18, 963.	2.8	19
39	Systematic identification of cancer-related long noncoding RNAs and aberrant alternative splicing of quintuple-negative lung adenocarcinoma through RNA-Seq. Lung Cancer, 2017, 109, 21-27.	2.0	26
40	Network embedding-based representation learning for single cell RNA-seq data. Nucleic Acids Research, 2017, 45, e166-e166.	14.5	54
41	The identification of switch-like alternative splicing exons among multiple samples with RNA-Seq data. PLoS ONE, 2017, 12, e0178320.	2.5	2
42	Identification and functional analysis of a novel *LHX1* mutation associated with congenital absence of the uterus and vagina. Oncotarget, 2017, 8, 8785-8790.	1.8	11
43	Epigenetic Switch Driven by DNA Inversions Dictates Phase Variation in Streptococcus pneumoniae. PLoS Pathogens, 2016, 12, e1005762.	4.7	149
44	Differential expression analyses for single‐cell RNA‐Seq: old questions on new data. Quantitative Biology, 2016, 4, 243-260.	0.5	31
45	CNV analysis in Chinese children of mental retardation highlights a sex differentiation in parental contribution to de novo and inherited mutational burdens. Scientific Reports, 2016, 6, 25954.	3.3	19
46	Reading the Underlying Information From Massive Metagenomic Sequencing Data. Proceedings of the IEEE, 2016, , 1-15.	21.3	12
47	An overview of major metagenomic studies on human microbiomes in health and disease. Quantitative Biology, 2016, 4, 192-206.	0.5	10
48	Exome Sequencing and Gene Prioritization Correct Misdiagnosis in a Chinese Kindred with Familial Amyloid Polyneuropathy. Scientific Reports, 2016, 6, 26362.	3.3	8
49	Precision Medicine: What Challenges Are We Facing?. Genomics, Proteomics and Bioinformatics, 2016, 14, 253-261.	6.9	15
50	Special collection of recent advances in next-generation bioinformatics, part II. Quantitative Biology, 2016, 4, 92-93.	0.5	0
51	Computational prediction of CRISPR cassettes in gut metagenome samples from Chinese type-2 diabetic patients and healthy controls. BMC Systems Biology, 2016, 10, 5.	3.0	12
52	dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic Acids Research, 2016, 44, D164-D171.	14.5	347
53	Special collection of recent advances in next‐generation bioinformatics. Quantitative Biology, 2016, 4, 20-21.	0.5	0
54	A survey of best practices for RNA-seq data analysis. Genome Biology, 2016, 17, 13.	8.8	1,898

#	Article	IF	Citations
55	Identification of important long non-coding RNAs and highly recurrent aberrant alternative splicing events in hepatocellular carcinoma through integrative analysis of multiple RNA-Seq datasets. Molecular Genetics and Genomics, 2016, 291, 1035-1051.	2.1	31
56	The Influence of the Global Gene Expression Shift on Downstream Analyses. PLoS ONE, 2016, 11, e0153903.	2.5	3
57	Annual periodicity in planktonic bacterial and archaeal community composition of eutrophic Lake Taihu. Scientific Reports, 2015, 5, 15488.	3.3	74
58	dslice: an R package for nonparametric testing of associations with application in QTL and gene set analysis. Bioinformatics, 2015, 31, 1842-1844.	4.1	1
59	qDNAmod: a statistical model-based tool to reveal intercellular heterogeneity of DNA modification from SMRT sequencing data. Nucleic Acids Research, 2014, 42, 13488-13499.	14.5	41
60	Exon expression QTL (eeQTL) analysis highlights distant genomic variations associated with splicing regulation. Quantitative Biology, 2014, 2, 71-79.	0.5	6
61	Sequence signatures of genes with accompanying antisense transcripts in Saccharomyces cerevisiae. Science China Life Sciences, 2014, 57, 52-58.	4.9	1
62	Resolving the genetic heterogeneity of prelingual hearing loss within one family: Performance comparison and application of two targeted next generation sequencing approaches. Journal of Human Genetics, 2014, 59, 599-607.	2.3	16
63	Exome Sequencing Identifies a Novel Frameshift Mutation of *MYO6* as the Cause of Autosomal Dominant Nonsyndromic Hearing Loss in a Chinese Family. Annals of Human Genetics, 2014, 78, 410-423.	0.8	10
