

Sayan Mukherjee

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/5609377/publications.pdf>

Version: 2024-02-01

87
papers

46,480
citations

70961

41
h-index

54797

84
g-index

101
all docs

101
docs citations

101
times ranked

85236
citing authors

#	ARTICLE	IF	CITATIONS
1	Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102, 15545-15550.	3.3	38,922
2	Choosing Multiple Parameters for Support Vector Machines. Machine Learning, 2002, 46, 131-159.	3.4	1,746
3	A Genomic Strategy to Refine Prognosis in Early-Stage Non-Small-Cell Lung Cancer. New England Journal of Medicine, 2006, 355, 570-580.	13.9	577
4	An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. Nature Genetics, 2005, 37, 48-55.	9.4	392
5	Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. American Journal of Human Genetics, 2016, 98, 456-472.	2.6	335
6	A phylogenetic transform enhances analysis of compositional microbiota data. ELife, 2017, 6, .	2.8	247
7	Estimating Dataset Size Requirements for Classifying DNA Microarray Data. Journal of Computational Biology, 2003, 10, 119-142.	0.8	244
8	Gene expression changes and molecular pathways mediating activity-dependent plasticity in visual cortex. Nature Neuroscience, 2006, 9, 660-668.	7.1	199
9	General conditions for predictivity in learning theory. Nature, 2004, 428, 419-422.	13.7	195
10	Optimal gene expression analysis by microarrays. Cancer Cell, 2002, 2, 353-361.	7.7	149
11	Probability measures on the space of persistence diagrams. Inverse Problems, 2011, 27, 124007.	1.0	132
12	Frochet Means for Distributions of Persistence Diagrams. Discrete and Computational Geometry, 2014, 52, 44-70.	0.4	118
13	Learning theory: stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization. Advances in Computational Mathematics, 2006, 25, 161-193.	0.8	112
14	Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets. PeerJ, 2017, 5, e2969.	0.9	105
15	A New Fully Automated Approach for Aligning and Comparing Shapes. Anatomical Record, 2015, 298, 249-276.	0.8	104
16	Comparative study of gene set enrichment methods. BMC Bioinformatics, 2009, 10, 275.	1.2	102
17	Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLoS Genetics, 2017, 13, e1006869.	1.5	100
18	Core and region-enriched networks of behaviorally regulated genes and the singing genome. Science, 2014, 346, 1256780.	6.0	97

#	ARTICLE	IF	CITATIONS
19	Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. <i>Genome Research</i> , 2012, 22, 386-397.	2.4	90
20	Naught all zeros in sequence count data are the same. <i>Computational and Structural Biotechnology Journal</i> , 2020, 18, 2789-2798.	1.9	87
21	Persistent homology transform for modeling shapes and surfaces. <i>Information and Inference</i> , 2014, 3, 310-344.	0.9	85
22	Evidence of Influence of Genomic DNA Sequence on Human X Chromosome Inactivation. <i>PLoS Computational Biology</i> , 2006, 2, e113.	1.5	84
23	An Integrated Approach to the Prediction of Chemotherapeutic Response in Patients with Breast Cancer. <i>PLoS ONE</i> , 2008, 3, e1908.	1.1	82
24	Age-Specific Differences in Oncogenic Pathway Deregulation Seen in Human Breast Tumors. <i>PLoS ONE</i> , 2008, 3, e1373.	1.1	81
25	Evidence-ranked motif identification. <i>Genome Biology</i> , 2010, 11, R19.	13.9	77
26	Melanoma Therapeutic Strategies that Select against Resistance by Exploiting MYC-Driven Evolutionary Convergence. <i>Cell Reports</i> , 2017, 21, 2796-2812.	2.9	77
27	Efficient Genome-Wide Sequencing and Low-Coverage Pedigree Analysis from Noninvasively Collected Samples. <i>Genetics</i> , 2016, 203, 699-714.	1.2	76
28	Genetics of gene expression responses to temperature stress in a sea urchin gene network. <i>Molecular Ecology</i> , 2012, 21, 4547-4562.	2.0	74
29	Cross Species Genomic Analysis Identifies a Mouse Model as Undifferentiated Pleomorphic Sarcoma/Malignant Fibrous Histiocytoma. <i>PLoS ONE</i> , 2009, 4, e8075.	1.1	71
30	Permutation Tests for Classification. <i>Lecture Notes in Computer Science</i> , 2005, , 501-515.	1.0	70
31	Cyclin-Dependent Kinases Are Regulators and Effectors of Oscillations Driven by a Transcription Factor Network. <i>Molecular Cell</i> , 2012, 45, 669-679.	4.5	66
32	Dissecting High-Dimensional Phenotypes with Bayesian Sparse Factor Analysis of Genetic Covariance Matrices. <i>Genetics</i> , 2013, 194, 753-767.	1.2	61
33	Modeling Cancer Progression via Pathway Dependencies. <i>PLoS Computational Biology</i> , 2008, 4, e28.	1.5	60
34	Differential expression analysis for RNAseq using Poisson mixed models. <i>Nucleic Acids Research</i> , 2017, 45, e106-e106.	6.5	60
35	Predicting Clinical Outcomes in Glioblastoma: An Application of Topological and Functional Data Analysis. <i>Journal of the American Statistical Association</i> , 2020, 115, 1139-1150.	1.8	60
36	Analysis of sample set enrichment scores: assaying the enrichment of sets of genes for individual samples in genome-wide expression profiles. <i>Bioinformatics</i> , 2006, 22, e108-e116.	1.8	58

