

Haixu Tang

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/1899162/publications.pdf>

Version: 2024-02-01

125
papers

8,598
citations

76031

42
h-index

62345

84
g-index

130
all docs

130
docs citations

130
times ranked

14108
citing authors

#	ARTICLE	IF	CITATIONS
1	The Ecoresponsive Genome of <i>Daphnia pulex</i> . <i>Science</i> , 2011, 331, 555-561.	6.0	1,086
2	FragGeneScan: predicting genes in short and error-prone reads. <i>Nucleic Acids Research</i> , 2010, 38, e191-e191.	6.5	760
3	Rate and molecular spectrum of spontaneous mutations in the bacterium <i>Escherichia coli</i> as determined by whole-genome sequencing. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2012, 109, E2774-83.	3.3	595
4	RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. <i>Bioinformatics</i> , 2012, 28, 125-126.	1.8	386
5	ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. <i>Bioinformatics</i> , 2017, 33, 3340-3347.	1.8	272
6	The transcriptional diversity of 25 <i>Drosophila</i> cell lines. <i>Genome Research</i> , 2011, 21, 301-314.	2.4	235
7	Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in <i>Arabidopsis</i> . <i>ELife</i> , 2015, 4, e09591.	2.8	228
8	De Novo Repeat Classification and Fragment Assembly. <i>Genome Research</i> , 2004, 14, 1786-1796.	2.4	195
9	Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. <i>Nature Genetics</i> , 2007, 39, 1361-1368.	9.4	192
10	Leaky Cauldron on the Dark Land. , 2017, 2017, 2421-2434.		169
11	A computational approach toward label-free protein quantification using predicted peptide detectability. <i>Bioinformatics</i> , 2006, 22, e481-e488.	1.8	166
12	MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. <i>Nature Communications</i> , 2021, 12, 3445.	5.8	152
13	Learning your identity and disease from research papers. , 2009, , .		147
14	Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the <i>Arabidopsis</i> epigenome. <i>Genes and Development</i> , 2012, 26, 1825-1836.	2.7	137
15	Determinants of spontaneous mutation in the bacterium <i>Escherichia coli</i> as revealed by whole-genome sequencing. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2015, 112, E5990-9.	3.3	135
16	RAPSearch: a fast protein similarity search tool for short reads. <i>BMC Bioinformatics</i> , 2011, 12, 159.	1.2	128
17	Regulation of adjacent yeast genes. <i>Trends in Genetics</i> , 2000, 16, 109-111.	2.9	127
18	Diverse CRISPRs Evolving in Human Microbiomes. <i>PLoS Genetics</i> , 2012, 8, e1002441.	1.5	125

#	ARTICLE	IF	CITATIONS
19	Whole-Genome Sequencing and Assembly with High-Throughput, Short-Read Technologies. PLoS ONE, 2007, 2, e484.	1.1	112
20	A novel method for multiple alignment of sequences with repeated and shuffled elements. Genome Research, 2004, 14, 2336-2346.	2.4	106
21	Insertion sequence-caused large-scale rearrangements in the genome of <i>Escherichia coli</i> . Nucleic Acids Research, 2016, 44, gkw647.	6.5	104
22	Interlaboratory Study on Differential Analysis of Protein Glycosylation by Mass Spectrometry: The ABRF Glycoprotein Research Multi-Institutional Study 2012. Molecular and Cellular Proteomics, 2013, 12, 2935-2951.	2.5	103
23	Automated interpretation of MS/MS spectra of oligosaccharides. Bioinformatics, 2005, 21, i431-i439.	1.8	101
24	Fragment assembly with double-barreled data. Bioinformatics, 2001, 17, S225-S233.	1.8	97
25	A Two-Step Process for Epigenetic Inheritance in Arabidopsis. Molecular Cell, 2014, 54, 30-42.	4.5	96
26	Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in <i>Escherichia coli</i> . Proceedings of the National Academy of Sciences of the United States of America, 2016, 113, 2176-2181.	3.3	94
27	Computational Framework for Identification of Intact Glycopeptides in Complex Samples. Analytical Chemistry, 2014, 86, 453-463.	3.2	92
28	Quantitative Glycomics Strategies. Molecular and Cellular Proteomics, 2013, 12, 874-884.	2.5	86
29	Efficient Genome-Wide, Privacy-Preserving Similar Patient Query based on Private Edit Distance. , 2015, , .		76
30	On the Mutational Topology of the Bacterial Genome. G3: Genes, Genomes, Genetics, 2013, 3, 399-407.	0.8	75
31	De novo identification of LTR retrotransposons in eukaryotic genomes. BMC Genomics, 2007, 8, 90.	1.2	73
32	A Bayesian Approach to Protein Inference Problem in Shotgun Proteomics. Journal of Computational Biology, 2009, 16, 1183-1193.	0.8	72
33	MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. Nucleic Acids Research, 2009, 37, e143-e143.	6.5	67
34	Improving confidence in detection and characterization of protein N-glycosylation sites and microheterogeneity. Rapid Communications in Mass Spectrometry, 2011, 25, 2007-2019.	0.7	64
35	Quantitative Measurement of Phosphoproteome Response to Osmotic Stress in Arabidopsis Based on Library-Assisted eXtracted Ion Chromatogram (LAXIC). Molecular and Cellular Proteomics, 2013, 12, 2354-2369.	2.5	62
36	Addressing Beacon re-identification attacks: quantification and mitigation of privacy risks. Journal of the American Medical Informatics Association: JAMIA, 2017, 24, 799-805.	2.2	62