64	Improvement of Dscam homophilic binding affinity throughout Drosophilaevolution. BMC Evolutionary Biology, 2014, 14, 186.	3.2	1
65	NURD: an implementation of a new method to estimate isoform expression from non-uniform RNA-seq data. BMC Bioinformatics, 2013, 14, 220.	2.6	15
66	Meeting report on RECOMB 2013 (the 17th Annual International Conference on Research in) Tj ETQq0 0 0 rgBT /Overlock 10 Tf 50 302	0.5	1
67	Alignment-free supervised classification of metagenomes by recursive SVM. BMC Genomics, 2013, 14, 641.	2.8	33
68	Identifying differentially spliced genes from two groups of RNA-seq samples. Gene, 2013, 518, 164-170.	2.2	48
69	Opportunities and methods for studying alternative splicing in cancer with RNA-Seq. Cancer Letters, 2013, 340, 179-191.	7.2	107
70	Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. Genome Research, 2013, 23, 129-141.	5.5	99
71	The Fetal Mouse Is a Sensitive Genotoxicity Model That Exposes Lentiviral-associated Mutagenesis Resulting in Liver Oncogenesis. Molecular Therapy, 2013, 21, 324-337.	8.2	21
72	Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic. PLoS Computational Biology, 2013, 9, e1002935.	3.2	67

#	Article	IF	Citations
73	Characterization and Prediction of Lysine (K)-Acetyl-Transferase Specific Acetylation Sites. Molecular and Cellular Proteomics, 2012, 11, M111.011080.	3.8	49
74	Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. Scientific Reports, 2012, 2, 936.	3.3	109
75	Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes. BMC Bioinformatics, 2012, 13, 182.	2.6	110
76	Comparison of metagenomic samples using sequence signatures. BMC Genomics, 2012, 13, 730.	2.8	74
77	Observations on shifted cumulative regulation. Genome Biology, 2011, 12, 404.	9.6	0
78	Observations on novel splice junctions from RNA sequencing data. Biochemical and Biophysical Research Communications, 2011, 409, 299-303.	2.1	16
79	Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. Nature Genetics, 2011, 43, 531-538.	21.4	516
80	Feature rescaling of support vector machines. Tsinghua Science and Technology, 2011, 16, 414-421.	6.1	3
81	Observations on potential novel transcripts from RNA-Seq data. Frontiers of Electrical and Electronic Engineering in China: Selected Publications From Chinese Universities, 2011, 6, 275-282.	0.6	1
82	Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. Bioinformatics, 2011, 27, 502-508.	4.1	91
83	Comparison of gene expression in hepatocellular carcinoma, liver development, and liver regeneration. Molecular Genetics and Genomics, 2010, 283, 485-492.	2.1	24
84	Bioinformatics study indicates possible microRNA-regulated pathways in the differentiation of breast cancer. Science Bulletin, 2010, 55, 927-936.	1.7	3
85	Predicted methylation landscape of all CpG islands on the human genome. Science Bulletin, 2010, 55, 2353-2358.	1.7	0
86	ISOFORM ABUNDANCE INFERENCE PROVIDES A MORE ACCURATE ESTIMATION OF GENE EXPRESSION LEVELS IN RNA-SEQ. Journal of Bioinformatics and Computational Biology, 2010, 08, 177-192.	0.8	33
87	DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics, 2010, 26, 136-138.	4.1	3,728
88	An overview of hepatocellular carcinoma study by omics-based methods. Acta Biochimica Et Biophysica Sinica, 2009, 41, 1-15.	2.0	51
89	Interaction between variants of two glycosyltransferase genes in IgA nephropathy. Kidney International, 2009, 76, 190-198.	5.2	27
90	Predicting the fate of microRNA target genes based on sequence features. Journal of Theoretical Biology, 2009, 261, 17-22.	1.7	5

#	ARTICLE	IF	CITATIONS
91	Pattern recognition methods in microarray based oncology study. Frontiers of Electrical and Electronic Engineering in China: Selected Publications From Chinese Universities, 2009, 4, 243-250.	0.6	0
92	CpG island methylation pattern in different human tissues and its correlation with gene expression. Biochemical and Biophysical Research Communications, 2009, 383, 421-425.	2.1	120
