Han Sun, PhD

hansun@stanford.edu • LinkedIn • GitHub • 408-800-1190

EXPERIENCE

Research Scientist/Senior Computational Biologist **2020 –**

Department of Pediatrics, School of Medicine, Stanford University, **Prof. Anna Gloyn**

Postdoc **2015 – 2019**

Department of Genetics, School of Medicine, Stanford University, **Prof. Lars Steinmetz**

EDUCATION

PhD, Bioinformatics **2009 – 2014**

University of Chinese Academy of Sciences, **Prof. Yixue Li** and **Prof. Lu Xie**

Bachelor, Software Engineering **2004 – 2008**

East China Normal University

SKILLS

- Python, R, C, git, conda, slurm, Linux, HPC, and cloud computing (AWS and GCP)
- Containerization (Docker and Singularity) and workflow management (Snakemake and Nextflow)
- Bioinformatics (blast, plink, bwa, samtools, STAR, featureCounts, GATK, seurat, scanpy)
- Biostatistics (hypothesis testing, confidence interval, power calculation, bayesian inference, DESeq2, statsmodels)
- Datasets (Ensembl, dbSNP, gnomAD, GTEx, COSMIC, TCGA, ENCODE, GEO, SRA, EGA, dbGaP, All of Us)
- Multiomics: Genomics - whole genome sequencing (WGS), whole exome sequencing (WES), genotyping (genotyping array); Transcriptomics - bulk and single cell RNA-Seq (10x Genomics/Smart-Seq2), long reads sequencing (PacBio/Nanopore); Proteomics - Label-free and Labelled tandem mass spectrometry (MS/MS); Epigenomics - bulk and single nucleus ATAC-Seq (chromatin accessibility), whole genome bisulfite sequencing and methylation array (methylation status), ChIP-Seq/CUT&RUN (transcription factor binding)
- Linkage analysis, quantitative trait locus mapping (QTL), genome-wide association studies (GWAS), imputation, permutation, fine-mapping, and colocalization
- Machine Learning (scikit-learn: linear model, generalized linear model, Random Forrest; dataset splitting, cross validation, overfitting) and Deep Learning (PyTorch/Tensorflow: CNN, ResNet, Transformer, GPT, etc.)

PROJECTS

Splicing-mediated transcript fusion and oncogene activation in multiple cancers

- Designed the project to explore splicing-mediated gene fusion events in cancer patients.
- Reanalyzed both whole exome sequencing (WES) and RNA-Seq data from over 10,000 patients across 33 cancer types from the TCGA project using a negative binomial-based generalized linear model.

- Discovered significant correlations between somatic mutations and the splicing of adjacent genes (transcript fusion), including the recurrent tumor-specific CDKN1A-RAB44 hit in bladder, stomach, and skin cancers.

- Lead the collaboration with other labs to validate the finding using CRISPR/Cas9 editing, RT-PCR, Sanger sequencing, qPCR, RNA-Seq, and phenotyping assays.

- Analyzed the data, interpreted the results, and prepared the manuscript. Proposed a splicing-mediated readthrough stabilization (SRS) model as a novel mechanism for oncogene activation.

RNA-binding protein RBM20-deficient dilated cardiomyopathy

- Identified a suspicious missense mutation in RBM20 via linkage analysis in a family affected by dilated cardiomyopathy through analyzing whole exome sequencing (WES) and whole genome sequencing (WGS) data.

- Collaborated with colleagues to validate the mutation using CRISPR/Cas9 editing in human induced pluripotent stem (iPS) cells and differentiating them into cardiomyocytes.

- Identified splicing targets of RBM20 using RNA-Seq data from humans and pigs given that RBM20 is an RNA-binding protein.

- By applying drug repositioning strategy in thousands of compounds from the Integrated Network-based Cellular Signatures (LINCS) project, we nominated retinoic acid and cardiac glycoside that may potentially rescue the phenotype.

- By integrating gene expression data from thousands of tissues in the The Genotype-Tissue Expression (GTEx) project, I proposed a model for the tissue-specific splicing of a mitochondrial inner membrane protein, IMMT, suggesting a molecular link between energy supply deficiencies and dilated cardiomyopathy.

