

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

66343

42
h-index

54911

84
g-index

130
all docs

130
docs citations

130
times ranked

12512
citing authors

#	ARTICLE	IF	CITATIONS
1	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.	1.6	2
2	A Reciprocal Best-hit Approach to Characterize Isomeric <i>N</i> -Glycans Using Tandem Mass Spectrometry. <i>Analytical Chemistry</i> , 2022, 94, 10003-10010.	6.5	1
3	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.	13.3	47
4	GlycoHybridSeq: Automated Identification of N-Linked Glycopeptides Using Electron Transfer/High-Energy Collision Dissociation (ET _h CD). <i>Journal of Proteome Research</i> , 2021, 20, 3345-3352.	3.7	9
5	MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. <i>Nature Communications</i> , 2021, 12, 3445.	12.8	152
6	GlycanGUI: Automated Glycan Annotation and Quantification Using Glucose Unit Index. <i>Frontiers in Chemistry</i> , 2021, 9, 707382.	3.6	1
7	Haplotype-based membership inference from summary genomic data. <i>Bioinformatics</i> , 2021, 37, i161-i168.	4.1	4
8	HySec-Flow: Privacy-Preserving Genomic Computing with SGX-based Big-Data Analytics Framework. , 2021, , .		2
9	Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. <i>Nature Communications</i> , 2020, 11, 5281.	12.8	57
10	Privacy-preserving construction of generalized linear mixed model for biomedical computation. <i>Bioinformatics</i> , 2020, 36, i128-i135.	4.1	15
11	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.	1.5	22
12	A Fast and Memory-Efficient Spectral Library Search Algorithm Using Locality-Sensitive Hashing. <i>Proteomics</i> , 2020, 20, e2000002.	2.2	10
13	Full-Spectrum Prediction of Peptides Tandem Mass Spectra using Deep Neural Network. <i>Analytical Chemistry</i> , 2020, 92, 4275-4283.	6.5	48
14	New complexities of SOS-induced "untargeted" mutagenesis in <i>Escherichia coli</i> as revealed by mutation accumulation and whole-genome sequencing. <i>DNA Repair</i> , 2020, 90, 102852.	2.8	9
15	Overlap detection on long, error-prone sequencing reads via smooth <i>q</i> -gram. <i>Bioinformatics</i> , 2020, 36, 4838-4845.	4.1	3
16	msCRUSH: Fast Tandem Mass Spectral Clustering Using Locality Sensitive Hashing. <i>Journal of Proteome Research</i> , 2019, 18, 147-158.	3.7	19
17	The Symmetrical Wave Pattern of Base-Pair Substitution Rates across the <i>Escherichia coli</i> Chromosome Has Multiple Causes. <i>MBio</i> , 2019, 10, .	4.1	20
18	8-plex LC-MS/MS Analysis of Permethylated <i>N</i> -Glycans Achieved by Using Stable Isotopic Iodomethane. <i>Analytical Chemistry</i> , 2019, 91, 11794-11802.	6.5	24

#	ARTICLE	IF	CITATIONS
19	A Meta-proteogenomic Approach to Peptide Identification Incorporating Assembly Uncertainty and Genomic Variation. <i>Molecular and Cellular Proteomics</i> , 2019, 18, S183-S192.	3.8	17
20	Identification of N-terminal protein processing sites by chemical labeling mass spectrometry. <i>Rapid Communications in Mass Spectrometry</i> , 2019, 33, 1015-1023.	1.5	2
21	Clonal reconstruction from time course genomic sequencing data. <i>BMC Genomics</i> , 2019, 20, 1002.	2.8	6
22	Algorithmic approaches to clonal reconstruction in heterogeneous cell populations. <i>Quantitative Biology</i> , 2019, 7, 255-265.	0.5	7
23	Constrained De Novo Sequencing of neo-Epitope Peptides Using Tandem Mass Spectrometry. <i>Lecture Notes in Computer Science</i> , 2018, 10812, 138-153.	1.3	6
24	Adaptation of <i>Escherichia coli</i> to long-term batch culture in various rich media. <i>Research in Microbiology</i> , 2018, 169, 145-156.	2.1	13
25	Computational identification of micro-structural variations and their proteogenomic consequences in cancer. <i>Bioinformatics</i> , 2018, 34, 1672-1681.	4.1	8
26	iDASH secure genome analysis competition 2017. <i>BMC Medical Genomics</i> , 2018, 11, 85.	1.5	15
27	Insertion sequence elements-mediated structural variations in bacterial genomes. <i>Mobile DNA</i> , 2018, 9, 29.	3.6	17
28	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.	2.9	20
29	A Secure Alignment Algorithm for Mapping Short Reads to Human Genome. <i>Journal of Computational Biology</i> , 2018, 25, 529-540.	1.6	4
30	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.	8.9	2
31	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.	2.9	31
32	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
33	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.	2.5	15
34	Addressing Beacon re-identification attacks: quantification and mitigation of privacy risks. <i>Journal of the American Medical Informatics Association: JAMIA</i> , 2017, 24, 799-805.	4.4	62
35	Adaptation of <i>Escherichia coli</i> to Long-Term Serial Passage in Complex Medium: Evidence of Parallel Evolution. <i>MSystems</i> , 2017, 2, .	3.8	42
36	A community effort to protect genomic data sharing, collaboration and outsourcing. <i>Npj Genomic Medicine</i> , 2017, 2, 33.	3.8	33

