# Sandrine Dudoit

## List of Publications by Citations

| | | | |
|---|---|---|---|
| **84** papers | **26,281** citations | **44** h-index | **94** g-index |
| **94** ext. papers | **30,753** ext. citations | **9.7** avg, IF | **6.69** L-index |

| # | Paper | IF | Citations |
|---|---|---|---|
| 84 | Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, **2004**, 5, R80 | 18.3 | 9148 |
| 83 | Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, **2002**, 30, e15 | 20.1 | 2571 |
| 82 | Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data. *Journal of the American Statistical Association*, **2002**, 97, 77-87 | 2.8 | 1778 |
| 81 | Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, **2010**, 11, 94 | 3.6 | 1132 |
| 80 | The developmental transcriptome of Drosophila melanogaster. *Nature*, **2011**, 471, 473-9 | 50.4 | 1094 |
| 79 | Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, **2014**, 32, 896-902 | 44.5 | 910 |
| 78 | Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*, **2010**, 330, 1787-97 | 33.3 | 892 |
| 77 | Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America*, **2002**, 99, 12877-82 | 11.5 | 852 |
| 76 | Multiple Hypothesis Testing in Microarray Experiments. *Statistical Science*, **2003**, 18, 71 | 2.4 | 614 |
| 75 | Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, **2018**, 19, 477 | 4.5 | 591 |
| 74 | Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Science Advances*, **2020**, 6, | 14.3 | 514 |
| 73 | Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*, **2010**, 38, e131 | 20.1 | 471 |
| 72 | GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, **2011**, 12, 480 | 3.6 | 458 |
| 71 | Survival ensembles. *Biostatistics*, **2006**, 7, 355-73 | 3.7 | 403 |
| 70 | A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, **2002**, 3, RESEARCH0036 | 18.3 | 384 |
| 69 | Resampling-based multiple testing for microarray data analysis. *Test*, **2003**, 12, 1-77 | 1.1 | 366 |
| 68 | Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, **2003**, 19, 1090-9 | 7.2 | 344 |

| | | | |
|---|---|---|---|
| 67 | Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **2002**, 99, 972-7 | 11.5 | 337 |
| 66 | A general and flexible method for signal extraction from single-cell RNA-seq data. *Nature Communications*, **2018**, 9, 284 | 17.4 | 306 |
| 65 | Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nature Methods*, **2017**, 14, 565-571 | 21.6 | 255 |
| 64 | Comparison of Methods for Image Analysis on cDNA Microarray Data. *Journal of Computational and Graphical Statistics*, **2002**, 11, 108-136 | 1.4 | 243 |
| 63 | Stage II colon cancer prognosis prediction by tumor gene expression profiling. *Journal of Clinical Oncology*, **2006**, 24, 4685-91 | 2.2 | 173 |
| 62 | Open Source Software for the Analysis of Microarray Data. *BioTechniques*, **2003**, 34, S45-S51 | 2.5 | 172 |
| 61 | Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Research*, **2011** , 21, 193-202 | 9.7 | 165 |
| 60 | Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell*, **2017**, 20, 817-830.e8 | 18.3 | 104 |
| 59 | Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biology*, **2018**, 19, 24 | 18.3 | 101 |
| 58 | Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, **2005**, 2, 131-154 | | 93 |
| 57 | Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications*, **2020**, 11, 1201 | 17.4 | 86 |
| 56 | Multiple testing. Part I. Single-step procedures for control of general type I error rates. *Statistical Applications in Genetics and Molecular Biology*, **2004**, 3, Article13 | 1.2 | 84 |
| 55 | Colon cancer prognosis prediction by gene expression profiling. *Oncogene*, **2005**, 24, 6155-64 | 9.2 | 83 |
| 54 | Polygenic and directional regulatory evolution across pathways in Saccharomyces. *Proceedings of the National Academy of Sciences of the United States of America*, **2010**, 107, 5058-63 | 11.5 | 76 |
| 53 | Exploration of global gene expression in human liver steatosis by high-density oligonucleotide microarray. *Laboratory Investigation*, **2006**, 86, 154-65 | 5.9 | 76 |
| 52 | Diverse transcriptional programs associated with environmental stress and hormones in the Arabidopsis receptor-like kinase gene family. *Molecular Plant*, **2009**, 2, 84-107 | 14.4 | 71 |
| 51 | A general framework for statistical performance comparison of evolutionary computation algorithms. *Information Sciences*, **2008**, 178, 2870-2879 | 7.7 | 71 |
| 50 | Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in Drosophila. *PLoS Genetics*, **2009**, 5, e1000525 | 6 | 68 |