#	ARTICLE	IF	CITATIONS
37	Citizen Science as a New Tool in Dog Cognition Research. PLoS ONE, 2015, 10, e0135176.	1.1	57
38	Gene Expression Programs of Human Smooth Muscle Cells: Tissue-Specific Differentiation and Prognostic Significance in Breast Cancers. PLoS Genetics, 2007, 3, e164.	1.5	56
39	Dynamic linear models guide design and analysis of microbiota studies within artificial human guts. Microbiome, 2018, 6, 202.	4.9	54
40	GSAASeqSP: A Toolset for Gene Set Association Analysis of RNA-Seq Data. Scientific Reports, 2014, 4, 6347.	1.6	53
41	Phylofactorization: a graph partitioning algorithm to identify phylogenetic scales of ecological data. Ecological Monographs, 2019, 89, e01353.	2.4	52
42	Characterizing the developmental pathways <i>TTF-1</i> , <i>NKX2-8</i> , and <i>PAX9</i> in lung cancer. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106, 5312-5317.	3.3	50
43	Measuring and mitigating PCR bias in microbiota datasets. PLoS Computational Biology, 2021, 17, e1009113.	1.5	43
44	Estimating Rates of Progression and Predicting Future Visual Fields in Glaucoma Using a Deep Variational Autoencoder. Scientific Reports, 2019, 9, 18113.	1.6	42
45	Random walks on simplicial complexes and harmonics. Random Structures and Algorithms, 2016, 49, 379-405.	0.6	39
46	Risk bounds for mixture density estimation. ESAIM - Probability and Statistics, 2005, 9, 220-229.	0.2	37
47	The topology of probability distributions on manifolds. Probability Theory and Related Fields, 2015, 161, 651-686.	0.9	37
48	The Information Geometry of Mirror Descent. IEEE Transactions on Information Theory, 2015, 61, 1451-1457.	1.5	33
49	Bayesian Approximate Kernel Regression With Variable Selection. Journal of the American Statistical Association, 2018, 113, 1710-1721.	1.8	33
50	Assessing the Radiation Response of Lung Cancer with Different Gene Mutations Using Genetically Engineered Mice. Frontiers in Oncology, 2013, 3, 72.	1.3	32
51	Genetic Effects on Mating Success and Partner Choice in a Social Mammal. American Naturalist, 2012, 180, 113-129.	1.0	31
52	Topological consistency via kernel estimation. Bernoulli, 2017, 23, .	0.7	30
53	Learning gradients on manifolds. Bernoulli, 2010, 16, .	0.7	28
54	The Use of Unlabeled Data in Predictive Modeling. Statistical Science, 2007, 22, 189.	1.6	27

