Yufeng Shen

Associate Professor of Systems Biology and Biomedical Informatics
Columbia University

1130 St Nicholas Ave, 803B, New York, NY 10032
Email: ys2411@cumc.columbia.edu Web: http://www.columbia.edu/~ys2411/

Appointments

Associate Professor, 7/2019–Present
Assistant Professor, 7/2011–6/2019
Departments of Systems Biology and Biomedical Informatics
Columbia University

Education

Ph.D., *Baylor College of Medicine*, Houston, Texas, 2007
Computational biology, Advisor: Dr. George Weinstock
B.Sc., *Peking University*, Beijing, 2000
Molecular Biology and Biochemistry, and Honor Science Program

Training

Postdoctoral Research Scientist, advised by Dr. Itsik Pe'er, 2008–2011
Department of Computer Science, Columbia University
Postdoctoral Fellow, advised by Drs. Richard Gibbs and George Weinstock, 2007–2008
Human Genome Sequencing Center, Baylor College of Medicine

Administrative Leadership and Academic Services

2011-present	Associate Director, JP Sulzberger Columbia Genome Center
2015-present	Adjunct member, Columbia Center of Translational Immunology
2016-present	Member of PhD Admissions Committee, Department of Systems Biology
2015-present	Co-organizer of Distinguished Speaker Seminar Series in Department of Systems Biology
2018-present	Affiliated Member, New York Genome Center

Research Support

Grants as PI or core director

2017-2022	National Institute of Health (NIH) R01 GM120609 *Integrate cancer genomics data in genetic studies and diagnosis of developmental disorders* (PI)
2019-2021	NIH R03 HL147197 *Integrate gene expression data to characterize the contribution of rare genetic risk factors to structural birth defects* (PI)
2019-2020	Simons Foundation Autism Research Initiative, *Maximizing autism gene discovery by machine learning and single cell data* (PI)
2017-2022	NIH P01 HD068250-06A1, *Gene Mutation and Rescue in Human Diaphragmatic Hernia*, (Role: MPI; PD: Donahoe).

2017-2022 NIH P01 HD093363, *Developmental Mechanisms of Trachea-Esophageal Birth Defects* (Role: Director of Genomics Core; PD: Zorn).

2016-2021 NIH U19 AI128949, *Human anti-viral immune responses in tissues and circulation.* (Role: Co-Director of Data Analysis Core; PD: Farber).

2018-2023 NIH P01 AI106697, *Tissue compartmentalization of human lymphocytes* (Role: PI of Bioinformatics Core; PD: Farber).

Previous grants

2017-2019 NIH R03 HL138352 *Genetic analysis of structural birth defects by integration of multiple diseases with epigenomic data and cancer mutations* (PI).

2013-2018 NIH P01 AI106697, *Tissue compartmentalization of human lymphocytes* (PI of Bioinformatics Core).

2011-2013 Amazon Web Services in Education Research Grant, *Cloud-based computational framework for analyzing large-scale high-throughput sequencing data.* (PI)

2018-2019 Amazon AWS Cloud Credits for Research program. Efficient cloud computing for bioinformatics analysis of large-scale genome sequencing data in genetic studies. (PI)

2015,16,17 NIH X01HL132366, X01HL140543, X01HL136998. *Genomic Analysis of Congenital Diaphragmatic Hernia*, (Role: MPI; PD: Chung WK).

Educational Contributions

Teaching

2015–present **G4017** *Deep Sequencing*, a graduate course about principles of high-throughput sequencing technologies, foundational statistical and computational methods in computational biology, and applications of sequencing in genetics and systems biology. Teach every fall semester at Columbia University Medical Center

2011-present Guest lectures in *Introduction to Precision Medicine*, *Introduction to Computer Applications in Health Care & Biomedicine*, *Functional Genetics Boot Camp: Computational Integration of Genome and Transcriptome Data.*

Graduate Student/Postdoctoral Fellow Mentoring

Previous trainees

PhD students

2015–2019 Alexander L. Hsieh (PhD student, DBMI): currently scientist at Broad Institute. Thesis: Mosaic mutations and congenital heart disease.

2014–2018 Hongjian Qi (PhD student, Applied Physics): currently data scientist at Robinhood Markets Inc. Thesis: The contribution of de novo mutations to genetic architecture of developmental disorders.