#	ARTICLE	IF	CITATIONS
37	Utilizing de Bruijn graph of metagenome assembly for metatranscriptome analysis. <i>Bioinformatics</i> , 2016, 32, 1001-1008.	1.8	59
38	Background Mutational Features of the Radiation-Resistant Bacterium <i>Deinococcus radiodurans</i> . <i>Molecular Biology and Evolution</i> , 2015, 32, 2383-2392.	3.5	58
39	Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. <i>Nature Communications</i> , 2020, 11, 5281.	5.8	57
40	A Novel Alignment Method and Multiple Filters for Exclusion of Unqualified Peptides To Enhance Label-Free Quantification Using Peptide Intensity in LC-MS/MS. <i>Journal of Proteome Research</i> , 2011, 10, 4799-4812.	1.8	55
41	BuildSummary: Using a Group-Based Approach To Improve the Sensitivity of Peptide/Protein Identification in Shotgun Proteomics. <i>Journal of Proteome Research</i> , 2012, 11, 1494-1502.	1.8	52
42	Testosterone Affects Neural Gene Expression Differently in Male and Female Juncos: A Role for Hormones in Mediating Sexual Dimorphism and Conflict. <i>PLoS ONE</i> , 2013, 8, e61784.	1.1	52
43	Full-Spectrum Prediction of Peptides Tandem Mass Spectra using Deep Neural Network. <i>Analytical Chemistry</i> , 2020, 92, 4275-4283.	3.2	48
44	Shotgun Protein Sequencing by Tandem Mass Spectra Assembly. <i>Analytical Chemistry</i> , 2004, 76, 7221-7233.	3.2	47
45	Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. <i>Nature Microbiology</i> , 2021, 6, 123-135.	5.9	47
46	N-Glycan Profiling by Microchip Electrophoresis to Differentiate Disease States Related to Esophageal Adenocarcinoma. <i>Analytical Chemistry</i> , 2012, 84, 3621-3627.	3.2	46
47	On the Accuracy and Limits of Peptide Fragmentation Spectrum Prediction. <i>Analytical Chemistry</i> , 2011, 83, 790-796.	3.2	45
48	Glycoproteomics: Identifying the Glycosylation of Prostate Specific Antigen at Normal and High Isoelectric Points by LC-MS/MS. <i>Journal of Proteome Research</i> , 2014, 13, 5570-5580.	1.8	44
49	Protecting genomic data analytics in the cloud: state of the art and opportunities. <i>BMC Medical Genomics</i> , 2016, 9, 63.	0.7	43
50	Adaptation of <i>Escherichia coli</i> to Long-Term Serial Passage in Complex Medium: Evidence of Parallel Evolution. <i>MSystems</i> , 2017, 2, .	1.7	42
51	RNAMotifScan: automatic identification of RNA structural motifs using secondary structural alignment. <i>Nucleic Acids Research</i> , 2010, 38, e176-e176.	6.5	41
52	The Importance of Peptide Detectability for Protein Identification, Quantification, and Experiment Design in MS/MS Proteomics. <i>Journal of Proteome Research</i> , 2010, 9, 6288-6297.	1.8	41
53	Mapping site-specific protein N-glycosylations through liquid chromatography/mass spectrometry and targeted tandem mass spectrometry. <i>Rapid Communications in Mass Spectrometry</i> , 2010, 24, 965-972.	0.7	39
54	Label-Free Glycopeptide Quantification for Biomarker Discovery in Human Sera. <i>Journal of Proteome Research</i> , 2014, 13, 4821-4832.	1.8	39

