Hongzhe Li

List of Publications by Year in Descending Order

119 papers	**9,999** citations	**38** h-index	**99** g-index
126 ext. papers	**12,597** ext. citations	**7.1** avg, IF	**6.2** L-index

#	Paper	IF	Citations
119	Transfer Learning for High-Dimensional Linear Regression: Prediction, Estimation and Minimax Optimality.. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2022**, 84, 149-173	3.9	3
118	Deep learning to predict the biosynthetic gene clusters in bacterial genomes.. *Journal of Molecular Biology*, **2022**, 167597	6.5	0
117	Optimal Permutation Recovery in Permuted Monotone Matrix Model. *Journal of the American Statistical Association*, **2021**, 116, 1358-1372	2.8	5
116	Statistical and Computational Methods for Analysis of Shotgun Metagenomics Sequencing Data. *Frontiers in Probability and the Statistical Sciences*, **2021**, 27-43		
115	MiRKAT: kernel machine regression-based global association tests for the microbiome. *Bioinformatics*, **2021**, 37, 1595-1597	7.2	3
114	Role of dietary fiber in the recovery of the human gut microbiome and its metabolome. *Cell Host and Microbe*, **2021**, 29, 394-407.e5	23.4	32
113	Menin-regulated Pbk controls high fat diet-induced compensatory beta cell proliferation. *EMBO Molecular Medicine*, **2021**, 13, e13524	12	2
112	Dietary Patterns and Growth From 12 to 24 Months of Age in African American Infants. *Current Developments in Nutrition*, **2021**, 5, 454-454	0.4	78
111	Hypothesis testing for phylogenetic composition: a minimum-cost flow perspective. *Biometrika*, **2021**, 108, 17-36	2	3
110	Optimal estimation of bacterial growth rates based on a permuted monotone matrix. *Biometrika*, **2021**, 108, 693-708	2	0
109	IFAA: Robust Association Identification and Inference for Absolute Abundance in Microbiome Analyses.. *Journal of the American Statistical Association*, **2021**, 116, 1595-1608	2.8	2
108	Global and Simultaneous Hypothesis Testing for High-Dimensional Logistic Regression Models. *Journal of the American Statistical Association*, **2021**, 116, 984-998	2.8	7
107	Maternal gut microbiota reflecting poor diet quality is associated with spontaneous preterm birth in a prospective cohort study. *American Journal of Clinical Nutrition*, **2021**, 113, 602-611	7	8
106	The Microbiome and p-Inulin in Hemodialysis: A Feasibility Study.. *Kidney360*, **2021**, 2, 445-455	1.8	0
105	A Compositional Mediation Model for Binary Outcome: Application to Microbiome Studies. *Bioinformatics*, **2021**,	7.2	2
104	A Randomized Trial Comparing the Specific Carbohydrate Diet to a Mediterranean Diet in Adults With Crohn's Disease. *Gastroenterology*, **2021**, 161, 837-852.e9	13.3	19
103	Mitochondrial dysfunction in inflammatory bowel disease alters intestinal epithelial metabolism of hepatic acylcarnitines. *Journal of Clinical Investigation*, **2021**, 131,	15.9	11

102 | Bayesian Balance-Regression in Microbiome Studies Using Stochastic Search **2021**, 347-362 | | 1

101 | INFORMATION CONTENT OF HIGH-ORDER ASSOCIATIONS OF THE HUMAN GUT MICROBIOTA NETWORK.. *Annals of Applied Statistics*, **2021**, 15, 1788-1807 | 2.1 |

100 | Emerging Priorities for Microbiome Research. *Frontiers in Microbiology*, **2020**, 11, 136 | 5.7 | 50

99 | Bacterial colonization reprograms the neonatal gut metabolome. *Nature Microbiology*, **2020**, 5, 838-847 | 26.6 | 37

98 | Castleman disease spectrum.. *Journal of Clinical Oncology*, **2020**, 38, 8548-8548 | 2.2 |

97 | Systematic integrated analysis of genetic and epigenetic variation in diabetic kidney disease. *Proceedings of the National Academy of Sciences of the United States of America*, **2020**, 117, 29013-29024 | 11.5 | 11

96 | Multisample estimation of bacterial composition matrices in metagenomics data. *Biometrika*, **2020**, 107, 75-92 | 2 | 15

95 | Optimal Estimation of Wasserstein Distance on a Tree With an Application to Microbiome Studies. *Journal of the American Statistical Association*, **2020**, 1-17 | 2.8 | 1

94 | Multi-omic Analysis of the Interaction between Clostridioides difficile Infection and Pediatric Inflammatory Bowel Disease. *Cell Host and Microbe*, **2020**, 28, 422-433.e7 | 23.4 | 20