93	Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development. Genomics, 2009, 93, 235-242.	2.9	62
94	Gene-set analysis identifies master transcription factors in developmental courses. Genomics, 2009, 94, 1-10.	2.9	12
95	Prediction of kinaseâ€specific phosphorylation sites with sequence features by a logâ€odds ratio approach. Proteins: Structure, Function and Bioinformatics, 2008, 70, 404-414.	2.6	52
96	Predicting features of breast cancer with gene expression patterns. Breast Cancer Research and Treatment, 2008, 108, 191-201.	2.5	199
97	Histone methylation marks play important roles in predicting the methylation status of CpG islands. Biochemical and Biophysical Research Communications, 2008, 374, 559-564.	2.1	45
98	Computing exact P-values for DNA motifs. Bioinformatics, 2007, 23, 531-537.	4.1	38
99	Putative Zinc Finger Protein Binding Sites Are Over-Represented in the Boundaries of Methylation-Resistant CpG Islands in the Human Genome. PLoS ONE, 2007, 2, e1184.	2.5	16
100	Neighbor number, valley seeking and clustering. Pattern Recognition Letters, 2007, 28, 173-180.	4.2	29
101	Finding distinct biclusters from background in gene expression matrices. Bioinformation, 2007, 2, 207-215.	0.5	0
102	Significance of Gene Ranking for Classification of Microarray Samples. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2006, 3, 312-320.	3.0	45
103	A New Method for Detecting Human Recombination Hotspots and Its Applications to the HapMap ENCODE Data. American Journal of Human Genetics, 2006, 79, 628-639.	6.2	20
104	Predicting methylation status of CpG islands in the human brain. Bioinformatics, 2006, 22, 2204-2209.	4.1	93
105	Primary transcripts and expressions of mammal intergenic microRNAs detected by mapping ESTs to their flanking sequences. Mammalian Genome, 2006, 17, 1033-1041.	2.2	33
106	Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. BMC Bioinformatics, 2006, 7, 197.	2.6	272
107	The effect of GeneChip gene definitions on the microarray study of cancers. BioEssays, 2006, 28, 739-746.	2.5	33
108	Symptom Combinations Associated with Outcome and Therapeutic Effects in a Cohort of Cases with SARS. The American Journal of Chinese Medicine, 2006, 34, 937-947.	3.8	34

#	Article	IF	Citations
109	Characterizing the Microenvironment Surrounding Phosphorylated Protein Sites. Genomics, Proteomics and Bioinformatics, 2005, 3, 213-217.	6.9	11
110	Discrimination and feature selection of geographic origins of traditional Chinese medicine herbs with NIR spectroscopy. Science Bulletin, 2005, 50, 179-184.	1.7	18
111	The effect of U1 snRNA binding free energy on the selection of 5â€² splice sites. Biochemical and Biophysical Research Communications, 2005, 333, 64-69.	2.1	9
112	The Effect of Haplotype-Block Definitions on Inference of Haplotype-Block Structure and htSNPs Selection. Molecular Biology and Evolution, 2004, 22, 148-159.	8.9	24
113	Evidence and characteristics of putative human Î± recombination hotspots. Human Molecular Genetics, 2004, 13, 2823-2828.	2.9	13
114	The impact of very short alternative splicing on protein structures and functions in the human genome. Trends in Genetics, 2004, 20, 232-236.	6.7	33
115	Molecular classification of liver cirrhosis in a rat model by proteomics and bioinformatics. Proteomics, 2004, 4, 3235-3245.	2.2	47
116	On Ï^-Learning. Journal of the American Statistical Association, 2003, 98, 724-734.	3.1	142
117	Multiple removal based on detection and estimation of localized coherent signal. Geophysics, 2003, 68, 745-750.	2.6	5
118	Array comparative genome hybridization for tumor classification and gene discovery in mouse models of malignant melanoma. Cancer Research, 2003, 63, 5352-6.	0.9	34
119	Kernel Nearest-Neighbor Algorithm. Neural Processing Letters, 2002, 15, 147-156.	3.2	104
120	AggEnhance: Aggregation Enhancement by Class Interior Points in Federated Learning with Non-IID Data. ACM Transactions on Intelligent Systems and Technology, 0, , .	4.5	2