#	ARTICLE	IF	CITATIONS
37	Leaky Cauldron on the Dark Land. , 2017, 2017, 2421-2434.		169
38	ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics, 2017, 33, 3340-3347.	4.1	272
39	STRScan: targeted profiling of short tandem repeats in whole-genome sequencing data. BMC Bioinformatics, 2017, 18, 398.	2.6	14
40	A Community Effort to Protect Genomic Data Sharing, Collaboration and Outsourcing. SSRN Electronic Journal, 2017, , .	0.4	0
41	Improving de novo metatranscriptome assembly via machine learning algorithms. International Journal of Computational Biology and Drug Design, 2017, 10, 91.	0.3	1
42	Protecting genomic data analytics in the cloud: state of the art and opportunities. BMC Medical Genomics, 2016, 9, 63.	1.5	43
43	XLSearch: a Probabilistic Database Search Algorithm for Identifying Cross-Linked Peptides. Journal of Proteome Research, 2016, 15, 1830-1841.	3.7	15
44	Automated Glycan Sequencing from Tandem Mass Spectra of N-Linked Glycopeptides. Analytical Chemistry, 2016, 88, 5725-5732.	6.5	31
45	Insertion sequence-caused large-scale rearrangements in the genome of <i>Escherichia coli</i> . Nucleic Acids Research, 2016, 44, gkw647.	14.5	104
46	Impact of Amidation on Peptide Fragmentation and Identification in Shotgun Proteomics. Journal of Proteome Research, 2016, 15, 3656-3665.	3.7	8
47	Computational Methods in Mass Spectrometry-Based Proteomics. Advances in Experimental Medicine and Biology, 2016, 939, 63-89.	1.6	12
48	Utilizing de Bruijn graph of metagenome assembly for metatranscriptome analysis. Bioinformatics, 2016, 32, 1001-1008.	4.1	59
49	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.	7.1	94
50	A Graph-Centric Approach for Metagenome-Guided Peptide and Protein Identification in Metaproteomics. PLoS Computational Biology, 2016, 12, e1005224.	3.2	36
51	On the privacy risks of sharing clinical proteomics data. AMIA Summits on Translational Science Proceedings, 2016, 2016, 122-31.	0.4	3
52	Efficient Genome-Wide, Privacy-Preserving Similar Patient Query based on Private Edit Distance. , 2015, , .		76
53	Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. ELife, 2015, 4, e09591.	6.0	228
54	Characterization of the Glycosylation Site of Human PSA Prompted by Missense Mutation using LC-MS/MS. Journal of Proteome Research, 2015, 14, 2872-2883.	3.7	19

