CURRICULUM VITAE

Updated: May 2018

Yufeng Shen

Assistant Professor of Systems Biology and Biomedical Informatics
Columbia University

Address: 1130 St Nicholas Ave, 803B, New York, NY 10032
Phone: 212-851-4662
Web: http://www.columbia.edu/~ys2411/
Email: ys2411@cumc.columbia.edu

Appointments

2011–present	Assistant Professor, Department of Systems Biology, Department of Biomedical Informatics, Columbia University
2011–present	Associate Director of Bioinformatics, JP Sulzberger Columbia Genome Center, Columbia University

Education

2007	Ph.D., Computational Biology, advised by Dr. George Weinstock Baylor College of Medicine
2000	B.Sc., Molecular Biology and Biochemistry, and Honor Science Program, Peking University, Beijing

Training

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

Research Support

Grants as PI or core director

2017-2022	NIH R01 GM120609 *Integrate cancer genomics data in genetic studies and diagnosis of developmental disorders*, Role: PI.
2017-2019	NIH R03 HL138352 *Genetic analysis of structural birth defects by integration of multiple diseases with epigenomic data and cancer mutations*, Role: PI.
2017-2022	NIH P01 HD068250-06A1, *Gene Mutation and Rescue in Human Diaphragmatic Hernia*, Role: MPI (PD: Donahoe P).
2017-2022	NIH P01 HD093363, *Developmental Mechanisms of Trachea-Esophageal Birth Defects*, Role: Director of Genomics Core (PD: Zorn A).
2016-2021	NIH U19 AI128949, *Human anti-viral immune responses in tissues and circulation.* Role: Co-Director of Data Analysis Core (PD: Farber DL)

2013-2018 NIH P01 AI106697, *Tissue compartmentalization of human lymphocytes*, Role: PI of Bioinformatics Core (PD: Farber DL).

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

Grants as co-Investigator

2017-2022 NIH U01 AI100119, *Development of lung T cell responses in infant respiratory immunity*. Role: Co-Investigator (PI: Farber DL)

2016-2021 NIH R01 DK52431-21A, *Molecular Genetic Analysis of Human Obesity*. Role: Co-Investigator (PI: Leibel R and Chung WK)

2015-2020 NIH U01 HG008680, *Columbia GENIE (GENomic Integration with Ehr)*, part of eMERGE network. Role: Co-Investigator (PD: Weng C)

2015-2019 NIH R01 HG008157, *Novel genetic association methods accounting for haplotype background*. Role: Co-Investigator (PI: Guan Y)

Pending grants

2018-2022 NIH R01 GM131005-01 *Developing new computational methods to interpret genome sequencing data*. Role: PI

Previous grants

2012-2017 NIH U01 HL098163, *Molecular approaches to gene identification in congenital heart disease*. Role: Co-Investigator (PD: Chung WK)

2014-2017 NYSTEM C029562, *Mapping development of human dendritic cells*. Role: Co-Investigator (PI: Liu K)

2011-2015 NIH DP3 DK094400, *Epigenetic events underlying Type I Diabetes*. Role: Co-Investigator (PD: Clynes RA)

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

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

Selected Professional Activities

2015–present Teaching: 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

2010–present Reviewer for academic journals (*Cell, Nature Biotechnology, New England Journal of Medicine, Genome Biology, Genome Research, Nucleic Acids Research, Bioinformatics, American Journal of Human Genetics, Journal of Immunology*, etc)

2013–present Editorial Board, *Scientific Reports*

2013–present Reviewer for grants (W. M. Keck Foundation, UK Medical Research Council, US-Israel Binational Science)

Graduate Student/Postdoctoral Fellow Mentoring

2018–present Dr. Lu Qiao (Postdoc, jointly with Dr. Wendy Chung)
2017–present Dr. Xueya Zhou (Postdoc, jointly with Dr. Wendy Chung)
2017–present Dr. Haicang Zhang (Postdoc)
2017–present Siying Chen (PhD student, Integrated Program)
2017–present Haoquan Zhao (PhD student, DBMI)
2017–present Chen Chen (PhD student, Biological Sciences)
2016–present Alexander Kitaygorodsky (PhD student, DBMI)
2016–present Na Zhu (Postdoc, jointed with Dr. Wendy Chung)
2015–present Alexander L. Hsieh (PhD student, DBMI)
2014–present Dr. Wenji Ma (Postdoc)
2014–present Hongjian Qi (PhD student, Applied Physics)

2015–2017 Xinwei Han (Postdoc): currently Scientist at Constellation Pharmaceuticals
2012–2017 Boris Grinshpun (PhD student): currently Postdoc fellow at EMD Serono
2016–2016 Elise Flynn (Rotation student, DSB): currently PhD student in Lappalainen group
2015–2016 Kyung In Kim (Postdoc): currently a Professor in Statistics at Inha University, Korea
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 Scientist at Flatiron Health
2014–2015 Hong Zhang (visiting PhD student): Fudan University
2012–2014 Badri Vardarajan (Postdoc): currently Assistant Professor at Columbia University
2012–2014 Patrick Cheung (Postdoc jointly with Chung and Wang groups): currently a Scientist at Philips Research
2012–2014 Edwin Lin (Master student): currently MD/PhD student at University of Utah
2011–2013 Casey Overby (Postdoc): currently Assistant Professor at Johns Hopkins
2013–2013 Andrew Chiang (Rotation student, DBMI), currently PhD student in Vitkup group
2013–2013 Jing He (Rotation PhD student, DBMI), current postdoc at Stanford
2011–2012 Yaping Feng (Postdoc): currently Research Scientist at Rutgers University