#	ARTICLE	IF	CITATIONS
55	A Cheeger-type inequality on simplicial complexes. <i>Advances in Applied Mathematics</i> , 2014, 56, 56-77.	0.4	27
56	Development and Assessment of Fully Automated and Globally Transitive Geometric Morphometric Methods, With Application to a Biological Comparative Dataset With High Interspecific Variation. <i>Anatomical Record</i> , 2018, 301, 636-658.	0.8	25
57	DNase-seq predicts regions of rotational nucleosome stability across diverse human cell types. <i>Genome Research</i> , 2013, 23, 1118-1129.	2.4	22
58	Local Homology Transfer and Stratification Learning. , 2012, , .		21
59	A unifying framework for interpreting and predicting mutualistic systems. <i>Nature Communications</i> , 2019, 10, 242.	5.8	21
60	Statistical inference for dynamical systems: A review. <i>Statistics Surveys</i> , 2015, 9, .	7.3	19
61	Impact of anxiety and depression on progression to glaucoma among glaucoma suspects. <i>British Journal of Ophthalmology</i> , 2021, 105, 1244-1249.	2.1	19
62	Statistical Analysis of Crystallization Database Links Protein Physico-Chemical Features with Crystallization Mechanisms. <i>PLoS ONE</i> , 2014, 9, e101123.	1.1	18
63	Synchrony and idiosyncrasy in the gut microbiome of wild baboons. <i>Nature Ecology and Evolution</i> , 2022, 6, 955-964.	3.4	18
64	Partial Factor Modeling: Predictor-Dependent Shrinkage for Linear Regression. <i>Journal of the American Statistical Association</i> , 2013, 108, 999-1008.	1.8	14
65	Novel Distal eQTL Analysis Demonstrates Effect of Population Genetic Architecture on Detecting and Interpreting Associations. <i>Genetics</i> , 2014, 198, 879-893.	1.2	14
66	On the reproducibility of results of pathway analysis in genome-wide expression studies of colorectal cancers. <i>Journal of Biomedical Informatics</i> , 2010, 43, 397-406.	2.5	12
67	A comparative study of covariance selection models for the inference of gene regulatory networks. <i>Journal of Biomedical Informatics</i> , 2013, 46, 894-904.	2.5	12
68	Genomic Features That Predict Allelic Imbalance in Humans Suggest Patterns of Constraint on Gene Expression Variation. <i>Molecular Biology and Evolution</i> , 2009, 26, 2047-2059.	3.5	9
69	Localized Sliced Inverse Regression. <i>Journal of Computational and Graphical Statistics</i> , 2010, 19, 843-860.	0.9	9
70	HOMINID: a framework for identifying associations between host genetic variation and microbiome composition. <i>GigaScience</i> , 2017, 6, 1-7.	3.3	9
71	A statistical pipeline for identifying physical features that differentiate classes of 3D shapes. <i>Annals of Applied Statistics</i> , 2021, 15, .	0.5	8
72	A topological data analytic approach for discovering biophysical signatures in protein dynamics. <i>PLoS Computational Biology</i> , 2022, 18, e1010045.	1.5	7

#	ARTICLE	IF	CITATIONS
73	Estimating variable structure and dependence in multitask learning via gradients. Machine Learning, 2011, 83, 265-287.	3.4	6
74	Kernel Sliced Inverse Regression: Regularization and Consistency. Abstract and Applied Analysis, 2013, 2013, 1-11.	0.3	6
75	A Digital Network Approach to Infer Sex Behavior in Emerging HIV Epidemics. PLoS ONE, 2014, 9, e101416.	1.1	6
76	Geometric Representations of Random Hypergraphs. Journal of the American Statistical Association, 2017, 112, 363-383.	1.8	6
77	Discovering genetic variants in Crohn's disease by exploring genomic regions enriched of weak association signals. Digestive and Liver Disease, 2011, 43, 623-631.	0.4	5
78	A Predictive Framework for Integrating Disparate Genomic Data Types Using Sample-Specific Gene Set Enrichment Analysis and Multi-Task Learning. PLoS ONE, 2012, 7, e44635.	1.1	5
79	The Geometry of Synchronization Problems and Learning Group Actions. Discrete and Computational Geometry, 2021, 65, 150-211.	0.4	4
80	Likelihood ratio statistics for gene set enrichment in Alzheimer's disease pathways. Alzheimer's and Dementia, 2021, 17, 561-573.	0.4	4
81	RS-SNP: a random-set method for genome-wide association studies. BMC Genomics, 2011, 12, 166.	1.2	1
82	Fast Moment Estimation for Generalized Latent Dirichlet Models. Journal of the American Statistical Association, 2018, 113, 1528-1540.	1.8	1
83	Label propagation defines signaling networks associated with recurrently mutated cancer genes. Scientific Reports, 2019, 9, 9401.	1.6	1
84	Bayesian Non-Parametric Factor Analysis for Longitudinal Spatial Surfaces. Bayesian Analysis, 2021, -1, .	1.6	1
85	Concerning Two Classes of Non-Diophantine Arithmetics. , 2022, 81, .		1
86	Response to Shen et al.. American Journal of Human Genetics, 2016, 99, 1220-1221.	2.6	0
87	Learning Subspaces of Different Dimensions. Journal of Computational and Graphical Statistics, 0, , 1-35.	0.9	0