#	ARTICLE	IF	CITATIONS
55	Background Mutational Features of the Radiation-Resistant Bacterium <i>Deinococcus radiodurans</i> . <i>Molecular Biology and Evolution</i> , 2015, 32, 2383-2392.	8.9	58
56	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.	14.5	15
57	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.	4.4	22
58	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.	7.1	135
59	Identification of Glycopeptides with Multiple Hydroxylysine O-Glycosylation Sites by Tandem Mass Spectrometry. <i>Journal of Proteome Research</i> , 2015, 14, 5099-5108.	3.7	13
60	Automated annotation and quantitation of glycans by liquid chromatography/electrospray ionization mass spectrometric analysis using the MultiGlycan-ESI computational tool. <i>Rapid Communications in Mass Spectrometry</i> , 2015, 29, 135-142.	1.5	35
61	Secure Genomic Computation through Site-Wise Encryption. <i>AMIA Summits on Translational Science Proceedings</i> , 2015, 2015, 227-31.	0.4	2
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.	1.7	28
63	Characterization of microbial associations in human oral microbiome. <i>Bio-Medical Materials and Engineering</i> , 2014, 24, 3737-3744.	0.6	3
64	Detection of Structural Variants Involving Repetitive Regions in the Reference Genome. <i>Journal of Computational Biology</i> , 2014, 21, 219-233.	1.6	14
65	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.	3.7	44
66	Bioinformatics Protocols in Glycomics and Glycoproteomics. <i>Current Protocols in Protein Science</i> , 2014, 76, 2.15.1-2.15.7.	2.8	9
67	A Two-Step Process for Epigenetic Inheritance in Arabidopsis. <i>Molecular Cell</i> , 2014, 54, 30-42.	9.7	96
68	Computational Framework for Identification of Intact Glycopeptides in Complex Samples. <i>Analytical Chemistry</i> , 2014, 86, 453-463.	6.5	92
69	Label-Free Glycopeptide Quantification for Biomarker Discovery in Human Sera. <i>Journal of Proteome Research</i> , 2014, 13, 4821-4832.	3.7	39
70	Improving phosphopeptide identification in shotgun proteomics by supervised filtering of peptide-spectrum matches. , 2013, , .		3
71	On the Mutational Topology of the Bacterial Genome. <i>G3: Genes, Genomes, Genetics</i> , 2013, 3, 399-407.	1.8	75
72	Probabilistic Inference of Biochemical Reactions in Microbial Communities from Metagenomic Sequences. <i>PLoS Computational Biology</i> , 2013, 9, e1002981.	3.2	17

#	ARTICLE	IF	CITATIONS
73	Quantitative Glycomics Strategies. <i>Molecular and Cellular Proteomics</i> , 2013, 12, 874-884.	3.8	86
74	Extending the coverage of spectral libraries: A neighbor-based approach to predicting intensities of peptide fragmentation spectra. <i>Proteomics</i> , 2013, 13, 756-765.	2.2	9
75	Quantitative Measurement of Phosphoproteome Response to Osmotic Stress in Arabidopsis Based on Library-Assisted eXtracted Ion Chromatogram (LAXIC). <i>Molecular and Cellular Proteomics</i> , 2013, 12, 2354-2369.	3.8	62
76	Interlaboratory Study on Differential Analysis of Protein Glycosylation by Mass Spectrometry: The ABRF Glycoprotein Research Multi-Institutional Study 2012. <i>Molecular and Cellular Proteomics</i> , 2013, 12, 2935-2951.	3.8	103
77	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.	2.5	52
78	Diverse CRISPRs Evolving in Human Microbiomes. <i>PLoS Genetics</i> , 2012, 8, e1002441.	3.5	125
79	RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. <i>Bioinformatics</i> , 2012, 28, 125-126.	4.1	386
80	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.	7.1	595
81	A de Bruijn Graph Approach to the Quantification of Closely-Related Genomes in a Microbial Community. <i>Journal of Computational Biology</i> , 2012, 19, 814-825.	1.6	18
82	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.	5.9	137
83	N-Glycan Profiling by Microchip Electrophoresis to Differentiate Disease States Related to Esophageal Adenocarcinoma. <i>Analytical Chemistry</i> , 2012, 84, 3621-3627.	6.5	46
84	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.	3.7	52
85	Protein identification problem from a Bayesian point of view. <i>Statistics and Its Interface</i> , 2012, 5, 21-37.	0.3	8
86	On the Accuracy and Limits of Peptide Fragmentation Spectrum Prediction. <i>Analytical Chemistry</i> , 2011, 83, 790-796.	6.5	45
87	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.	3.7	55
88	The Ecoresponsive Genome of <i>Daphnia pulex</i> . <i>Science</i> , 2011, 331, 555-561.	12.6	1,086
89	RAPSearch: a fast protein similarity search tool for short reads. <i>BMC Bioinformatics</i> , 2011, 12, 159.	2.6	128
90	Improving confidence in detection and characterization of protein N-glycosylation sites and microheterogeneity. <i>Rapid Communications in Mass Spectrometry</i> , 2011, 25, 2007-2019.	1.5	64

