

# Pengyi Yang

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/7706880/publications.pdf>

Version: 2024-02-01

86  
papers

4,210  
citations

159358

30  
h-index

138251

58  
g-index

114  
all docs

114  
docs citations

114  
times ranked

6549  
citing authors

#	ARTICLE	IF	CITATIONS
1	A Review of Ensemble Methods in Bioinformatics. <i>Current Bioinformatics</i> , 2010, 5, 296-308.	0.7	399
2	Dynamic Adipocyte Phosphoproteome Reveals that Akt Directly Regulates mTORC2. <i>Cell Metabolism</i> , 2013, 17, 1009-1020.	7.2	352
3	Global Phosphoproteomic Analysis of Human Skeletal Muscle Reveals a Network of Exercise-Regulated Kinases and AMPK Substrates. <i>Cell Metabolism</i> , 2015, 22, 922-935.	7.2	333
4	DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. <i>Nature Communications</i> , 2015, 6, 6910.	5.8	204
5	Ensemble deep learning in bioinformatics. <i>Nature Machine Intelligence</i> , 2020, 2, 500-508.	8.3	162
6	Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. <i>Molecular Cell</i> , 2014, 55, 708-722.	4.5	149
7	Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. <i>EMBO Journal</i> , 2014, 33, 878-889.	3.5	136
8	scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2019, 116, 9775-9784.	3.3	130
9	Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. <i>Cell Systems</i> , 2019, 8, 427-445.e10.	2.9	111
10	Impact of similarity metrics on single-cell RNA-seq data clustering. <i>Briefings in Bioinformatics</i> , 2019, 20, 2316-2326.	3.2	107
11	An integrative systems genetic analysis of mammalian lipid metabolism. <i>Nature</i> , 2019, 567, 187-193.	13.7	101
12	Mitochondrial CoQ deficiency is a common driver of mitochondrial oxidants and insulin resistance. <i>ELife</i> , 2018, 7, .	2.8	91
13	Global redox proteome and phosphoproteome analysis reveals redox switch in Akt. <i>Nature Communications</i> , 2019, 10, 5486.	5.8	89
14	mTORC1 Is a Major Regulatory Node in the FGF21 Signaling Network in Adipocytes. <i>Cell Reports</i> , 2016, 17, 29-36.	2.9	88
15	Intragenic Enhancers Attenuate Host Gene Expression. <i>Molecular Cell</i> , 2017, 68, 104-117.e6.	4.5	85
16	scClassify: sample size estimation and multiscale classification of cells using single and multiple reference. <i>Molecular Systems Biology</i> , 2020, 16, e9389.	3.2	79
17	Sample Subset Optimization Techniques for Imbalanced and Ensemble Learning Problems in Bioinformatics Applications. <i>IEEE Transactions on Cybernetics</i> , 2014, 44, 445-455.	6.2	70
18	CiteFuse enables multi-modal analysis of CITE-seq data. <i>Bioinformatics</i> , 2020, 36, 4137-4143.	1.8	63

#	ARTICLE	IF	CITATIONS
19	Selective photodegradation of 1-methylimidazole-2-thiol by the magnetic and dual conductive imprinted photocatalysts based on TiO <sub>2</sub> /Fe <sub>3</sub> O <sub>4</sub> /MWCNTs. <i>Chemical Engineering Journal</i> , 2014, 240, 244-252.	6.6	61
20	NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. <i>Nature Communications</i> , 2019, 10, 3072.	5.8	53
21	Ensemble-Based Wrapper Methods for Feature Selection and Class Imbalance Learning. <i>Lecture Notes in Computer Science</i> , 2013, , 544-555.	1.0	48
22	PhosR enables processing and functional analysis of phosphoproteomic data. <i>Cell Reports</i> , 2021, 34, 108771.	2.9	48
23	Decoding the function of bivalent chromatin in development and cancer. <i>Genome Research</i> , 2021, 31, 2170-2184.	2.4	48
24	Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data. <i>Genome Biology</i> , 2022, 23, 49.	3.8	48
25	Gene-gene interaction filtering with ensemble of filters. <i>BMC Bioinformatics</i> , 2011, 12, S10.	1.2	47
26	A multi-filter enhanced genetic ensemble system for gene selection and sample classification of microarray data. <i>BMC Bioinformatics</i> , 2010, 11, S5.	1.2	46
27	Evaluating stably expressed genes in single cells. <i>GigaScience</i> , 2019, 8, .	3.3	44
28	A particle swarm based hybrid system for imbalanced medical data sampling. <i>BMC Genomics</i> , 2009, 10, S34.	1.2	42
29	Autoencoder-based cluster ensembles for single-cell RNA-seq data analysis. <i>BMC Bioinformatics</i> , 2019, 20, 660.	1.2	39
30	AdaSampling for Positive-Unlabeled and Label Noise Learning With Bioinformatics Applications. <i>IEEE Transactions on Cybernetics</i> , 2019, 49, 1932-1943.	6.2	37
31	Feature selection revisited in the single-cell era. <i>Genome Biology</i> , 2021, 22, 321.	3.8	37
32	Positive-unlabeled ensemble learning for kinase substrate prediction from dynamic phosphoproteomics data. <i>Bioinformatics</i> , 2016, 32, 252-259.	1.8	34
33	Knowledge-Based Analysis for Detecting Key Signaling Events from Time-Series Phosphoproteomics Data. <i>PLoS Computational Biology</i> , 2015, 11, e1004403.	1.5	32
34	Integrative analysis identifies co-dependent gene expression regulation of BRG1 and CHD7 at distal regulatory sites in embryonic stem cells. <i>Bioinformatics</i> , 2017, 33, 1916-1920.	1.8	31
35	Integrated Glycoproteomics Identifies a Role of N-Glycosylation and Galectin-1 on Myogenesis and Muscle Development. <i>Molecular and Cellular Proteomics</i> , 2021, 20, 100030.	2.5	31
36	Global phosphoproteomics reveals DYRK1A regulates CDK1 activity in glioblastoma cells. <i>Cell Death Discovery</i> , 2021, 7, 81.	2.0	31

