Xiao Fan

Phone number: +1 (507) 369 3864
Email: xiaofan.hit@gmail.com
Website: http://xiaofanhit.wix.com/home

Work experience

- 2021.8 – present Columbia University (in the labs of Drs. Wendy Chung and Yufeng Shen)

Associate research scientist in Pediatrics Department

- 2018.7 – 2021.7 Columbia University (in the labs of Drs. Wendy Chung and Yufeng Shen)

Postdoctoral research scientist in Pediatrics Department

- 2016.8 – 2018.6, Mayo Clinic (in the lab of Dr. Iftikhar Kullo)

Postdoctoral fellow in Cardiovascular Department

Education

- 2011.9 – 2016.6, University of Alberta (in the lab of Dr. Lukasz Kurgan)

Ph.D. in Electrical and Computer Engineering, GPA: 4.0/4.0

- 2009.9 – 2011.7, Harbin Institute of Technology (in the lab of Dr. Ye Zhang)

M.Sc. in Information Engineering, GPA: 89/100

- 2005.9 – 2009.7, Ocean University of China

B.Sc. in Electronic Information Science and Technology, GPA: 86/100

Research experience

1. Genetic association (journal articles 5, 8-10, conference article 2-5, 7, 8)
- Study genetic causes of common conditions using genome-wide association study and phenome-wide association study.
- Discover novel genes/loci for rare inherited diseases especially for cardiovascular diseases and cancer using next generation sequencing data.

2. Variant pathogenicity (journal articles 1, 2, 4, 6, 11, conference article 1 and 6)
- Specialized in variant classification based on American College of Medical Genetics guideline.
- Improve variant interpretation by leveraging phenotypic and expression data.

3. RNA sequencing (journal articles 7, 13, 16, 18)
- discovery of novel microRNA from short RNA sequencing data and expression differential analysis.
- Identification of bacteria using 16s rRNA sequencing data and expression differential analysis.

4. Machine learning (journal 14, 15, 17, 19-21, conference article 1, 9-11)
- Image enhancement and classification
- Prediction of intrinsically disordered protein
- Protein crystallization prediction
- MicroRNA target prediction
- Disease diagnosis using electronic medical record (EMR)
- Prediction of variant pathogenicity

Skills

- Machine learning: feature transformation and selection, clustering, regression, classification.
- Deep learning: PyTorch, keras, sklearn.
- Next-generation sequencing analysis: DNA-seq, RNA-seq, microRNA sequencing and 16s rRNA

sequencing.

- Bioinformatics database: NCBI, EMBL, UCSC, GTEx, GEO, Uniprot, ClinVar, gnomAD, etc.
- Bioinformatics tools: GATK, BLAST, VEP, ANNOVAR, samtools, PLINK, KING, etc.
- Programming languages: C++, R, Bash and Python.

Awards & Grants

- Principal investigator for NIH Pathway to Independence Award (K99/R00 Independent), 2021-2026
- Co-investigator for the American Heart Association Institute for Precision Cardiovascular Medicine Data Ming Grants, 2017
- Andrew Stewart Memorial Graduate Prize (up to 20 doctoral students annually), 2015
- Alberta Innovates Graduate Student Scholarship, 2014-2015
- Distinguished Graduate Dissertation Award, 2011
- Distinguished Undergraduate Dissertation Award, 2009
- Excellent Completion in Innovative Experimental Program, 2007-2009

Publications

1. H. Zhang, M.S. Xu, **X. Fan**, et al. Predicting functional effect of missense variants using graph attention neural networks, *Nature Machine Intelligence*, 2022, revision.
2. S.M. Ware, et al. The genetic architecture of pediatric cardiomyopathy, *AJHG*, 2022 Feb.
3. Z. Wang, **X. Fan,** et al. Non-cancer-related pathogenic germline variants and expression consequences in ten-thousand cancer genomes, *Genome Med.*, 2021 Sep.
4. **X. Fan**, et al. Penetrance of breast cancer susceptibility genes from the eMERGE III Network, *JCNI Cancer Spectr.*, 2021 May.
5. E. Breidbart, et al. Frequency and characterization of mutations in genes in a large cohort of patients referred to MODY registry, *J Pediatr Endocrinol Metab.*, 2021 April.
6. S. Saadatagah, et al. Genetic basis of hypercholesterolemia in adults, *npj Genom. Med.*, 2021 April.
7. L. Boyle, et al. Genotype and defects in microtubule-based motility correlate with clinical severity in KIF1A Associated Neurological Disorder, *HGG Advances*, 2021 April; 2(2).
8. R. Cuella-Martin, **X. Fan**, et al. Functional interrogation of genetic variants of the DNA damage response with CRISPR-dependent base editing screens, *Cell*. 2021 Feb; 184(4):1081-1097.
9. J. Bain, et al. Detailed Clinical and Psychological Phenotype of the X-linked HNRNPH2-related Neurodevelopmental Disorder. *Neurol Genet.* 2021 Jan; 7(1):e551.
10. E. Rosenthal, et al. Association between triglycerides, known risk SNVs and conserved rare variation in SLC25A40 in a multi-ancestry cohort. *BMC Med Genet*. 2021 Jan; 14(1):11.
11. A. Vosoughi, et al. Common germline-somatic variant interactions in advanced urothelial cancer. *Nat Commun*. 2020 Dec; 11(1):6195.
12. J. Groenendyk, **X. Fan**, et al. Endoplasmic reticulum and the microRNA environment in the cardiovascular system. *Can J Physiol Pharmacol*. 2019 Jun; 97(6):515-27.
13. **X. Fan**, M.S. Safarova, et al**.** Targeted Sequencing Study to Uncover Shared Genetic Susceptibility Between Peripheral Artery Disease and Coronary Heart Disease. *ATVB.* 2019 June; 39(6):1227-33.
14. S. Gordon, et all. DENND5B Regulates Intestinal Triglyceride Absorption and Body Mass. *Sci. Rep.* 2019 Mar; 9(1):3597.