93 | A Zero-Inflated Latent Dirichlet Allocation Model for Microbiome Studies. *Frontiers in Genetics*, **2020**, 11, 602594 | 4.5 | 2

92 | Estimation of Heterogeneous Restricted Mean Survival Time Using Random Forest. *Frontiers in Genetics*, **2020**, 11, 587378 | 4.5 |

91 | An Integrated Gaussian Graphical Model to evaluate the impact of exposures on metabolic networks. *Computers in Biology and Medicine*, **2019**, 114, 103417 | 7 | 2

90 | Pathologic gene network rewiring implicates PPP1R3A as a central regulator in pressure overload heart failure. *Nature Communications*, **2019**, 10, 2760 | 17.4 | 11

89 | Kidney cytosine methylation changes improve renal function decline estimation in patients with diabetic kidney disease. *Nature Communications*, **2019**, 10, 2461 | 17.4 | 30

88 | Integrative analysis of genetical genomics data incorporating network structures. *Biometrics*, **2019**, 75, 1063-1075 | 1.8 | 4

87 | Generalized linear models with linear constraints for microbiome compositional data. *Biometrics*, **2019**, 75, 235-244 | 1.8 | 21

86 | Molecular analysis of the endobronchial stent microbial biofilm reveals bacterial communities that associate with stent material and frequent fungal constituents. *PLoS ONE*, **2019**, 14, e0217306 | 3.7 | 9

85 | Large Covariance Estimation for Compositional Data Via Composition-Adjusted Thresholding. *Journal of the American Statistical Association*, **2019**, 114, 759-772 | 2.8 | 16

84	Optimal Estimation of Genetic Relatedness in High-Dimensional Linear Models. *Journal of the American Statistical Association*, **2019**, 114, 358-369	2.8	8
83	Statistical and Computational Methods in Microbiome and Metagenomics **2019**, 977-550		5
82	Joint testing and false discovery rate control in high-dimensional multivariate regression. *Biometrika*, **2018**, 105, 249-269	2	3
81	Two-sample tests of high-dimensional means for compositional data. *Biometrika*, **2018**, 105, 115-132	2	17
80	Medication class enrichment analysis: a novel algorithm to analyze multiple pharmacologic exposures simultaneously using electronic health record data. *Journal of the American Medical Informatics Association: JAMIA*, **2018**, 25, 780-789	8.6	2
79	Comparative analyses of fecal microbiota in Chinese isolated Yao population, minority Zhuang and rural Han by 16sRNA sequencing. *Scientific Reports*, **2018**, 8, 1142	4.9	22
78	A GLM-based latent variable ordination method for microbiome samples. *Biometrics*, **2018**, 74, 448-457	1.8	17
77	Reframing the Biological Basis of Neuroprotection Using Functional Genomics: Differentially Weighted, Time-Dependent Multifactor Pathogenesis of Human Ischemic Brain Damage. *Frontiers in Neurology*, **2018**, 9, 497	4.1	4
76	Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nature Methods*, **2018**, 15, 1041-1044	21.6	20
75	FXR-Dependent Modulation of the Human Small Intestinal Microbiome by the Bile Acid Derivative Obeticholic Acid. *Gastroenterology*, **2018**, 155, 1741-1752.e5	13.3	54
74	Conditional Regression Based on a Multivariate Zero-Inflated Logistic-Normal Model for Microbiome Relative Abundance Data. *Statistics in Biosciences*, **2018**, 10, 587-608	1.5	18
73	Sociability Deficits and Altered Amygdala Circuits in Mice Lacking Pcdh10, an Autism Associated Gene. *Biological Psychiatry*, **2017**, 81, 193-202	7.9	38
72	A Model-Based Approach For Species Abundance Quantification Based On Shotgun Metagenomic Data. *Statistics in Biosciences*, **2017**, 9, 13-27	1.5	7
71	Sparse simultaneous signal detection for identifying genetically controlled disease genes. *Journal of the American Statistical Association*, **2017**, 112, 1032-1046	2.8	5
70	Genetic-Variation-Driven Gene-Expression Changes Highlight Genes with Important Functions for Kidney Disease. *American Journal of Human Genetics*, **2017**, 100, 940-953	11	52
69	A model for paired-multinomial data and its application to analysis of data on a taxonomic tree. *Biometrics*, **2017**, 73, 1266-1278	1.8	11
68	Optimal detection of weak positive latent dependence between two sequences of multiple tests. *Journal of Multivariate Analysis*, **2017**, 160, 169-184	1.4	2
67	A role for bacterial urease in gut dysbiosis and Crohn's disease. *Science Translational Medicine*, **2017**, 9,	17.5	92