2017-2018 Chen Chen (PhD student, Biological Sciences): currently postdoc fellow at Illumina Inc. Thesis: The Generation and Functional Consequence of Human Genetic Mutations.

2012–2017 Boris Grinshpun (PhD student, Integrated Program): currently Postdoc fellow at EMD Serono. Thesis: diversity of T cell receptors in healthy human and patients.

Postdoctoral scientists

2014–2019 Wenji Ma: currently Assistant Professor at Chinese Academy of Sciences
2015–2017 Xinwei Han: currently Scientist at Constellation Pharmaceuticals
2012–2014 Badri Vardarajan: currently Assistant Professor at Columbia University
2011–2013 Casey Overby: currently Assistant Professor at Johns Hopkins University
2011–2012 Yaping Feng: currently Research Scientist at Rutgers University

Master or undergraduate students:
2018-2018 Bonnie Wang (Undergraduate): currently at Columbia College
2017-2017 Emily Gao (Undergraduate from Cornell): currently at FaceBook
2017-2017 Dana Wintner (Undergraduate from Cornell).
2014–2016 Aleksandar Obradovic (Undergraduate): currently MD/PhD student at CUMC
2014–2015 Jonathan Packer (Undergraduate): currently PhD student at Univ. of Washington
2012–2015 Daniel Backenroth (Master student): currently Director at Johnson & Johnson
2012–2014 Edwin Lin (Master student): currently MD/PhD student at University of Utah

Current Trainees
2019-present Yige Zhao (PhD student, DSB)
2019-present Dr. Zicheng Wang (Postdoc)
2019-present Dr. Renjie Tan (Postdoc)
2018–present Bulat Ziganshin (PhD student, Department of Genetics and Development)
2018–present Dr. Xiao Fan (Postdoc)
2018–present Dr. Lu Qiao (Postdoc)
2017–present Dr. Haicang Zhang (Postdoc)
2017–present Dr. Xueya Zhou (Postdoc, jointly mentored by Dr. Wendy Chung)
2017–present Siying Chen (PhD student, Integrated Program)
2016–present Alexander Kitaygorodsky (PhD student, DBMI)
2016–present Na Zhu (Postdoc, jointed mentored by Dr. Wendy Chung)

PhD Thesis committee
2017–present Xiangtian Tan, Biological Sciences (Califano group)
2016–2019 Filip Cvetkovski, Immunology (Farber group)
2016–2019 Michelle Miron, Integrated Program (Farber group)
2014–2017 Jing He, DBMI (Califano group)
2013–2016 Yajing Angela Xie, Integrated Program (Allikmets group)
2014–2015 Eugenia Lyashenko, DBMI (Vitkup group)
2014–2015 Rachel Melamed, DBMI (Rabadan group)
2013–2015 Naomi Yudanin, Immunology (Farber group)
2013–2014 Zachary W. Carpenter, Integrated Program (Rabadan group)
2012–2013 Jiyang Yu, DBMI (Califano group)

Publications

# indicates corresponding authorship
* indicated co-first authorship
Underlined indicates lab members
& indicates consortium member