Mutation and regulatory elements underlying neonatal diabetes

- Identified a suspicious intronic mutation in neonatal diabetes by analyzing whole genome sequencing (WGS) and array-based genotyping data from mice.

- Collaborated with colleagues to successfully validate that the mutation is causing diabetes by introducing it into both B6 and NOD mice using CRISPR/Cas9 editing.

- Proposed a mechanism where the mutation creates a novel splice site, resulting in a longer transcript but truncated protein, which has been verified as well by RT-PCR.

- Working with bulk and single-cell data (scRNA-Seq and scATAC-Seq) throughout development to investigate potential extra regulatory elements and understand why the phenotype is specific to pancreas.

- Working on large-scale QTL mapping (eQTL, caQTL, and pQTL) on multiomics datasets of human islets to study the genetic architecture of diabetes.

Applying machine learning/deep learning in omics data

- Applied machine learning models (Random Forrest, XGBoost, SVM, etc.) to predict ancestry from genotyping data.

- Developed deep learning models (CNN, ResNet, GPT, Llama2) in predicting selenocysteine and tissue-specific splicing from DNA sequences.

CERTIFICATIONS

- Machine Learning Specialization, Andrew Ng, Coursera

- Deep Learning Specialization, Andrew Ng, Coursera

PATENTS

- Methods of Treatment, Genetic Screening, and Disease Models for Heart Conditions Associated with RBM20 Deficiency. Francesca Briganti, Lars M. Steinmetz, Han Sun, and Wu Wei. WO 2020/092171.

PUBLICATIONS

1. **Han Sun**, Huiying Yan, Kathryn Bieging-Rolett, Michelle Nguyen, William F. Mueller, Zhuanfen Cheng, Hong Zeng, Laura Attardi, Wu Wei, and Lars M. Steinmetz et al. "CDKN1A-RAB44 transcript fusion and oncogene activation in cancers.", bioRxiv, **co-first author**.

2. Francesca Briganti, **Han Sun**, Wu Wei, Jingyan Wu, Chenchen Zhu, Martin Liss, Ioannis Karakikes et al. "iPSC Modeling of RBM20-Deficient DCM Identifies Upregulation of RBM20 as a Therapeutic Strategy." Cell Reports 32, no. 10 (2020): 108117, **co-first author**.

3. **Han Sun**, Seth A Sharp, Nicole A J Krentz, Anna Gloyn et al., "Deep learning prediction of selenocystine element and its application in diabetes", in preparation, **co-first author**.

4. Russ-Silsby, James, Yunkyeong Lee, Varsha Rajesh, **Han Sun**, Mahsa Amoli, Nasser Ali Mirhosseini, Tushar Godbole, Matthew B. Johnson et al. "Complete Loss of PAX4 causes Transient Neonatal Diabetes in Humans." Molecular Metabolism (2025): 102201.

5. de Winter, Twan JJ, Miha Sovrovic, Esmee Dekker, Natascha de Graaf, **Han Sun**, James D. Johnson, Anna L. Gloyn, Francoise Carlotti, Eelco JP de Koning, and Anna Alemany. "Cell type-specific eQTL detection from single-cell RNA-seq reveals post-transcriptional regulatory mechanisms in human islets." bioRxiv (2025): 2025-01.

6. Evans-Molina, Carmella, Yasminye D. Pettway, Diane C. Saunders, Seth A. Sharp, Thomas SR Bate, **Han Sun**, Heather Durai et al. "Heterogeneous endocrine cell composition defines human islet functional phenotypes." bioRxiv (2024): 2024-11.

7. Kolic, Jelena, WenQing Grace Sun, Haoning Howard Cen, Jessica D. Ewald, Jason C. Rogalski, Shugo Sasaki, **Han Sun**, et al. "Proteomic predictors of individualized nutrient-specific insulin secretion in health and disease." Cell Metabolism 36, no. 7 (2024): 1619-1633.