#	ARTICLE	IF	CITATIONS
55	A Graph-Centric Approach for Metagenome-Guided Peptide and Protein Identification in Metaproteomics. <i>PLoS Computational Biology</i> , 2016, 12, e1005224.	1.5	36
56	Automated annotation and quantitation of glycans by liquid chromatography/electrospray ionization mass spectrometric analysis using the MultiGlycanâ€ESI computational tool. <i>Rapid Communications in Mass Spectrometry</i> , 2015, 29, 135-142.	0.7	35
57	AN ORFOME ASSEMBLY APPROACH TO METAGENOMICS SEQUENCES ANALYSIS. <i>Journal of Bioinformatics and Computational Biology</i> , 2009, 07, 455-471.	0.3	33
58	A community effort to protect genomic data sharing, collaboration and outsourcing. <i>Npj Genomic Medicine</i> , 2017, 2, 33.	1.7	33
59	Identifying repeat domains in large genomes. <i>Genome Biology</i> , 2006, 7, R7.	13.9	31
60	Automated Glycan Sequencing from Tandem Mass Spectra of N-Linked Glycopeptides. <i>Analytical Chemistry</i> , 2016, 88, 5725-5732.	3.2	31
61	Determinants of Base-Pair Substitution Patterns Revealed by Whole-Genome Sequencing of DNA Mismatch Repair Defective <i>Escherichia coli</i> . <i>Genetics</i> , 2018, 209, 1029-1042.	1.2	31
62	Potential for sexual conflict assessed via testosterone-mediated transcriptional changes in liver and muscle of a songbird. <i>Journal of Experimental Biology</i> , 2014, 217, 507-17.	0.8	28
63	8-plex LCâ€MS/MS Analysis of Permethylated <i>N</i> -Glycans Achieved by Using Stable Isotopic Iodomethane. <i>Analytical Chemistry</i> , 2019, 91, 11794-11802.	3.2	24
64	Rational design of a more stable penicillin G acylase against organic cosolvent. <i>Journal of Molecular Catalysis B: Enzymatic</i> , 2002, 18, 285-290.	1.8	23
65	Genome Assembly, Rearrangement, and Repeats. <i>Chemical Reviews</i> , 2007, 107, 3391-3406.	23.0	23
66	A machine-learning approach to combined evidence validation of genome assemblies. <i>Bioinformatics</i> , 2008, 24, 744-750.	1.8	23
67	LTR retroelements in the genome of <i>Daphnia pulex</i> . <i>BMC Genomics</i> , 2010, 11, 425.	1.2	23
68	Choosing blindly but wisely: differentially private solicitation of DNA datasets for disease marker discovery. <i>Journal of the American Medical Informatics Association: JAMIA</i> , 2015, 22, 100-108.	2.2	22
69	iDASH secure genome analysis competition 2018: blockchain genomic data access logging, homomorphic encryption on GWAS, and DNA segment searching. <i>BMC Medical Genomics</i> , 2020, 13, 98.	0.7	22
70	ADVANCEMENT IN PROTEIN INFERENCE FROM SHOTGUN PROTEOMICS USING PEPTIDE DETECTABILITY. , 2006, , .		22
71	Proteomic Changes in the Photoreceptor Outer Segment upon Intense Light Exposure. <i>Journal of Proteome Research</i> , 2010, 9, 1173-1181.	1.8	21
72	A machine learning approach to predicting peptide fragmentation spectra. <i>Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing</i> , 2006, , 219-30.	0.7	21

