Jeffrey T Leek

List of Publications by Year
in descending order

78 papers	20,526 citations	66234 42 h-index	74018 75 g-index
106 all docs	106 docs citations	106 times ranked	35552 citing authors

#	Article	IF	Citations
1	Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols, 2016, 11, 1650-1667.	5.5	4,743
2	The <tt>sva</tt> package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics, 2012, 28, 882-883.	1.8	3,912
3	Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics, 2010, 11, 733-739.	7.7	1,641
4	Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genetics, 2007, 3, e161.	1.5	1,599
5	Ballgown bridges the gap between transcriptome assembly and expression analysis. Nature Biotechnology, 2015, 33, 243-246.	9.4	716
6	Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature, 2011, 478, 519-523.	13.7	644
7	Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. International Journal of Epidemiology, 2012, 41, 200-209.	0.9	567
8	Significance analysis of time course microarray experiments. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102, 12837-12842.	3.3	534
9	svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Research, 2014, 42, e161-e161.	6.5	460
10	Reproducible RNA-seq analysis using recount2. Nature Biotechnology, 2017, 35, 319-321.	9.4	395
11	A general framework for multiple testing dependence. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105, 18718-18723.	3.3	302
12	Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nature Neuroscience, 2018, 21, 1117-1125.	7.1	300
13	EDGE: extraction and analysis of differential gene expression. Bioinformatics, 2006, 22, 507-508.	1.8	279
14	Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature, 2009, 462, 358-362.	13.7	277
15	Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biology, 2010, 11, R83.	13.9	268
16	Polyester: simulating RNA-seq datasets with differential transcript expression. Bioinformatics, 2015, 31, 2778-2784.	1.8	250
17	Transparency and reproducibility in artificial intelligence. Nature, 2020, 586, E14-E16.	13.7	233
18	Sequencing technology does not eliminate biological variability. Nature Biotechnology, 2011, 29, 572-573.	9.4	193

#	Article	IF	Citations
19	What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. Perspectives on Psychological Science, 2016, 11, 539-544.	5.2	168
20	Statistics: P values are just the tip of the iceberg. Nature, 2015, 520, 612-612.	13.7	157
21	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.	3.3	152
22	ReCount: A multi-experiment resource of analysis-ready RNA-seq gene count datasets. BMC Bioinformatics, 2011, 12, 449.	1.2	144
23	Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. Nature Neuroscience, 2015, 18, 154-161.	7.1	142
24	The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. Biostatistics, 2007, 8, 414-432.	0.9	125
25	On the design and analysis of gene expression studies in human populations. Nature Genetics, 2007, 39, 807-808.	9.4	121
26	An estimate of the science-wise false discovery rate and application to the top medical literature. Biostatistics, 2014, 15, 1-12.	0.9	120
27	recount3: summaries and queries for large-scale RNA-seq expression and splicing. Genome Biology, 2021, 22, 323.	3.8	103
28	Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction. Bioinformatics, 2014, 30, 2757-2763.	1.8	102
29	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.	3.3	95
30	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.	3.8	94
31	Significance analysis and statistical dissection of variably methylated regions. Biostatistics, 2012, 13, 166-178.	0.9	92
32	Test set bias affects reproducibility of gene signatures. Bioinformatics, 2015, 31, 2318-2323.	1.8	90
33	Inflammatory Molecular Signature Associated With Infectious Agents in Psychosis. Schizophrenia Bulletin, 2014, 40, 963-972.	2.3	88
34	Evolution of cellular morpho-phenotypes in cancer metastasis. Scientific Reports, 2016, 5, 18437.	1.6	81
35	What is the question?. Science, 2015, 347, 1314-1315.	6.0	69
36	Addressing confounding artifacts in reconstruction of gene co-expression networks. Genome Biology, 2019, 20, 94.	3.8	68

