Jeffrey T Leek

List of Publications by Year
in descending order

Source: https://exaly.com/author-pdf/3771243/publications.pdf

Version: 2024-02-01

78	20,526	66343	74163
papers	citations	42 h-index	75 g-index

106	106	106	35552
all docs	docs citations	times ranked	citing authors

#	Article	IF	Citations
1	Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space. Cell Genomics, 2022, 2, 100085.	6.5	59
2	Transcriptional profile of platelets and iPSC-derived megakaryocytes from whole-genome and RNA sequencing. Blood, 2021, 137, 959-968.	1.4	21
3	Gene and protein expression in human megakaryocytes derived from induced pluripotent stem cells. Journal of Thrombosis and Haemostasis, 2021, 19, 1783-1799.	3.8	6
4	recount3: summaries and queries for large-scale RNA-seq expression and splicing. Genome Biology, 2021, 22, 323.	8.8	103
5	Diagnosing Data Analytic Problems in the Classroom. Journal of Statistics and Data Science Education, 2021, 29, 267-276.	1.6	3
6	The Democratization of Data Science Education. American Statistician, 2020, 74, 1-7.	1.6	21
7	Transparency and reproducibility in artificial intelligence. Nature, 2020, 586, E14-E16.	27.8	233
8	Methods for correcting inference based on outcomes predicted by machine learning. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117, 30266-30275.	7.1	28
9	Recounting the FANTOM CAGE-Associated Transcriptome. Genome Research, 2020, 30, 1073-1081.	5.5	35
10	A visual tool for defining reproducibility and replicability. Nature Human Behaviour, 2019, 3, 650-652.	12.0	26
11	Addressing confounding artifacts in reconstruction of gene co-expression networks. Genome Biology, 2019, 20, 94.	8.8	68
12	Improving the value of public RNA-seq expression data by phenotype prediction. Nucleic Acids Research, 2018, 46, e54-e54.	14.5	49
13	How to Share Data for Collaboration. American Statistician, 2018, 72, 53-57.	1.6	14
14	Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nature Neuroscience, 2018, 21, 1117-1125.	14.8	300
15	A direct approach to estimating false discovery rates conditional on covariates. PeerJ, 2018, 6, e6035.	2.0	60
16	Rail-RNA: scalable analysis of RNA-seq splicing and coverage. Bioinformatics, 2017, 33, 4033-4040.	4.1	57
17	Reproducible RNA-seq analysis using recount2. Nature Biotechnology, 2017, 35, 319-321.	17.5	395
18	qSVA framework for RNA quality correction in differential expression analysis. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114, 7130-7135.	7.1	95

#	Article	IF	Citations
19	Flexible expressed region analysis for RNA-seq with<tt>derfinder</tt>. Nucleic Acids Research, 2017, 45, e9-e9.	14.5	54
20	Is Most Published Research Really False?. Annual Review of Statistics and Its Application, 2017, 4, 109-122.	7.0	44
21	Evolution of cellular morpho-phenotypes in cancer metastasis. Scientific Reports, 2016, 5, 18437.	3.3	81
22	Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive. Genome Biology, 2016, 17, 266.	8.8	94
23	Genomic and clinical predictors for improving estimator precision in randomized trials of breast cancer treatments. Contemporary Clinical Trials Communications, 2016, 3, 48-54.	1.1	1
24	What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. Perspectives on Psychological Science, 2016, 11, 539-544.	9.0	168
25	Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols, 2016, 11, 1650-1667.	12.0	4,743
26	BatchQC: interactive software for evaluating sample and batch effects in genomic data. Bioinformatics, 2016, 32, 3836-3838.	4.1	50
27	Rail-dbGaP: analyzing dbGaP-protected data in the cloud with Amazon Elastic MapReduce. Bioinformatics, 2016, 32, 2551-2553.	4.1	5
28	Discussion of â€œvisualizing statistical models: Removing the blindfoldâ€• Statistical Analysis and Data Mining, 2015, 8, 240-241.	2.8	0
29	Practical impacts of genomic data â€œcleaningâ€•on biological discovery using surrogate variable analysis. BMC Bioinformatics, 2015, 16, 372.	2.6	51
30	Statistics: P values are just the tip of the iceberg. Nature, 2015, 520, 612-612.	27.8	157
31	Reproducible research can still be wrong: Adopting a prevention approach. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112, 1645-1646.	7.1	152
32	What is the question?. Science, 2015, 347, 1314-1315.	12.6	69
33	Ballgown bridges the gap between transcriptome assembly and expression analysis. Nature Biotechnology, 2015, 33, 243-246.	17.5	716
34	Polyester: simulating RNA-seq datasets with differential transcript expression. Bioinformatics, 2015, 31, 2778-2784.	4.1	250
35	Test set bias affects reproducibility of gene signatures. Bioinformatics, 2015, 31, 2318-2323.	4.1	90
36	Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. Nature Neuroscience, 2015, 18, 154-161.	14.8	142