#	ARTICLE	IF	CITATIONS
73	The Spectrum of Replication Errors in the Absence of Error Correction Assayed Across the Whole Genome of <i>Escherichia coli</i> . <i>Genetics</i> , 2018, 209, 1043-1054.	1.2	20
74	The Symmetrical Wave Pattern of Base-Pair Substitution Rates across the <i>Escherichia coli</i> Chromosome Has Multiple Causes. <i>MBio</i> , 2019, 10, .	1.8	20
75	Characterization of the Glycosylation Site of Human PSA Prompted by Missense Mutation using LC-MS/MS. <i>Journal of Proteome Research</i> , 2015, 14, 2872-2883.	1.8	19
76	msCRUSH: Fast Tandem Mass Spectral Clustering Using Locality Sensitive Hashing. <i>Journal of Proteome Research</i> , 2019, 18, 147-158.	1.8	19
77	A <i>de Bruijn</i> Graph Approach to the Quantification of Closely-Related Genomes in a Microbial Community. <i>Journal of Computational Biology</i> , 2012, 19, 814-825.	0.8	18
78	Probabilistic Inference of Biochemical Reactions in Microbial Communities from Metagenomic Sequences. <i>PLoS Computational Biology</i> , 2013, 9, e1002981.	1.5	17
79	Insertion sequence elements-mediated structural variations in bacterial genomes. <i>Mobile DNA</i> , 2018, 9, 29.	1.3	17
80	A Meta-proteogenomic Approach to Peptide Identification Incorporating Assembly Uncertainty and Genomic Variation. <i>Molecular and Cellular Proteomics</i> , 2019, 18, S183-S192.	2.5	17
81	Combinatorial Libraries of Synthetic Peptides as a Model for Shotgun Proteomics. <i>Analytical Chemistry</i> , 2010, 82, 6559-6568.	3.2	16
82	DNA sequence templates adjacent nucleosome and ORC sites at gene amplification origins in <i>Drosophila</i> . <i>Nucleic Acids Research</i> , 2015, 43, 8746-8761.	6.5	15
83	XLSearch: a Probabilistic Database Search Algorithm for Identifying Cross-Linked Peptides. <i>Journal of Proteome Research</i> , 2016, 15, 1830-1841.	1.8	15
84	Insertion polymorphisms of mobile genetic elements in sexual and asexual populations of <i>Daphnia pulex</i> . <i>Genome Biology and Evolution</i> , 2017, 9, evw302.	1.1	15
85	iDASH secure genome analysis competition 2017. <i>BMC Medical Genomics</i> , 2018, 11, 85.	0.7	15
86	Privacy-preserving construction of generalized linear mixed model for biomedical computation. <i>Bioinformatics</i> , 2020, 36, i128-i135.	1.8	15
87	Detection of Structural Variants Involving Repetitive Regions in the Reference Genome. <i>Journal of Computational Biology</i> , 2014, 21, 219-233.	0.8	14
88	STRScan: targeted profiling of short tandem repeats in whole-genome sequencing data. <i>BMC Bioinformatics</i> , 2017, 18, 398.	1.2	14
89	Independent Mammalian Genome Contractions Following the KT Boundary. <i>Genome Biology and Evolution</i> , 2009, 1, 2-12.	1.1	13
90	Identification of Glycopeptides with Multiple Hydroxylysine O-Glycosylation Sites by Tandem Mass Spectrometry. <i>Journal of Proteome Research</i> , 2015, 14, 5099-5108.	1.8	13

#	ARTICLE	IF	CITATIONS
91	Adaptation of Escherichia coli to long-term batch culture in various rich media. <i>Research in Microbiology</i> , 2018, 169, 145-156.	1.0	13
92	A New Estimator of Significance of Correlation in Time Series Data. <i>Journal of Computational Biology</i> , 2001, 8, 463-470.	0.8	12
93	Computational Methods in Mass Spectrometry-Based Proteomics. <i>Advances in Experimental Medicine and Biology</i> , 2016, 939, 63-89.	0.8	12
94	On the estimation of false positives in peptide identifications using decoy search strategy. <i>Proteomics</i> , 2009, 9, 194-204.	1.3	11
95	A Fast and Memory-Efficient Spectral Library Search Algorithm Using Locality-Sensitive Hashing. <i>Proteomics</i> , 2020, 20, e2000002.	1.3	10
96	Correcting Base-Assignment Errors in Repeat Regions of Shotgun Assembly. <i>IEEE/ACM Transactions on Computational Biology and Bioinformatics</i> , 2007, 4, 54-64.	1.9	9
97	Extending the coverage of spectral libraries: A neighbor-based approach to predicting intensities of peptide fragmentation spectra. <i>Proteomics</i> , 2013, 13, 756-765.	1.3	9
98	Bioinformatics Protocols in Glycomics and Glycoproteomics. <i>Current Protocols in Protein Science</i> , 2014, 76, 2.15.1-2.15.7.	2.8	9
99	New complexities of SOS-induced ϕ -mutagenesis in Escherichia coli as revealed by mutation accumulation and whole-genome sequencing. <i>DNA Repair</i> , 2020, 90, 102852.	1.3	9
100	GlycoHybridSeq: Automated Identification of N-Linked Glycopeptides Using Electron Transfer/High-Energy Collision Dissociation (ETHcD). <i>Journal of Proteome Research</i> , 2021, 20, 3345-3352.	1.8	9
101	Impact of Amidination on Peptide Fragmentation and Identification in Shotgun Proteomics. <i>Journal of Proteome Research</i> , 2016, 15, 3656-3665.	1.8	8
102	Computational identification of micro-structural variations and their proteogenomic consequences in cancer. <i>Bioinformatics</i> , 2018, 34, 1672-1681.	1.8	8
103	Protein identification problem from a Bayesian point of view. <i>Statistics and Its Interface</i> , 2012, 5, 21-37.	0.2	8
104	Algorithmic approaches to clonal reconstruction in heterogeneous cell populations. <i>Quantitative Biology</i> , 2019, 7, 255-265.	0.3	7
105	Constrained De Novo Sequencing of neo-Epitope Peptides Using Tandem Mass Spectrometry. <i>Lecture Notes in Computer Science</i> , 2018, 10812, 138-153.	1.0	6
106	Clonal reconstruction from time course genomic sequencing data. <i>BMC Genomics</i> , 2019, 20, 1002.	1.2	6
107	A computational approach to characterizing bond linkages of glycan isomers using matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry. <i>Rapid Communications in Mass Spectrometry</i> , 2008, 22, 3561-3569.	0.7	4
108	A Secure Alignment Algorithm for Mapping Short Reads to Human Genome. <i>Journal of Computational Biology</i> , 2018, 25, 529-540.	0.8	4