Selected Publications in Peer-reviewed Journals

1. Hsieh A, Morton SU, Willcox JAL, Gorham JM, Tai AC, Qi H, DePalma S, et al, Chung WK, Seidman CE, Seidman JG, **Shen Y**, (2020) EM-mosaic detects mosaic point mutations that contribute to congenital heart disease, *Genome Medicine*, PMID: 32349777
2. Richter F*, Morton SU*, Kim SW*, Kitaygorodsky A*, Wasson L, Chen KM, Zhou J, Qi H, Patel N, DePalma S, …, Srivastava D#, Tristani-Firouzi M#, Troyanskaya OG#, Dickel DE#, **Shen Y#,** Seidman JG#, Seidman CE#, Gelb BD#, (2020) Role of noncoding de novo variants in congenital heart disease, *Nature Genetics*, in press
3. Feliciano P*, Zhou X*, Astrovskaya I*, Turner TN*, Wang T, Brueggeman L, Barnard R, Hsieh A, Snyder LG, Muzny DM, Sabo A; SPARK Consortium, Gibbs RA, Eichler EE, O'Roak BJ, Michaelson JJ, Volfovsky N, **Shen Y**, Chung WK, (2019) Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *npj Genomic Medicine*, PMID: 31452935
4. Ma W, Lee J, Backenroth D, Zhou YJ, Bush E, Sims P, Liu K, **Shen Y**, (2019) Single cell RNA-Seq reveals pre-cDCs fate determined by transcription factor combinatorial dose, *BMC Mol Cell Biol*, PMID: 31253076
5. Han X*, Chen S*, Flynn E, Wu S, Wintner D, **Shen Y#**, (2018) Distinct epigenomic patterns are associated with haploinsufficiency and predict risk genes of developmental disorders, *Nature Communications*, PMID: 29849042
6. Qi H*, Yu L*, Zhou X*, Kitaygorodsky A, et al, Chung WK#, **Shen Y#**, (2018) Genetic analysis of *de novo* variants reveals sex differences in complex and isolated congenital diaphragmatic hernia and indicates *MYRF* as a candidate gene. *PLOS Genetics*. PMID: 30532227
7. SPARK Consortium[&], (2018) SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research, *Neuron,* PMID: 29420931
8. DeWolf S*, Grinshpun B*, Savage T, Lau SP, Obradovic A, Shonts B, Yang S, Morris H, Zuber J, Winchester R, Sykes M, **Shen Y#** (2018) Quantifying size and diversity of the human T cell alloresponse, *JCI Insight*, PMID: 30089728
9. Zhu N, Welch CL, Wang J, Allen PM, Gonzaga-Jauregui C, Ma L, King AK, Krishnan U, Rosenzweig EB, Ivy DD, Austin ED, Hamid R, Pauciulo MW, Lutz KA, Nichols WC, Reid JG, Overton JD, Baras A, Dewey FE, **Shen Y**, Chung WK, (2018) Rare variants in SOX17 are associated with pulmonary arterial hypertension with congenital heart disease. *Genome Medicine*, PMID: 30029678
10. Zhu N, Gonzaga-Jauregui C, Welch CL, Ma L, Qi H, King AK, Krishnan U, Rosenzweig EB, Ivy DD, Austin ED, Hamid R, Nichols WC, Pauciulo MW, Lutz KA, Sawle A, Reid JG, Overton JD, Baras A, Dewey F, **Shen Y**, Chung WK. (2018) Exome Sequencing in Children With Pulmonary Arterial Hypertension Demonstrates Differences Compared With Adults. *Circulation: Genomic and Precision Medicine*, PMID: 29631995
11. Lee J*, Zhou YJ*, Ma W*, Zhang W, Aljoufi A, Luh T, Lucero K, Liang D, Thomsen M, Bhagat G, **Shen Y#**, Liu K#, (2017) Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors. *Nature Immunology*, PMID: 28650480
12. Zhang C, **Shen Y#**, (2017) A cell type-specific expression signature predicts haploinsufficient autism-susceptibility genes. *Human Mutation*, PMID: 27860035

13. Kumar BV*, Ma W*, Miron M, Granot T, Guyer RS, Carpenter DJ, Senda T, Sun X, Ho SH, Lerner H, Friedman AL, **Shen Y#**, Farber DL#, (2017) Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites, *Cell Reports*, PMID: 28930685
14. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, Hung WC, Haider S, Zhang J, Knight J, Bjornson RD, Castaldi C, Tikhonoa IR, Bilguvar K, Mane SM, Sanders SJ, Mital S, Russell MW, Gaynor JW, Deanfield J, Giardini A, Porter GA Jr, Srivastava D, Lo CW, **Shen Y**, Watkins WS, Yandell M, Yost HJ, Tristani-Firouzi M, Newburger JW, Roberts AE, Kim R, Zhao H, Kaltman JR, Goldmuntz E, Chung WK, Seidman JG, Gelb BD, Seidman CE, Lifton RP, Brueckner M, (2017) Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands, *Nature Genetics*, PMID: 28991257
15. Longoni M, High FA, Qi H, Joy MP, Hila R, Coletti CM, Wynn J, Loscertales M, Shan L, Bult CJ, Wilson JM, **Shen Y**, Chung WK, Donahoe PK, (2017) Genome-wide enrichment of damaging de novo variants in patients with isolated and complex congenital diaphragmatic hernia. *Human Genetics*, PMID: 28303347
16. Chen C, Qi H, **Shen Y**, Pickrell J, Przeworski M. (2017) Contrasting Determinants of Mutation Rates in Germline and Soma. *Genetics*, PMID: 28733365
17. Qi H, Dong C, Chung WK, Wang K, **Shen Y#**, (2016) Deep Genetic Connection between Cancer and Developmental Disorders. *Human Mutation*, PMID: 27363847