66   A general framework for association analysis of microbial communities on a taxonomic tree. *Bioinformatics*, **2017**, 33, 1278-1285   7.2   17

65   Regression analysis for microbiome compositional data. *Annals of Applied Statistics*, **2016**, 10,   2.1   55

64   Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. *Gut*, **2016**, 65, 63-72   19.2   307

63   Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants. *JAMA Pediatrics*, **2016**, 170, 212-9   8.3   170

62   Correction to the paper Optimal False Discovery Rate Control for Dependent Data *Statistics and Its Interface*, **2016**, 9, 33-35   0.4   2

61   Joint Estimation of Multiple High-dimensional Precision Matrices. *Statistica Sinica*, **2016**, 26, 445-464   0.7   22

60   A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, **2016**, 32, 2611-7   7.2   117

59   Activation of basolateral amygdala in juvenile C57BL/6J mice during social approach behavior. *Neuroscience*, **2016**, 335, 184-94   3.9   16

58   Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *Annual Review of Statistics and Its Application*, **2015**, 2, 73-94   7.6   158

57   The Ovarian Cancer Chemokine Landscape Is Conducive to Homing of Vaccine-Primed and CD3/CD28-Costimulated T Cells Prepared for Adoptive Therapy. *Clinical Cancer Research*, **2015**, 21, 2840-50   12.9   42

56   Regularization Methods for High-Dimensional Instrumental Variables Regression With an Application to Genetical Genomics. *Journal of the American Statistical Association*, **2015**, 110, 270-288   2.8   38

55   Testing in Microbiome-Profiling Studies with MiRKAT, the Microbiome Regression-Based Kernel Association Test. *American Journal of Human Genetics*, **2015**, 96, 797-807   11   161

54   Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics*, **2015**, 31, 2461-8   7.2   193

53   Genetic sharing and heritability of paediatric age of onset autoimmune diseases. *Nature Communications*, **2015**, 6, 8442   17.4   46

52   Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell Host and Microbe*, **2015**, 18, 489-500   23.4   446

51   glmgraph: an R package for variable selection and predictive modeling of structured genomic data. *Bioinformatics*, **2015**, 31, 3991-3   7.2   6

50   Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nature Medicine*, **2015**, 21, 1018-27   50.5   143

49   Comparative Effectiveness of Nutritional and Biological Therapy in North American Children with Active Crohn's Disease. *Inflammatory Bowel Diseases*, **2015**, 21, 1786-93   4.5   100

48	Nonparametric Tests for Differential Histone Enrichment with ChIP-Seq Data. *Cancer Informatics*, **2015**, 14, 11-22	2.4	2
47	Genome wide nucleosome mapping for HSV-1 shows nucleosomes are deposited at preferred positions during lytic infection. *PLoS ONE*, **2015**, 10, e0117471	3.7	15
46	SPARSE SEGMENT IDENTIFICATIONS WITH APPLICATIONS TO DNA COPY NUMBER VARIATION ANALYSIS **2015**, 863-887		
45	A change-point model for identifying 3'UTR switching by next-generation RNA sequencing. *Bioinformatics*, **2014**, 30, 2162-70	7.2	30
44	Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology*, **2014**, 147, 1055-63.e8	13.3	464
43	A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer Cell*, **2014**, 26, 344-357	24.3	303
42	Isoform Expression Analysis Based on RNA-seq Data **2014**, 247-259		
41	Direct estimation of differential networks. *Biometrika*, **2014**, 101, 253-268	2	65
40	Variable selection in regression with compositional covariates. *Biometrika*, **2014**, 101, 785-797	2	112
39	Parametric modeling of whole-genome sequencing data for CNV identification. *Biostatistics*, **2014**, 15, 427-41	3.7	5
38	More powerful genetic association testing via a new statistical framework for integrative genomics. *Biometrics*, **2014**, 70, 881-90	1.8	23
37	A Hierarchical Bayesian Model for Estimating and Inferring Differential Isoform Expression for Multi-Sample RNA-Seq Data. *Statistics in Biosciences*, **2013**, 5, 119-137	1.5	7
36	A logistic normal multinomial regression model for microbiome compositional data analysis. *Biometrics*, **2013**, 69, 1053-63	1.8	68
35	Systems biology approaches to epidemiological studies of complex diseases. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, **2013**, 5, 677-86	6.6	7
34	Covariate-Adjusted Precision Matrix Estimation with an Application in Genetical Genomics. *Biometrika*, **2013**, 100, 139-156	2	53
33	Adjusting for High-dimensional Covariates in Sparse Precision Matrix Estimation by ℓ_1-Penalization. *Journal of Multivariate Analysis*, **2013**, 116, 365-381	1.4	16
32	LEARNING LOCAL DIRECTED ACYCLIC GRAPHS BASED ON MULTIVARIATE TIME SERIES DATA. *Annals of Applied Statistics*, **2013**, 7, 1249-1835	2.1	2
31	VARIABLE SELECTION FOR SPARSE DIRICHLET-MULTINOMIAL REGRESSION WITH AN APPLICATION TO MICROBIOME DATA ANALYSIS. *Annals of Applied Statistics*, **2013**, 7,	2.1	129