15. M.S. Safarova, B.A. Satterfield, **X. Fan**, et al**.** A Phenome-Wide Association Study to Discover Pleiotropic Effects of *PCSK9*, *APOB* and *LDLR*. *npj Genom Med*. 2019 Feb; 4(1):3.
16. I.J. Kullo, J. Olson, **X. Fan**, et al. The Return of Actionable Variants Empiric (RAVE) Study, a Mayo Clinic Genomic Medicine Implementation Study: Design and Initial Results. *Mayo Clin Proc.* 2018 Nov; 93(11):1600-10.
17. **X. Fan**, L. Kurgan. Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Brief Bioinform*. 2015 Sept; 16(5):780-94.
18. L.A. Ferreira, **X. Fan**, et al. Analyzing the effects of protecting osmolytes on solute-water interactions by solvatochromic comparison method: II. Globular proteins. *RSC Adv.* 2015 July; 5:59780-91.
19. Z. Peng, J. Yan, **X. Fan**, et al. Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci*. 2015 Jan; 72(1):137-51.
20. J. Groenendyk, **X. Fan**, et al. Genome-wide analysis of thapsigargin-induced microRNAs and their targets in NIH3T3 cells. *Genom Data*. 2014 Dec; 2:325-7.
21. M.J. Mizianty, **X. Fan**, et al. Covering complete proteomes with X-ray structures: A current snapshot. *Acta Crystallogr D Biol Crystallogr*. 2014 Nov; 70(11):2781-93.
22. J. Groenendyk, Z. Peng, E. Dudek, **X. Fan**, et al. Interplay between the oxidoreductase PDIA6 and microRNA-322 controls the response to disrupted endoplasmic reticulum calcium homeostasis. *Sci Signal*. 2014 Jun; 7(329):ra54.
23. **X. Fan**, B. Xue, et al. The intrinsic disorder status of the human hepatitis C virus proteome. *Mol Biosyst*. 2014 Jun; 10(6):1345-63.
24. L.A. Ferreira, **X. Fan**, et al. Structural features important for differences in protein partitioning in aqueous dextran-polyethylene glycol two-phase systems of different ionic compositions. *Biochim Biophys Acta*. 2014 Mar; 1844(3):694-704.
25. **X. Fan**, L. Kurgan. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J Biomol Struct Dyn*. 2014; 32(3):448-64.

Conference and commentary articles:

1. **X. Fan**, et al. Characteristics of Pathogenicity for In-frame Insertion and Deletion Variants. *2022 NHGRI Annual Training & Career Development Meeting*.
2. **X. Fan**, et al. Modeling the ACMG/AMP Guidelines as a Quantitative Approach. *2019 American Society of Human Genetics (ASHG) Meeting*.
3. M. Jose, **X. Fan**, et al. Differences in Prevalence of Phenotypically and Genotypically Ascertained Familial Hypercholesterolemia in a cohort with Hypercholesterolemia. *2018 American College of Cardiology (ACC) Scientific Sessions*, Volume 71, Issue 11 Supplement, March 2018.
4. H. Sandhyavenu, **X. Fan**, et al. Association of Rare Functional Variants in Triglyceride Related Genes with Circulating Triglyceride Levels. *2018 American College of Cardiology (ACC) Scientific Sessions*, Volume 71, Issue 11 Supplement, March 2018
5. B. Satterfield, **X. Fan**, et al. A Phenome-Wide Association Study to Discover Pleiotropic Effects of *LDLR* and *APOB*. *American College of Medical Genetics and Genomics (ACMG) meeting 2018*.
6. I.J. Kullo, **X. Fan**, K. Ding. Genetic Risk, Lifestyle, and Coronary Artery Disease. *N Engl J Med*. 2017 Mar; 376(12):1192-3.

7. **X. Fan**, M.S. Safarova, et al. Targeted sequencing of 109 genes in the eMERGEseq panel uncovers novel variants and genes influencing triglyceride levels. *2017 American Society of Human Genetics (ASHG) Meeting*.
8. **X. Fan**, M.S. Safarova, et al. Phenotype-assisted Interpretation of Variants in Genes Causal for Familial Hypercholesterolemia. *Circulation*. 2017;136:A14644.
9. M.S. Safarova, **X. Fan**, et al. A phenome-wide Association Study to Assess Pleiotropic Effects of *LPA*. *Circulation. 2017;136:A20688*.
10. Z. Ye, **X. Fan**, et al. Risk factors associated with early-versus late-accelerated growth patterns in patients with abdominal aortic aneurysm. Journal of the American College of Cardiology. 2017 Mar 21; 69: (11)2024.
11. **X. Fan**, L. Kurgan. Accurate high-throughput prediction of non-canonical microRNA targets. *2015 RiboWest*.
12. Y. Zhang, **X. Fan**, et al. Linear spectral unmixing with generalized constraint for hyperspectral imagery. *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 4106-4109.
13. **X. Fan**, Y. Zhang, et al. A robust spectral target recognition method for hyperspectral data based on combined spectral signatures. *2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 4328-4331.