    A commentary on this paper was published in the same issue: by Zek et al (2016) Cancer Genetics May Aid Diagnostics of Developmental Disorders. *Human Mutation*, PMID: 27623404
18. Sims JS, Grinshpun B, Feng Y, Ung TH, Neira JA, Samanamud JL, Canoll P, **Shen Y#**, Sims PA #, Bruce JN#, (2016) Diversity and divergence of the glioma-infiltrating T-cell receptor repertoire. *Proc Natl Acad Sci U S A*, PMID: 27261081
19. Thome JJ*, Grinshpun B*, et al, **Shen Y**, Farber DL, (2016) Long-term maintenance of human naive T cells through in situ homeostasis in lymphoid tissue sites, *Science Immunology*, PMID: 28361127
20. Castel SE, Mohammadi P, Chung WK, **Shen Y**, Lappalainen T, (2016) Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nature Communications*, PMID: 27605262
21. DeWolf S, **Shen Y**, Sykes M. (2016) A New Window into the Human Alloresponse. *Transplantation*, PMID: 26760572
22. Bain JM, Cho MT, Telegrafi A, Wilson A, Brooks S, Botti C, Gowans G, Autullo LA, Krishnamurthy V, Willing MC, Toler TL, Ben-Zev B, Elpeleg O, **Shen Y**, Retterer K, Monaghan KG, Chung WK, (2016) Variants in HNRNPH2 on the X chromosome are associated with a neurodevelopmental disorder in females, *American Journal of Human Genetics*, PMID: 27545675
23. Homsy J*, Zaidi S*, **Shen Y***, Ware JS, et al. (2015) De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*, PMID: 26785492
24. Morris H, De Wolf S, Robins H, Sprangers B, Locascio SA, Shonts BA, Kawai T, Wong W, Yang S, Zuber J, **Shen Y#**, Sykes M# (2015) Tracking donor-reactive T cells: evidence for clonal deletion in tolerant kidney transplant patients, *Science Translational Medicine*, PMID: 25632034.

25. Yu L, Sawle AD, Wynn J, Aspelund G, Stolar CJ, Arkovitz MS, Potoka D, Azarow KS, Mychaliska GB, **Shen Y#**, Chung WK#, (2015) Increased burden of de novo predicted deleterious variants in complex congenital diaphragmatic hernia, *Human Molecular Genetics* PMID: 26034137
26. Ostrowski EA#, **Shen Y**, Tian X, Sucgang R, et al, Strassmann JE, Queller DC. (2015) Genomic signatures of cooperation and conflict in the social amoeba. *Current Biology*, PMID: 26051890
27. Backenroth D, Homsy J, Murillo LR, Glessner J, Lin E, Brueckner M, Lifton R, Goldmuntz E, Chung WK, **Shen Y#**, (2014) CANOES: Detecting rare copy number variants from whole exome sequencing data, *Nucleic Acids Research,* PMID: 24771342
28. Thome JJC, Yudanin N, Ohmura Y, Kubota M, Grinshpun B, Sathaliyawala T, Kat T, Lerner H, **Shen Y**, Farber DL, (2014) Spatial Map of Human T Cell Compartmentalization and Maintenance over Decades of Life, *Cell*, PMID: 25417158
29. Gao F, Lin E, Feng Y, Mack WJ, **Shen Y#**, Wang K# (2013) Characterizing immunoglobulin repertoire from whole blood by a personal genome sequencer. *PLoS One,* PMID: 24058670
30. **Shen Y#**, Song R, and Pe'er I#, (2011) Coverage Tradeoffs and Power Estimation in the Design of Whole-Genome Sequencing Experiments for Detecting Association, *Bioinformatics*, PMID: 21636589
31. **Shen Y#**, Gu Y, Pe'er I# (2011) A hidden Markov model for copy number variant prediction from whole genome resequencing data. *BMC Bioinformatics,* PMID: 21989326
32. **Shen Y#**, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, Liu Y, Weinstock GM, Wheeler DA, Gibbs RA, Yu F#. (2010) A SNP discovery method to assess variant allele probability from next generation resequencing data. *Genome Research.* PMID: 20019143
33. Wheeler DA*, Srinivasan M*, Egholm M*, **Shen Y***, et al, (2008) The complete genome of an individual by massively parallel DNA sequencing, *Nature*, PMID: 18421352