#	ARTICLE	IF	CITATIONS
37	Direction pathway analysis of large-scale proteomics data reveals novel features of the insulin action pathway. <i>Bioinformatics</i> , 2014, 30, 808-814.	1.8	29
38	CNOT3-Dependent mRNA Deadenylation Safeguards the Pluripotent State. <i>Stem Cell Reports</i> , 2016, 7, 897-910.	2.3	29
39	scDC: single cell differential composition analysis. <i>BMC Bioinformatics</i> , 2019, 20, 721.	1.2	29
40	A benchmark study of simulation methods for single-cell RNA sequencing data. <i>Nature Communications</i> , 2021, 12, 6911.	5.8	29
41	KinasePA: Phosphoproteomics data annotation using hypothesis driven kinase perturbation analysis. <i>Proteomics</i> , 2016, 16, 1868-1871.	1.3	27
42	Clonal evolution in liver cancer at single-cell and single-variant resolution. <i>Journal of Hematology and Oncology</i> , 2021, 14, 22.	6.9	25
43	A genetic ensemble approach for gene-gene interaction identification. <i>BMC Bioinformatics</i> , 2010, 11, 524.	1.2	22
44	Data Independent Acquisition Proteomic Analysis Can Discriminate between Actinic Keratosis, Bowen's Disease, and Cutaneous Squamous Cell Carcinoma. <i>Journal of Investigative Dermatology</i> , 2020, 140, 212-222.e11.	0.3	22
45	A hierarchical approach to removal of unwanted variation for large-scale metabolomics data. <i>Nature Communications</i> , 2021, 12, 4992.	5.8	22
46	Akt phosphorylates insulin receptor substrate to limit PI3K-mediated PIP3 synthesis. <i>ELife</i> , 2021, 10, .	2.8	21
47	Positive unlabeled learning via wrapper-based adaptive sampling. , 2017, , .		20
48	Enhanced structure and function of human pluripotent stem cell-derived beta-cells cultured on extracellular matrix. <i>Stem Cells Translational Medicine</i> , 2021, 10, 492-505.	1.6	19
49	Signaling Heterogeneity is Defined by Pathway Architecture and Intercellular Variability in Protein Expression. <i>IScience</i> , 2021, 24, 102118.	1.9	19
50	Differentiation of brain and retinal organoids from confluent cultures of pluripotent stem cells connected by nerve-like axonal projections of optic origin. <i>Stem Cell Reports</i> , 2022, 17, 1476-1492.	2.3	19
51	scReClassify: post hoc cell type classification of single-cell rNA-seq data. <i>BMC Genomics</i> , 2019, 20, 913.	1.2	18
52	MiR-93-5p is a novel predictor of coronary in-stent restenosis. <i>Heart Asia</i> , 2019, 11, e011134.	1.1	17
53	Improving X!Tandem on Peptide Identification from Mass Spectrometry by Self-Boosted Percolator. <i>IEEE/ACM Transactions on Computational Biology and Bioinformatics</i> , 2012, 9, 1273-1280.	1.9	16
54	Uncovering cell identity through differential stability with Cepo. <i>Nature Computational Science</i> , 2021, 1, 784-790.	3.8	16