#	ARTICLE	IF	CITATIONS
91	Bioinformatic Approaches in Glycomics and Glycoproteomics. <i>Current Proteomics</i> , 2011, 8, 309-324.	0.3	1
92	The transcriptional diversity of 25 <i>Drosophila</i> cell lines. <i>Genome Research</i> , 2011, 21, 301-314.	5.5	235
93	LTR retroelements in the genome of <i>Daphnia pulex</i> . <i>BMC Genomics</i> , 2010, 11, 425.	2.8	23
94	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.	1.5	39
95	RNAMotifScan: automatic identification of RNA structural motifs using secondary structural alignment. <i>Nucleic Acids Research</i> , 2010, 38, e176-e176.	14.5	41
96	FragGeneScan: predicting genes in short and error-prone reads. <i>Nucleic Acids Research</i> , 2010, 38, e191-e191.	14.5	760
97	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.	3.7	41
98	Combinatorial Libraries of Synthetic Peptides as a Model for Shotgun Proteomics. <i>Analytical Chemistry</i> , 2010, 82, 6559-6568.	6.5	16
99	Proteomic Changes in the Photoreceptor Outer Segment upon Intense Light Exposure. <i>Journal of Proteome Research</i> , 2010, 9, 1173-1181.	3.7	21
100	AN ORFOME ASSEMBLY APPROACH TO METAGENOMICS SEQUENCES ANALYSIS. <i>Journal of Bioinformatics and Computational Biology</i> , 2009, 07, 455-471.	0.8	33
101	A Bayesian Approach to Protein Inference Problem in Shotgun Proteomics. <i>Journal of Computational Biology</i> , 2009, 16, 1183-1193.	1.6	72
102	Learning your identity and disease from research papers. , 2009, , .		147
103	MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. <i>Nucleic Acids Research</i> , 2009, 37, e143-e143.	14.5	67
104	On the estimation of false positives in peptide identifications using decoy search strategy. <i>Proteomics</i> , 2009, 9, 194-204.	2.2	11
105	Independent Mammalian Genome Contractions Following the KT Boundary. <i>Genome Biology and Evolution</i> , 2009, 1, 2-12.	2.5	13
106	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.	1.5	4
107	A machine-learning approach to combined evidence validation of genome assemblies. <i>Bioinformatics</i> , 2008, 24, 744-750.	4.1	23
108	Correcting Base-Assignment Errors in Repeat Regions of Shotgun Assembly. <i>IEEE/ACM Transactions on Computational Biology and Bioinformatics</i> , 2007, 4, 54-64.	3.0	9

#	ARTICLE	IF	CITATIONS
109	Genome Assembly, Rearrangement, and Repeats. Chemical Reviews, 2007, 107, 3391-3406.	47.7	23
110	Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. Nature Genetics, 2007, 39, 1361-1368.	21.4	192
111	De novo identification of LTR retrotransposons in eukaryotic genomes. BMC Genomics, 2007, 8, 90.	2.8	73
112	Whole-Genome Sequencing and Assembly with High-Throughput, Short-Read Technologies. PLoS ONE, 2007, 2, e484.	2.5	112
113	A Computational Approach for the Identification of Site-Specific Protein Glycosylations Through Ion-Trap Mass Spectrometry. Lecture Notes in Computer Science, 2007, , 96-107.	1.3	3
114	Identifying repeat domains in large genomes. Genome Biology, 2006, 7, R7.	9.6	31
115	A computational approach toward label-free protein quantification using predicted peptide detectability. Bioinformatics, 2006, 22, e481-e488.	4.1	166
116	ADVANCEMENT IN PROTEIN INFERENCE FROM SHOTGUN PROTEOMICS USING PEPTIDE DETECTABILITY. , 2006, , .		22
117	A machine learning approach to predicting peptide fragmentation spectra. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 2006, , 219-30.	0.7	21
118	Automated interpretation of MS/MS spectra of oligosaccharides. Bioinformatics, 2005, 21, i431-i439.	4.1	101
119	A novel method for multiple alignment of sequences with repeated and shuffled elements. Genome Research, 2004, 14, 2336-2346.	5.5	106
120	De Novo Repeat Classification and Fragment Assembly. Genome Research, 2004, 14, 1786-1796.	5.5	195
121	Shotgun Protein Sequencing by Tandem Mass Spectra Assembly. Analytical Chemistry, 2004, 76, 7221-7233.	6.5	47
122	Rational design of a more stable penicillin G acylase against organic cosolvent. Journal of Molecular Catalysis B: Enzymatic, 2002, 18, 285-290.	1.8	23
123	Fragment assembly with double-barreled data. Bioinformatics, 2001, 17, S225-S233.	4.1	97
124	A New Estimator of Significance of Correlation in Time Series Data. Journal of Computational Biology, 2001, 8, 463-470.	1.6	12
125	Regulation of adjacent yeast genes. Trends in Genetics, 2000, 16, 109-111.	6.7	127