Other Publications in Peer-reviewed Journals

34. Watkins WS, Hernandez EJ, Wesolowski S, Bisgrove BW, Sunderland RT, Lin E, Lemmon G, Demarest BL, Miller TA, Bernstein D, Brueckner M, Chung WK, Gelb BD, Goldmuntz E, Newburger JW, Seidman CE, **Shen Y**, Yost HJ, Yandell M, Tristani-Firouzi M. (2019) De novo and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes. *Nature Communications*, PMID: 31624253
35. Manheimer KB, Richter F, Edelmann LJ, D'Souza SL, Shi L, **Shen Y**, Homsy J, Boskovski MT, Tai AC, Gorham J, Yasso C, Goldmuntz E, Brueckner M, Lifton RP, Chung WK, Seidman CE, Seidman JG, Gelb BD (2018) Robust identification of mosaic variants in congenital heart disease. *Hum Genet* 137: 183-193. doi: 10.1007/s00439-018-1871-6
36. Savage TM, Shonts BA, Obradovic A, Dewolf S, Lau S, Zuber J, Simpson MT, Berglund E, Fu J, Yang S, Ho SH, Tang Q, Turka LA, **Shen** Y, Sykes M. (2018) Early expansion of donor-specific Tregs in tolerant kidney transplant recipients, *JCI Insight*, PMID: 30429370
37. Bohnen MS, Ma L, Zhu N, Qi H, et al, **Shen Y**, Nichols CG, Kass RS, Chung WK. (2018) Loss-of-Function ABCC8 Mutations in Pulmonary Arterial Hypertension, *Circulation: Genomic and Precision Medicine*, doi: 10.1161/CIRCGEN.118.002087