#	ARTICLE	IF	CITATIONS
55	Improved Akt reporter reveals intra- and inter-cellular heterogeneity and oscillations in signal transduction. <i>Journal of Cell Science</i> , 2017, 130, 2757-2766.	1.2	15
56	Sample Subset Optimization for Classifying Imbalanced Biological Data. <i>Lecture Notes in Computer Science</i> , 2011, , 333-344.	1.0	15
57	Transcriptional network dynamics during the progression of pluripotency revealed by integrative statistical learning. <i>Nucleic Acids Research</i> , 2020, 48, 1828-1842.	6.5	14
58	Unraveling Kinase Activation Dynamics Using Kinase-Substrate Relationships from Temporal Large-Scale Phosphoproteomics Studies. <i>PLoS ONE</i> , 2016, 11, e0157763.	1.1	14
59	A clustering based hybrid system for biomarker selection and sample classification of mass spectrometry data. <i>Neurocomputing</i> , 2010, 73, 2317-2331.	3.5	13
60	A dynamic wavelet-based algorithm for pre-processing tandem mass spectrometry data. <i>Bioinformatics</i> , 2010, 26, 2242-2249.	1.8	13
61	Hybrid Methods to Select Informative Gene Sets in Microarray Data Classification. , 2007, , 810-814.		13
62	An Agent-Based Hybrid System for Microarray Data Analysis. <i>IEEE Intelligent Systems</i> , 2009, 24, 53-63.	4.0	11
63	Trafficking regulator of GLUT4-1 (TRARG1) is a GSK3 substrate. <i>Biochemical Journal</i> , 2022, 479, 1237-1256.	1.7	11
64	ISL1 Regulates Peroxisome Proliferator-Activated Receptor $\hat{\text{A}}$ Activation and Early Adipogenesis via Bone Morphogenetic Protein 4-Dependent and -Independent Mechanisms. <i>Molecular and Cellular Biology</i> , 2014, 34, 3607-3617.	1.1	10
65	Temporal ordering of omics and multiomic events inferred from time-series data. <i>Npj Systems Biology and Applications</i> , 2020, 6, 22.	1.4	10
66	OCAP: an open comprehensive analysis pipeline for iTRAQ. <i>Bioinformatics</i> , 2012, 28, 1404-1405.	1.8	9
67	scREMOTE: Using multimodal single cell data to predict regulatory gene relationships and to build a computational cell reprogramming model. <i>NAR Genomics and Bioinformatics</i> , 2022, 4, lqac023.	1.5	9
68	NoisyGOA: Noisy GO annotations prediction using taxonomic and semantic similarity. <i>Computational Biology and Chemistry</i> , 2016, 65, 203-211.	1.1	8
69	Hierarchical kernel mixture models for the prediction of AIDS disease progression using HIV structural gp120 profiles. <i>BMC Genomics</i> , 2010, 11, S22.	1.2	7
70	Protocol for the processing and downstream analysis of phosphoproteomic data with PhosR. <i>STAR Protocols</i> , 2021, 2, 100585.	0.5	7
71	Re-Fraction: A Machine Learning Approach for Deterministic Identification of Protein Homologues and Splice Variants in Large-scale MS-based Proteomics. <i>Journal of Proteome Research</i> , 2012, 11, 3035-3045.	1.8	6
72	Computational systems approach towards phosphoproteomics and their downstream regulation. <i>Proteomics</i> , 2023, 23, e2200068.	1.3	6

#	ARTICLE	IF	CITATIONS
73	QCMAP: An Interactive Web Tool for Performance Diagnosis and Prediction of LC-MS Systems. Proteomics, 2019, 19, 1900068.	1.3	5
74	Computational approaches for direct cell reprogramming: from the bulk omics era to the single cell era. Briefings in Functional Genomics, 2022, 21, 270-279.	1.3	4
75	Time-resolved phosphoproteome and proteome analysis reveals kinase signaling on master transcription factors during myogenesis. Science, 2022, 25, 104489.	1.9	4
76	Modeling and simulation of Parallel Current Mode controlled boost converter. , 2008, , .		3
77	Research and simulation of parallel current-mode controlled buck converter. , 2008, , .		3
78	A Clustering Based Hybrid System for Mass Spectrometry Data Analysis. Lecture Notes in Computer Science, 2008, , 98-109.	1.0	3
79	Multiagent Framework for Bio-data Mining. Lecture Notes in Computer Science, 2009, , 200-207.	1.0	3
80	An Uncertainty Visual Analytics Framework for fMRI Functional Connectivity. Neuroinformatics, 2019, 17, 211-223.	1.5	3
81	Defining cell identity beyond the premise of differential gene expression. Cell Regeneration, 2021, 10, 20.	1.1	3
82	A Hybrid Approach to Selecting Susceptible Single Nucleotide Polymorphisms for Complex Disease Analysis. , 2008, , .		2
83	Genetic Algorithm-Based Multi-objective Optimisation for QoS-Aware Web Services Composition. Lecture Notes in Computer Science, 2010, , 549-554.	1.0	2
84	Functional analysis of the stable phosphoproteome reveals cancer vulnerabilities. Bioinformatics, 2022, 38, 1956-1963.	1.8	2
85	Mouse organogenesis atlas at single-cell resolution. Cell, 2022, 185, 1625-1627.	13.5	1
86	New Parallel Algorithms for All Pairwise Computation on Large HPC Clusters. , 2019, , .		0