49 — Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq. *Cell Systems*, **2019**, 8, 315-328.e8 — 10.6 — 66

48 — Oracle inequalities for multi-fold cross validation. *Statistics & Risk Modeling*, **2006**, 24, 351-371 — 66

47 — Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, **2004**, 3, Article15 — 1.2 — 62

46 — Ischemic preconditioning modulates the expression of several genes, leading to the overproduction of IL-1Ra, iNOS, and Bcl-2 in a human model of liver ischemia-reperfusion. *FASEB Journal*, **2005**, 19, 1617-26 — 0.9 — 60

45 — More power via graph-structured tests for differential expression of gene networks. *Annals of Applied Statistics*, **2012**, 6, — 2.1 — 57

44 — Asymptotic optimality of likelihood-based cross-validation. *Statistical Applications in Genetics and Molecular Biology*, **2004**, 3, Article4 — 1.2 — 56

43 — The establishment of gene silencing at single-cell resolution. *Nature Genetics*, **2009**, 41, 800-6 — 36.3 — 53

42 — The cross-validated adaptive epsilon-net estimator. *Statistics & Risk Modeling*, **2006**, 24, 373-395 — 52

41 — Tree-based multivariate regression and density estimation with right-censored data. *Journal of Multivariate Analysis*, **2004**, 90, 154-177 — 1.4 — 45

40 — A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*, **2021**, 598, 86-102 — 50.4 — 44

39 — Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Statistical Applications in Genetics and Molecular Biology*, **2004**, 3, Article14 — 1.2 — 43

38 — Injury Activates Transient Olfactory Stem Cell States with Diverse Lineage Capacities. *Cell Stem Cell*, **2017**, 21, 775-790.e9 — 18 — 42

37 — Silencing of odorant receptor genes by G protein βγ signaling ensures the expression of one odorant receptor per olfactory sensory neuron. *Neuron*, **2014**, 81, 847-59 — 13.9 — 41

36 — A score test for the linkage analysis of qualitative and quantitative traits based on identity by descent data from sib-pairs. *Biostatistics*, **2000**, 1, 1-26 — 3.7 — 41

35 — An untargeted metabolomics method for archived newborn dried blood spots in epidemiologic studies. *Metabolomics*, **2017**, 13, 1 — 4.7 — 38

34 — Multiple testing methods for ChIP-Chip high density oligonucleotide array data. *Journal of Computational Biology*, **2006**, 13, 579-613 — 1.7 — 34

33 — Novel low abundance and transient RNAs in yeast revealed by tiling microarrays and ultra high-throughput sequencing are not conserved across closely related yeast species. *PLoS Genetics*, **2008**, 4, e1000299 — 6 — 29

32 — clusterExperiment and RSEC: A Bioconductor package and framework for clustering of single-cell and other large gene expression datasets. *PLoS Computational Biology*, **2018**, 14, e1006378 — 5 — 26