38. Berko ER, Cho MT, Eng C, Shao Y, Sweetser DA, Waxler J, Robin NH, Brewer F, Donkervoort S, Mohassel P, Bonnemann CG, Bialer M, Moore C, Wolfe LA, Tifft CJ, **Shen Y**, Retterer K, Millan F, Chung WK (2017) De novo missense variants in HECW2 are associated with neurodevelopmental delay and hypotonia. *J Med Genet* 54: 84-86. doi: 10.1136/jmedgenet-2016-103943
39. McKean DM, Homsy J, Wakimoto H, Patel N, Gorham J, DePalma SR, Ware JS, Zaidi S, Ma W, Patel N, Lifton RP, Chung WK, Kim R, **Shen Y**, Brueckner M, Goldmuntz E, Sharp AJ, Seidman CE, Gelb BD, Seidman JG (2016) Loss of RNA expression and allele-specific expression associated with congenital heart disease. *Nat Commun* 7: 12824. doi: 10.1038/ncomms12824
40. Shi L, Guo Y, Dong C, Huddleston J, Yang H, Han X, Fu A, Li Q, Li N, Gong S, Lintner KE, Ding Q, Wang Z, Hu J, Wang D, Wang F, Wang L, Lyon GJ, Guan Y, **Shen Y**, Evgrafov OV, Knowles JA, Thibaud-Nissen F, Schneider V, Yu CY, Zhou L, Eichler EE, So KF, Wang K (2016) Long-read sequencing and de novo assembly of a Chinese genome. *Nat Commun* 7: 12065. doi: 10.1038/ncomms12065
41. Liu X, Jia Y, Stoopler MB, **Shen Y**, Cheng H, Chen J, Mansukhani M, Koul S, Halmos B, Borczuk AC (2016) Next-Generation Sequencing of Pulmonary Sarcomatoid Carcinoma Reveals High Frequency of Actionable MET Gene Mutations. *J Clin Oncol* 34: 794-802. doi: 10.1200/JCO.2015.62.0674
42. Nicoletti P, Werk AN, Sawle A, **Shen Y**, et al, (2016) HLA-DRB1*16: 01-DQB1*05: 02 is a novel genetic risk factor for flupirtine-induced liver injury. *Pharmacogenet Genomics* 26: 218-24. doi: 10.1097/FPC.0000000000000209
43. Beck DB, Cho MT, Millan F, Yates C, Hannibal M, O'Connor B, Shinawi M, Connolly AM, Waggoner D, Halbach S, Angle B, Sanders V, **Shen Y**, Retterer K, Begtrup A, Bai R, Chung WK (2016) A recurrent de novo CTBP1 mutation is associated with developmental delay, hypotonia, ataxia, and tooth enamel defects. *Neurogenetics* 17: 173-8. doi: 10.1007/s10048-016-0482-4
44. Nicoletti P, Bansal M, Lefebvre C, Guarnieri P, **Shen Y**, Pe'er I, Califano A, Floratos A (2015) ABC transporters and the proteasome complex are implicated in susceptibility to Stevens-Johnson syndrome and toxic epidermal necrolysis across multiple drugs. *PLoS One* 10: e0131038. doi: 10.1371/journal.pone.0131038
45. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, **Shen Y**, Floratos A, Sham PC, Li MJ, Wang J, Cardon LR, Whittaker JC, Sanseau P (2015) The support of human genetic evidence for approved drug indications. *Nat Genet* 47: 856-60. doi: 10.1038/ng.3314
46. Guo Y, Ding X, **Shen Y**, Lyon GJ, Wang K (2015) SeqMule: automated pipeline for analysis of human exome/genome sequencing data. *Sci Rep* 5: 14283. doi: 10.1038/srep14283
47. Jiang K, Sun X, Chen Y, **Shen Y**, Jarvis JN (2015) RNA sequencing from human neutrophils reveals distinct transcriptional differences associated with chronic inflammatory states. *BMC Med Genomics* 8: 55. doi: 10.1186/s12920-015-0128-7
48. Vardarajan BN, Zhang Y, Lee JH, Cheng R, Bohm C, Ghani M, Reitz C, Reyes-Dumeyer D, **Shen Y**, Rogaeva E, St George-Hyslop P, Mayeux R (2015) Coding mutations in SORL1 and Alzheimer disease. *Ann Neurol* 77: 215-27. doi: 10.1002/ana.24305
49. Shang L, Cho MT, Retterer K, Folk L, Humberson J, Rohena L, Sidhu A, Saliganan S, Iglesias A, Vitazka P, Juusola J, O'Donnell-Luria AH, **Shen Y**, Chung WK (2015) Mutations in ARID2 are associated with intellectual disabilities. *Neurogenetics*. doi: 10.1007/s10048-015-0454-0