#	Article	IF	Citations
37	regionReport: Interactive reports for region-based analyses. F1000Research, 2015, 4, 105.	1.6	5
38	regionReport: Interactive reports for region-level and feature-level genomic analyses. F1000Research, 2015, 4, 105.	1.6	4
39	svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Research, 2014, 42, e161-e161.	14.5	460
40	An estimate of the science-wise false discovery rate and application to the top medical literature. Biostatistics, 2014, 15, 1-12.	1.5	120
41	Differential expression analysis of RNA-seq data at single-base resolution. Biostatistics, 2014, 15, 413-426.	1.5	56
42	Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction. Bioinformatics, 2014, 30, 2757-2763.	4.1	102
43	Inflammatory Molecular Signature Associated With Infectious Agents in Psychosis. Schizophrenia Bulletin, 2014, 40, 963-972.	4.3	88
44	Removing batch effects for prediction problems with frozen surrogate variable analysis. PeerJ, 2014, 2, e561.	2.0	50
45	A randomized trial in a massive online open course shows people donâ€™t know what a statistically significant relationship looks like, but they can learn. PeerJ, 2014, 2, e589.	2.0	11
46	Measurement, Summary, and Methodological Variation in RNA-sequencing. , 2014, , 115-128.		0
47	SVAw - a web-based application tool for automated surrogate variable analysis of gene expression studies. Source Code for Biology and Medicine, 2013, 8, 8.	1.7	6
48	A simple and reproducible breast cancer prognostic test. BMC Genomics, 2013, 14, 336.	2.8	45
49	A Decisionâ€Theory Approach to Interpretable Set Analysis for Highâ€Dimensional Data. Biometrics, 2013, 69, 614-623.	1.4	5
50	Sequestration: inadvertently killing biomedical research to score political points. Genome Biology, 2013, 14, 109.	9.6	2
51	Gene set bagging for estimating the probability a statistically significant result will replicate. BMC Bioinformatics, 2013, 14, 360.	2.6	7
52	Significance analysis and statistical dissection of variably methylated regions. Biostatistics, 2012, 13, 166-178.	1.5	92
53	The <tt>sva</tt> package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics, 2012, 28, 882-883.	4.1	3,912
54	Keep a way open for tailored treatments. Nature, 2012, 484, 318-318.	27.8	5

#	Article	IF	Citations
55	The practical effect of batch on genomic prediction. Statistical Applications in Genetics and Molecular Biology, 2012, 11, Article 10.	0.6	25
56	Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. International Journal of Epidemiology, 2012, 41, 200-209.	1.9	567
57	A statistical approach to selecting and confirming validation targets in -omics experiments. BMC Bioinformatics, 2012, 13, 150.	2.6	13
58	Gene expression anti-profiles as a basis for accurate universal cancer signatures. BMC Bioinformatics, 2012, 13, 272.	2.6	41
59	Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature, 2011, 478, 519-523.	27.8	644
60	Sequencing technology does not eliminate biological variability. Nature Biotechnology, 2011, 29, 572-573.	17.5	193
61	Asymptotic Conditional Singular Value Decomposition for High-Dimensional Genomic Data. Biometrics, 2011, 67, 344-352.	1.4	47
62	ReCount: A multi-experiment resource of analysis-ready RNA-seq gene count datasets. BMC Bioinformatics, 2011, 12, 449.	2.6	144
63	The Joint Null Criterion for Multiple Hypothesis Tests. Statistical Applications in Genetics and Molecular Biology, 2011, 10, .	0.6	13
64	A computationally efficient modular optimal discovery procedure. Bioinformatics, 2011, 27, 509-515.	4.1	14
65	Dissecting Inflammatory Complications in Critically Injured Patients by Within-Patient Gene Expression Changes: A Longitudinal Clinical Genomics Study. PLoS Medicine, 2011, 8, e1001093.	8.4	51
66	Cooperation between Referees and Authors Increases Peer Review Accuracy. PLoS ONE, 2011, 6, e26895.	2.5	28
67	Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics, 2010, 11, 733-739.	16.3	1,641
68	Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biology, 2010, 11, R83.	9.6	268
69	The <tt>tspair</tt> package for finding top scoring pair classifiers in <tt>R</tt>. Bioinformatics, 2009, 25, 1203-1204.	4.1	31
70	Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature, 2009, 462, 358-362.	27.8	277
71	A general framework for multiple testing dependence. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105, 18718-18723.	7.1	302
72	Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genetics, 2007, 3, e161.	3.5	1,599

#	Article	IF	Citations
73	The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. Biostatistics, 2007, 8, 414-432.	1.5	125
74	On the design and analysis of gene expression studies in human populations. Nature Genetics, 2007, 39, 807-808.	21.4	121
75	EDGE: extraction and analysis of differential gene expression. Bioinformatics, 2006, 22, 507-508.	4.1	279
76	Significance analysis of time course microarray experiments. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102, 12837-12842.	7.1	534
77	Capturing Heterogeneity in Gene Expression Studies by "Surrogate Variable Analysis". PLoS Genetics, 2005, preprint, e161.	3.5	3
78	Strategies for cellular deconvolution in human brain RNA sequencing data. F1000Research, 0, 10, 750.	1.6	4