30	MATCHCLIP: locate precise breakpoints for copy number variation using CIGAR string by matching soft clipped reads. *Frontiers in Genetics*, **2013**, 4, 157	4.5	12
29	Robust Detection and Identification of Sparse Segments in Ultra-High Dimensional Data Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2012**, 74, 773-797	3.9	16
28	U-statistics in genetic association studies. *Human Genetics*, **2012**, 131, 1395-401	6.3	11
27	High-Dimensional Heteroscedastic Regression with an Application to eQTL Data Analysis. *Biometrics*, **2012**, 68, 316-326	1.8	25
26	Robust Gaussian graphical modeling via l1 penalization. *Biometrics*, **2012**, 68, 1197-206	1.8	18
25	Introduction to Special Issue Methods for Analysis of Graphs and Networks and Their Applications in Biosciences *Statistics in Biosciences*, **2012**, 4, 1-2	1.5	5
24	Hidden Markov Random Field Models for Network-Based Analysis of Genomic Data **2011**, 353-368		
23	Linking long-term dietary patterns with gut microbial enterotypes. *Science*, **2011**, 334, 105-8	33.3	3898
22	A SPARSE CONDITIONAL GAUSSIAN GRAPHICAL MODEL FOR ANALYSIS OF GENETICAL GENOMICS DATA. *Annals of Applied Statistics*, **2011**, 5, 2630-2650	2.1	86
21	The Sparse Laplacian Shrinkage Estimator for High-Dimensional Regression. *Annals of Statistics*, **2011**, 39, 2021-2046	3.2	56
20	A penalized likelihood approach for bivariate conditional normal models for dynamic co-expression analysis. *Biometrics*, **2011**, 67, 299-308	1.8	10
19	Network-based empirical Bayes methods for linear models with applications to genomic data. *Journal of Biopharmaceutical Statistics*, **2010**, 20, 209-22	1.3	8
18	Optimal Sparse Segment Identification with Application in Copy Number Variation Analysis. *Journal of the American Statistical Association*, **2010**, 105, 1156-1166	2.8	46
17	A hidden Markov random field model for genome-wide association studies. *Biostatistics*, **2010**, 11, 139-50	3.7	33
16	VARIABLE SELECTION AND REGRESSION ANALYSIS FOR GRAPH-STRUCTURED COVARIATES WITH AN APPLICATION TO GENOMICS. *Annals of Applied Statistics*, **2010**, 4, 1498-1516	2.1	78
15	Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS ONE*, **2010**, 5, e15216	3.7	262
14	A Bayesian Approach for Graph-constrained Estimation for High-dimensional Regression **2010**, 1, 255-272		
13	Copy number variation at 1q21.1 associated with neuroblastoma. *Nature*, **2009**, 459, 987-91	50.4	285

12	U-statistics-based tests for multiple genes in genetic association studies. *Annals of Human Genetics*, **2008**, 72, 821-33	2.2 32
11	Vertex Clustering in Random Graphs via Reversible Jump Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics*, **2008**, 17, 388-409	1.4 2
10	Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, **2008**, 24, 1175-82	7.2 395
9	Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *New England Journal of Medicine*, **2008**, 358, 2585-93	59.2 224
8	A hidden spatial-temporal Markov random field model for network-based analysis of time course gene expression data. *Annals of Applied Statistics*, **2008**, 2,	2.1 45
7	A Markov random field model for network-based analysis of genomic data. *Bioinformatics*, **2007**, 23, 1537-44	7.2 189
6	Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics*, **2006**, 7, 302-17	3.7 113
5	Statistical Methods for Inference of Genetic Networks and Regulatory Modules143-167	1
4	Censored Data Regression in High-Dimensional and Low-Sample-Size Settings for Genomic Applications385-403	1
3	Inference for High-Dimensional Linear Mixed-Effects Models: A Quasi-Likelihood Approach. *Journal of the American Statistical Association*,1-12	2.8 4
2	Interaction Networks in Microbiome Studies1-13	
1	Transfer Learning in Large-Scale Gaussian Graphical Models with False Discovery Rate Control. *Journal of the American Statistical Association*,1-13	2.8 1