#	ARTICLE	IF	CITATIONS
109	Haplotype-based membership inference from summary genomic data. <i>Bioinformatics</i> , 2021, 37, i161-i168.	1.8	4
110	Improving phosphopeptide identification in shotgun proteomics by supervised filtering of peptide-spectrum matches. , 2013, , .		3
111	Characterization of microbial associations in human oral microbiome. <i>Bio-Medical Materials and Engineering</i> , 2014, 24, 3737-3744.	0.4	3
112	Overlap detection on long, error-prone sequencing reads via smooth q -gram. <i>Bioinformatics</i> , 2020, 36, 4838-4845.	1.8	3
113	A Computational Approach for the Identification of Site-Specific Protein Glycosylations Through Ion-Trap Mass Spectrometry. <i>Lecture Notes in Computer Science</i> , 2007, , 96-107.	1.0	3
114	On the privacy risks of sharing clinical proteomics data. <i>AMIA Summits on Translational Science Proceedings</i> , 2016, 2016, 122-31.	0.4	3
115	A Maximum-Likelihood Approach to Estimating the Insertion Frequencies of Transposable Elements from Population Sequencing Data. <i>Molecular Biology and Evolution</i> , 2018, 35, 2560-2571.	3.5	2
116	Identification of N-terminal protein processing sites by chemical labeling mass spectrometry. <i>Rapid Communications in Mass Spectrometry</i> , 2019, 33, 1015-1023.	0.7	2
117	HySec-Flow: Privacy-Preserving Genomic Computing with SGX-based Big-Data Analytics Framework. , 2021, , .		2
118	Secure Genomic Computation through Site-Wise Encryption. <i>AMIA Summits on Translational Science Proceedings</i> , 2015, 2015, 227-31.	0.4	2
119	Real-time Protection of Genomic Data Sharing in Beacon Services. <i>AMIA Summits on Translational Science Proceedings</i> , 2018, 2017, 45-54.	0.4	2
120	Locality-Sensitive Hashing-Based k -Mer Clustering for Identification of Differential Microbial Markers Related to Host Phenotype. <i>Journal of Computational Biology</i> , 2022, 29, 738-751.	0.8	2
121	Bioinformatic Approaches in Glycomics and Glycoproteomics. <i>Current Proteomics</i> , 2011, 8, 309-324.	0.1	1
122	GlycanGUI: Automated Glycan Annotation and Quantification Using Glucose Unit Index. <i>Frontiers in Chemistry</i> , 2021, 9, 707382.	1.8	1
123	Improving de novo metatranscriptome assembly via machine learning algorithms. <i>International Journal of Computational Biology and Drug Design</i> , 2017, 10, 91.	0.3	1
124	A Reciprocal Best-hit Approach to Characterize Isomeric N -Glycans Using Tandem Mass Spectrometry. <i>Analytical Chemistry</i> , 2022, 94, 10003-10010.	3.2	1
125	A Community Effort to Protect Genomic Data Sharing, Collaboration and Outsourcing. <i>SSRN Electronic Journal</i> , 2017, , .	0.4	0