PhD Thesis committee:

2017–present Xiangtian Tan, Biological Sciences (Califano group)
2016–present Filip Cvetkovski, Immunology (Farber group)
2016–present 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)

Research Interest

I'm interested in computational genomics and genetics of human diseases. My group focuses on three areas: (a) Develop computational methods for analyzing large-scale genomic data. We are developing computational methods to accurately detect genetic variants from sequencing data, predict genetic effect

of DNA mutations, and predict gene dosage sensitivity based on gene expression and epigenomic profiles. (b) Genetics of human diseases. We are working on projects to identify genetic causes of human diseases using statistical and computational approaches with integration of biological mechanism. (c) Study clonal dynamics of immune cells using high-throughput sequencing and computational modeling.

Publications

Selected Publications in Peer-reviewed Journals

1. 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, *bioRxiv* 205849. Accepted to publication at *Nature Communications*
2. SPARK Consortium&, (2018) SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research, *Neuron,* PMID: 29420931
3. 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
4. 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
5. Zhang C, **Shen Y#**, (2017) A cell type-specific expression signature predicts haploinsufficient autism-susceptibility genes. *Human Mutation*, PMID: 27860035
6. 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
7. 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
8. 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
9. Chen C, Qi H, **Shen Y**, Pickrell J, Przeworski M. (2017) Contrasting Determinants of Mutation Rates in Germline and Soma. *Genetics*, PMID: 28733365
10. Qi H, Dong C, Chung WK, Wang K, **Shen Y#**, (2016) Deep Genetic Connection between Cancer and Developmental Disorders. *Human Mutation*, PMID: 27363847
11. 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. *PNAS*, PMID: 27261081

12. Thome JJ*, Grinshpun B*, Kumar BV, Kubota M, Ohmura Y, Lerner H, Sempowski GD, **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
13. 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
14. DeWolf S, **Shen Y**, Sykes M. (2016) A New Window into the Human Alloresponse. *Transplantation*, PMID: 26760572
15. 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
16. 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
17. 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.
18. 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
19. 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
20. 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
21. 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
22. 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
23. **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
24. **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
25. **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
26. 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

27. 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

28. 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
29. 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
30. 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
31. 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
32. 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
33. 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
34. 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
35. 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
36. 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
37. 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
38. 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
39. 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
40. 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

41. 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
42. 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
43. 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
44. Yu L, Bennett JT, Wynn J, Carvill GL, Cheung YH, **Shen Y**, Mychaliska GB, Azarow KS, Crombleholme TM, Chung DH, Potoka D, Warner BW, Bucher B, Lim FY, Pietsch J, Stolar C, Aspelund G, Arkovitz MS, University of Washington Center for Mendelian G, 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
45. Yu L, Wynn J, Cheung YH, **Shen Y**, Mychaliska GB, Crombleholme TM, Azarow KS, Lim FY, Chung DH, Potoka D, Warner BW, Bucher B, Stolar C, Aspelund G, Arkovitz MS, 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
46. 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
47. 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
48. 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
49. DeStefano GM, Fantauzzo KA, Petukhova L, Kurban M, Tadin-Strapps M, Levy B, Warburton D, Cirulli ET, Han Y, Sun X, **Shen Y**, Shirazi M, Jobanputra V, Cepeda-Valdes R, Cesar Salas-Alanis J, 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
50. **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

51. 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
52. 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
53. 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
54. Daly AK, Donaldson PT, Bhatnagar P, **Shen Y**, Pe'er I, Floratos A, Daly MJ, Goldstein DB, John S, Nelson MR, Graham J, Park BK, Dillon JF, Bernal W, Cordell HJ, Pirmohamed M, Aithal GP, 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
55. 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
56. **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
57. 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
58. Rhesus Macaque Genome Sequencing Consortium& (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316: 222-34. doi: 10.1126/science.1139247
59. 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
60. Sea Urchin Genome Sequencing Consortium& (2006) The genome of the sea urchin Strongylocentrotus purpuratus. *Science* 314: 941-52. doi: 10.1126/science.1133609
61. 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
62. **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

63. 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
64. Qi H*, Yu L*, Zhou X*, Kitaygorodsky A, et al, Chung WK#, **Shen Y#**, (2017) Genetic analysis of *de novo* variants reveals sex differences in complex and isolated congenital diaphragmatic hernia and indicates *MYRF* as a candidate gene, *bioRxiv*, 206037

# indicates corresponding authorship

* indicated co-first authorship

Underlined indicates lab members

& indicates consortium member

Google Scholar: https://scholar.google.com/citations?user=e3UEBXAAAAAJ&sortby=pubdate