#	Article	IF	Citations
37	A direct approach to estimating false discovery rates conditional on covariates. PeerJ, 2018, 6, e6035.	0.9	60
38	Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space. Cell Genomics, 2022, 2, 100085.	3.0	59
39	Rail-RNA: scalable analysis of RNA-seq splicing and coverage. Bioinformatics, 2017, 33, 4033-4040.	1.8	57
40	Differential expression analysis of RNA-seq data at single-base resolution. Biostatistics, 2014, 15, 413-426.	0.9	56
41	Flexible expressed region analysis for RNA-seq with<tt>derfinder</tt>. Nucleic Acids Research, 2017, 45, e9-e9.	6.5	54
42	Practical impacts of genomic data â€œcleaningâ€•on biological discovery using surrogate variable analysis. BMC Bioinformatics, 2015, 16, 372.	1.2	51
43	Dissecting Inflammatory Complications in Critically Injured Patients by Within-Patient Gene Expression Changes: A Longitudinal Clinical Genomics Study. PLoS Medicine, 2011, 8, e1001093.	3.9	51
44	BatchQC: interactive software for evaluating sample and batch effects in genomic data. Bioinformatics, 2016, 32, 3836-3838.	1.8	50
45	Removing batch effects for prediction problems with frozen surrogate variable analysis. PeerJ, 2014, 2, e561.	0.9	50
46	Improving the value of public RNA-seq expression data by phenotype prediction. Nucleic Acids Research, 2018, 46, e54-e54.	6.5	49
47	Asymptotic Conditional Singular Value Decomposition for High-Dimensional Genomic Data. Biometrics, 2011, 67, 344-352.	0.8	47
48	A simple and reproducible breast cancer prognostic test. BMC Genomics, 2013, 14, 336.	1.2	45
49	Is Most Published Research Really False?. Annual Review of Statistics and Its Application, 2017, 4, 109-122.	4.1	44
50	Gene expression anti-profiles as a basis for accurate universal cancer signatures. BMC Bioinformatics, 2012, 13, 272.	1.2	41
51	Recounting the FANTOM CAGE-Associated Transcriptome. Genome Research, 2020, 30, 1073-1081.	2.4	35
52	The <tt>tspair</tt> package for finding top scoring pair classifiers in <tt>R</tt>. Bioinformatics, 2009, 25, 1203-1204.	1.8	31
53	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.	3.3	28
54	Cooperation between Referees and Authors Increases Peer Review Accuracy. PLoS ONE, 2011, 6, e26895.	1.1	28

#	Article	IF	Citations
55	A visual tool for defining reproducibility and replicability. Nature Human Behaviour, 2019, 3, 650-652.	6.2	26
56	The practical effect of batch on genomic prediction. Statistical Applications in Genetics and Molecular Biology, 2012, 11, Article 10.	0.2	25
57	The Democratization of Data Science Education. American Statistician, 2020, 74, 1-7.	0.9	21
58	Transcriptional profile of platelets and iPSC-derived megakaryocytes from whole-genome and RNA sequencing. Blood, 2021, 137, 959-968.	0.6	21
59	A computationally efficient modular optimal discovery procedure. Bioinformatics, 2011, 27, 509-515.	1.8	14
60	How to Share Data for Collaboration. American Statistician, 2018, 72, 53-57.	0.9	14
61	The Joint Null Criterion for Multiple Hypothesis Tests. Statistical Applications in Genetics and Molecular Biology, 2011, 10, .	0.2	13
62	A statistical approach to selecting and confirming validation targets in -omics experiments. BMC Bioinformatics, 2012, 13, 150.	1.2	13
63	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.	0.9	11
64	Gene set bagging for estimating the probability a statistically significant result will replicate. BMC Bioinformatics, 2013, 14, 360.	1.2	7
65	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
66	Gene and protein expression in human megakaryocytes derived from induced pluripotent stem cells. Journal of Thrombosis and Haemostasis, 2021, 19, 1783-1799.	1.9	6
67	Keep a way open for tailored treatments. Nature, 2012, 484, 318-318.	13.7	5
68	A Decisionâ€Theory Approach to Interpretable Set Analysis for Highâ€Dimensional Data. Biometrics, 2013, 69, 614-623.	0.8	5
69	Rail-dbGaP: analyzing dbGaP-protected data in the cloud with Amazon Elastic MapReduce. Bioinformatics, 2016, 32, 2551-2553.	1.8	5
70	regionReport: Interactive reports for region-based analyses. F1000Research, 2015, 4, 105.	0.8	5
71	Strategies for cellular deconvolution in human brain RNA sequencing data. F1000Research, 0, 10, 750.	0.8	4
72	regionReport: Interactive reports for region-level and feature-level genomic analyses. F1000Research, 2015, 4, 105.	0.8	4

#	Article	IF	Citations
73	Capturing Heterogeneity in Gene Expression Studies by "Surrogate Variable Analysis". PLoS Genetics, 2005, preprint, e161.	1.5	3
74	Diagnosing Data Analytic Problems in the Classroom. Journal of Statistics and Data Science Education, 2021, 29, 267-276.	0.9	3
75	Sequestration: inadvertently killing biomedical research to score political points. Genome Biology, 2013, 14, 109.	13.9	2
76	Genomic and clinical predictors for improving estimator precision in randomized trials of breast cancer treatments. Contemporary Clinical Trials Communications, 2016, 3, 48-54.	0.5	1
77	Discussion of â€œvisualizing statistical models: Removing the blindfoldâ€• Statistical Analysis and Data Mining, 2015, 8, 240-241.	1.4	0
78	Measurement, Summary, and Methodological Variation in RNA-sequencing. , 2014, , 115-128.		0