50. Glessner J, Bick AG, Ito K, Homsy J, Rodriguez-Murillo L, Fromer M, Mazaika EJ, Vardarajan B, Italia MJ, Leipzig J, DePalma S, Golhar R, Sanders SJ, Yamrom B, Ronemus M, Iossifov I, Willsey AJ, State MW, Kaltman JR, White PS, **Shen Y**, Warburton D, Brueckner M, Seidman C, Goldmuntz E, Gelb BD, Lifton R, Seidman JG, Hakonarson H, Chung WK (2014) Increased Frequency of De Novo Copy Number Variations in Congenital Heart Disease by Integrative Analysis of SNP Array and Exome Sequence Data. *Circ Res.* doi: 10.1161/CIRCRESAHA.115.304458
51. Johannesson B, Sagi I, Gore A, Paull D, Yamada M, Golan-Lev T, Li Z, LeDuc C, **Shen Y**, Stern S, Xu N, Ma H, Kang E, Mitalipov S, Sauer MV, Zhang K, Benvenisty N, Egli D (2014) Comparable frequencies of coding mutations and loss of imprinting in human pluripotent cells derived by nuclear transfer and defined factors. *Cell Stem Cell* 15: 634-42. doi: 10.1016/j.stem.2014.10.002
52. Leduc CA, Crouch EE, Wilson A, Lefkowitch J, Wamelink MM, Jakobs C, Salomons GS, Sun X, **Shen Y**, Chung WK (2014) Novel association of early onset hepatocellular carcinoma with transaldolase deficiency. *JIMD Rep* 12: 121-7. doi: 10.1007/8904_2013_254
53. Gill R, Cheung YH, **Shen Y**, Lanzano P, Mirza NM, Ten S, Maclaren NK, Motaghedi R, Han JC, Yanovski JA, Leibel RL, Chung WK (2014) Whole-exome sequencing identifies novel LEPR mutations in individuals with severe early onset obesity. *Obesity* (Silver Spring) 22: 576-84. doi: 10.1002/oby.20492
54. Yu L, Bennett JT, Wynn J, Carvill GL, Cheung YH, **Shen Y**, Mychaliska GB, et al, Mefford H, Chung WK (2014) Whole exome sequencing identifies de novo mutations in GATA6 associated with congenital diaphragmatic hernia. *J Med Genet* 51: 197-202. doi: 10.1136/jmedgenet-2013-101989
55. Yu L, Wynn J, Cheung YH, **Shen Y**, Mychaliska GB, et al, Chung WK (2013) Variants in GATA4 are a rare cause of familial and sporadic congenital diaphragmatic hernia. *Hum Genet* 132: 285-92. doi: 10.1007/s00439-012-1249-0
56. Behr ER, Ritchie MD, Tanaka T, Kaab S, Crawford DC, Nicoletti P, Floratos A, Sinner MF, Kannankeril PJ, Wilde AA, Bezzina CR, Schulze-Bahr E, Zumhagen S, Guicheney P, Bishopric NH, Marshall V, Shakir S, Dalageorgou C, Bevan S, Jamshidi Y, Bastiaenen R, Myerburg RJ, Schott JJ, Camm AJ, Steinbeck G, Norris K, Altman RB, Tatonetti NP, Jeffery S, Kubo M, Nakamura Y, **Shen Y**, George AL, Jr., Roden DM (2013) Genome wide analysis of drug-induced torsades de pointes: lack of common variants with large effect sizes. *PLoS One* 8: e78511. doi: 10.1371/journal.pone.0078511
57. Boland MR, Hripcsak G, **Shen Y**, Chung WK, Weng C (2013) Defining a comprehensive verotype using electronic health records for personalized medicine. *J Am Med Inform Assoc* 20: e232-8. doi: 10.1136/amiajnl-2013-001932
58. Overby CL, Pathak J, Gottesman O, Haerian K, Perotte A, Murphy S, Bruce K, Johnson S, Talwalkar J, **Shen Y**, Ellis S, Kullo I, Chute C, Friedman C, Bottinger E, Hripcsak G, Weng C (2013) A collaborative approach to developing an electronic health record phenotyping algorithm for drug-induced liver injury. *J Am Med Inform Assoc* 20: e243-52. doi: 10.1136/amiajnl-2013-001930
59. DeStefano GM, Fantauzzo KA, Petukhova L, Kurban M, Tadin-Strapps M, Levy B, Warburton D, Cirulli ET, Han Y, Sun X, **Shen Y**, et al, Christiano AM (2013) Position effect on FGF13 associated with X-linked congenital generalized hypertrichosis. *Proc Natl Acad Sci U S A* 110: 7790-5. doi: 10.1073/pnas.1216412110