| | | | |
|---|---|---|---|
| 31 | Gene expression profiling of nonneoplastic mucosa may predict clinical outcome of colon cancer patients. *Diseases of the Colon and Rectum*, **2005**, 48, 2238-48 | 3.1 | 25 |
| 30 | A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature*, **2021**, 598, 103-110 | 50.4 | 23 |
| 29 | An integrated transcriptomic and epigenomic atlas of mouse primary motor cortex cell types | | 23 |
| 28 | Metabolomics of neonatal blood spots reveal distinct phenotypes of pediatric acute lymphoblastic leukemia and potential effects of early-life nutrition. *Cancer Letters*, **2019**, 452, 71-78 | 9.9 | 21 |
| 27 | Resampling-based empirical Bayes multiple testing procedures for controlling generalized tail probability and expected value error rates: focus on the false discovery rate and simulation study. *Biometrical Journal*, **2008**, 50, 716-44 | 1.5 | 20 |
| 26 | ZINB-WaVE: A general and flexible method for signal extraction from single-cell RNA-seq data | | 19 |
| 25 | Untargeted adductomics of Cys34 modifications to human serum albumin in newborn dried blood spots. *Analytical and Bioanalytical Chemistry*, **2019**, 411, 2351-2362 | 4.4 | 19 |
| 24 | Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-seq | | 16 |
| 23 | Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics | | 15 |
| 22 | Untargeted lipidomic features associated with colorectal cancer in a prospective cohort. *BMC Cancer*, **2018**, 18, 996 | 4.8 | 13 |
| 21 | A fine-scale linkage-disequilibrium measure based on length of haplotype sharing. *American Journal of Human Genetics*, **2006**, 78, 615-28 | 11 | 12 |
| 20 | Microarray Expression Profiling Identifies Genes with Altered Expression in HDL-Deficient Mice. *Genome Research*, **2000**, 10, 2022-2029 | 9.7 | 12 |
| 19 | A multimodal cell census and atlas of the mammalian primary motor cortex | | 12 |
| 18 | Bioconductor workflow for single-cell RNA sequencing: Normalization, dimensionality reduction, clustering, and lineage inference. *F1000Research*, **2017**, 6, 1158 | 3.6 | 11 |
| 17 | Supervised detection of regulatory motifs in DNA sequences. *Statistical Applications in Genetics and Molecular Biology*, **2003**, 2, Article5 | 1.2 | 10 |
| 16 | Cys34 Adductomics Links Colorectal Cancer with the Gut Microbiota and Redox Biology. *Cancer Research*, **2019**, 79, 6024-6031 | 10.1 | 10 |
| 15 | Comparing Segmentation Methods for Genome Annotation Based on RNA-Seq Data. *Journal of Agricultural, Biological, and Environmental Statistics*, **2014**, 19, 101-118 | 1.9 | 9 |
| 14 | Loss-based estimation with cross-validation. *SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining*, **2003**, 5, 56-68 | 4.6 | 9 |

| 13 | Asymptotically optimal model selection method with right censored outcomes. *Bernoulli*, **2004**, 10, 1011 | 1.6 | 9 |

| 12 | Trajectory-based differential expression analysis for single-cell sequencing data | | 9 |

| 11 | Untargeted adductomics of newborn dried blood spots identifies modifications to human serum albumin associated with childhood leukemia. *Leukemia Research*, **2020**, 88, 106268 | 2.7 | 9 |

| 10 | A deletion/substitution/addition algorithm for classification neural networks, with applications to biomedical data. *Journal of Statistical Planning and Inference*, **2008**, 138, 464-488 | 0.8 | 7 |

| 9 | Power and robustness of a score test for linkage analysis of quantitative traits using identity by descent data on sib pairs. *Genetic Epidemiology*, **2001**, 20, 415-31 | 2.6 | 6 |

| 8 | Untargeted metabolomics of newborn dried blood spots reveals sex-specific associations with pediatric acute myeloid leukemia. *Leukemia Research*, **2021**, 106, 106585 | 2.7 | 4 |

| 7 | Power of a score test for quantitative trait linkage analysis of relative pairs. *Genetic Epidemiology*, **2000**, 19 Suppl 1, S85-91 | 2.6 | 3 |

| 6 | clusterExperiment and RSEC: A Bioconductor package and framework for clustering of single-cell and other large gene expression datasets | | 2 |

| 5 | Exploring High-Dimensional Biological Data with Sparse Contrastive Principal Component Analysis | | 1 |

| 4 | Klf5 establishes bi-potential cell fate by dual regulation of ICM and TE specification genes. *Cell Reports*, **2021**, 37, 109982 | 10.6 | 0 |

| 3 | cvCovEst: Cross-validated covariance matrix estimator selection and evaluation in R. *Journal of Open Source Software*, **2021**, 6, 3273 | 5.2 | 0 |

| 2 | Loss-Based Estimation with Evolutionary Algorithms and Cross-Validation. *Adaptation, Learning, and Optimization*, **2010**, 453-484 | 0.7 | |

| 1 | scPCA: A toolbox for sparse contrastive principal component analysis in R. *Journal of Open Source Software*, **2020**, 5, 2079 | 5.2 | |