60. **Shen Y**, Nicoletti P, Floratos A, Pirmohamed M, Molokhia M, Geppetti P, Benemei S, Giomi B, Schena D, Vultaggio A, Stern R, Daly MJ, John S, Nelson MR, Pe'er I, International Serious Adverse Events C (2012) Genome-wide association study of serious blistering skin rash caused by drugs. *Pharmacogenomics J* 12: 96-104. doi: 10.1038/tpj.2010.84
61. Nicoletti P, Cartsos VM, Palaska PK, **Shen Y**, Floratos A, Zavras AI (2012) Genomewide pharmacogenetics of bisphosphonate-induced osteonecrosis of the jaw: the role of RBMS3. *Oncologist* 17: 279-87. doi: 10.1634/theoncologist.2011-0202
62. Urban TJ, **Shen Y**, Stolz A, Chalasani N, et al, (2012) Limited contribution of common genetic variants to risk for liver injury due to a variety of drugs. *Pharmacogenet Genomics* 22: 784-95. doi: 10.1097/FPC.0b013e3283589a76
63. Lucena MI, Molokhia M**, Shen Y**, Urban TJ, et al, (2011) Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and II alleles. *Gastroenterology* 141: 338-47. doi: 10.1053/j.gastro.2011.04.001
64. Daly AK, Donaldson PT, Bhatnagar P, **Shen Y**, et al, Day CP, Study D, International SAEC (2009) HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat Genet* 41: 816-9. doi: 10.1038/ng.379
65. Liu Y, Qin X, Song XZ, Jiang H, **Shen Y**, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, Zhang L, Sodergren E, Havlak P, Worley KC, Weinstock GM, Gibbs RA (2009) Bos taurus genome assembly. *BMC Genomics* 10: 180. doi: 10.1186/1471-2164-10-180
66. **Shen Y**, Sarin S, Liu Y, Hobert O, Pe'er I (2008) Comparing platforms for C. elegans mutant identification using high-throughput whole-genome sequencing. *PLoS One* 3: e4012. doi: 10.1371/journal.pone.0004012
67. Wang X, Gao H, **Shen** Y, Weinstock GM, Zhou J, Palzkill T (2008) A high-throughput percentage-of-binding strategy to measure binding energies in DNA-protein interactions: application to genome-scale site discovery. *Nucleic Acids Res* 36: 4863-71. doi: 10.1093/nar/gkn477
68. Rhesus Macaque Genome Sequencing Consortium[&] (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316: 222-34. doi: 10.1126/science.1139247
69. Sodergren E, **Shen Y**, Song X, Zhang L, Gibbs RA, Weinstock GM (2006) Shedding genomic light on Aristotle's lantern. *Dev Biol* 300: 2-8. doi: 10.1016/j.ydbio.2006.10.005
70. Sea Urchin Genome Sequencing Consortium[&] (2006) The genome of the sea urchin Strongylocentrotus purpuratus. *Science* 314: 941-52. doi: 10.1126/science.1133609
71. Kong Y, **Shen Y**, Warth TE, Ma J (2002) Conformational pathways in the gating of Escherichia coli mechanosensitive channel. *Proc Natl Acad Sci U S A* 99: 5999-6004. doi: 10.1073/pnas.092051099
72. **Shen Y**, Kong Y, Ma J (2002) Intrinsic flexibility and gating mechanism of the potassium channel KcsA. *Proc Natl Acad Sci U S A* 99: 1949-53. doi: 10.1073/pnas.042650399

Preprints under peer-review

73. Qi H*, Chen C*, Zhang H, Long JJ, Chung WK, Guan Y, **Shen Y#**. (2018) MVP: predicting pathogenicity of missense variants by deep learning. *bioRxiv*, 259390

H-index as of May 2020: 38
(Google Scholar:
https://scholar.google.com/citations?user=e3UEBXAAAAAJ&sortby=pubdate)

Invited Talks

2012 Joint Statistical Meetings, American Statistical Association, San Diego, CA
2013 Mount Sinai School of Medicine, New York, NY
2014 Weill Cornell Medical College, New York, NY
2014 University of Southern California, Los Angeles, CA
2015 Columbia Epidemiology Seminar, New York, NY
2016 American Society of Human Genetics Annual Meeting, Vancouver, Canada
2017 Society for Developmental Biology 11th Structural Birth Defects Meeting, Bethesda, MD
2017 Massachusetts General Hospital, Harvard Medical School, Boston, MA
2018 Baylor College of Medicine, Houston, TX
2018 International Conference on Intelligent Biology and Medicine, Los Angeles, CA
2018 Cold Spring Harbor Asia Meeting, Autism & Neurodevelopment Disorders – from Genetic Discoveries to Interventions, Suzhou, China
2018 Columbia Global Centers, Beijing Symposium on Precision Medicine, Beijing, China

Awards and honors

2000 Graduation with Distinction, Peking University, Beijing
1999 Novo Nordisk Scholarship, Peking University, Beijing
1997, 98, 99 Merit Student, Peking University, Beijing

Professional Organizations and Societies

2009-present Member, American Society of Human Genetics
2013-present Editorial Board, Scientific Reports
2013-present Reviewer for grants (W. M. Keck Foundation, UK Medical Research Council, US-Israel Binational Science)
2007- present Referee for Peer-reviewed Journals *(*such as Cell, Nature Biotechnology, New England Journal of Medicine, Genome Biology, Genome Research, Nucleic Acids Research, Bioinformatics, American Journal of Human Genetics, eLife, Journal of Immunology, Bioinformatics)