8. Ewald, Jessica D., Yao Lu, Cara E. Ellis, Jessica Worton, Jelena Kolic, **Han Sun**, Shugo Sasaki, Dahai Zhang et al. "HumanIslets. com: Improving accessibility, integration, and usability of human research islet data." Cell Metabolism (2024).

9. Rottner, Antje K., Yingying Ye, Elena Navarro-Guerrero, Varsha Rajesh, Alina Pollner, **Han Sun**, Romina J. Bevacqua, Jing Yang et al. "A genome-wide CRISPR screen identifies CALCOCO2 as a regulator of beta cell function influencing type 2 diabetes risk." Nature Genetics 55, no. 1 (2023): 54-65.

10. Lau, Hwee Hui, Nicole AJ Krentz, Fernando Abaitua, Marta Perez-Alcantara, **Han Sun**, Jun-Wei Chan, Jila Ajeian, Soumita Ghosh et al. "PAX4 loss of function increases diabetes risk by altering human pancreatic endocrine cell development." Nature Communications 14, no. 1 (2023): 6119.

11. Mattis, Katia K., Nicole AJ Krentz, Christoph Metzendorf, Fernando Abaitua, Aliya F. Spigelman, **Han Sun**, Jennifer M. Ikle et al. "Loss of RREB1 in pancreatic beta cells reduces cellular insulin content and affects endocrine cell gene expression." Diabetologia 66, no. 4 (2023): 674-694.

12. Cortez, Briana N., Hui Pan, Samuel Hinthorn, **Han Sun**, Nicola Neretti, Anna L. Gloyn, and Cristina Aguayo-Mazzucato. "Heterogeneity of increased biological age in type 2 diabetes correlates with differential tissue DNA methylation, biological variables, and pharmacological treatments." GeroScience (2023): 1-21.

13. Torres, Jason M., **Han Sun**, Vibe Nylander, Damien J. Downes, Martijn van de Bunt, Mark I. McCarthy, Jim R. Hughes, and Anna L. Gloyn. "Inferring causal genes at type 2 diabetes GWAS loci through chromosome interactions

in islet cells." Wellcome Open Research 8 (2023).

14. Alghamdi, Tamadher A., Nicole AJ Krentz, Nancy Smith, Aliya F. Spigelman, Varsha Rajesh, Alokkumar Jha, **Han Sun**, Mourad Ferdaoussi et al. "Zmiz1 is required for mature β-cell function and mass expansion upon high fat feeding." Molecular Metabolism 66 (2022): 101621.

15. Chenchen Zhu, Jingyan Wu, **Han Sun**, Francesca Briganti, Benjamin Meder, Wu Wei, and Lars M. Steinmetz. "Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes." Nature Communications 12, no. 1 (2021): 1-9.

16. Benedikt Rauscher, William F Mueller, Sandra Clauder-Münster, Petra Jakob, M Saiful Islam, **Han Sun**, Sonja Ghidelli-Disse, Markus Boesche, Marcus Bantscheff, Hannah Pflaumer, Paul Collier, Bettina Haase, Songjie Chen, Rene Hoffman, Guangwen Wang, Vladimir Benes, Gerard Drewes, Michael Snyder, Lars M Steinmetz. "Patient-derived gene and protein expression signatures of NGLY1 deficiency", Journal of Biochemistry, 2021

17. Semih Calamak, Menekse Ermis, **Han Sun**, Saiful Islam, Michael Sikora, Michelle Nguyen, Vasif Hasirci, Lars M. Steinmetz, and Utkan Demirci. "A Circulating Bioreactor Reprograms Cancer Cells Toward a More Mesenchymal Niche." Advanced Biosystems 4, no. 2 (2020): 1900139, **co-first author**.

18. Jay W Schneider, Saji Oommen, Muhammad Y Qureshi, Sean C Goetsch, Rhianna S Sundsbak, Wei Guo, Mingming Sun, **Han Sun**, Dennis A Webster, Alex W Coutts, Francesca Briganti, Wu Wei, Lars Steinmetz, Daniel F Carlson, and Timothy J. Nelson et al. "Dysregulated ribonucleoprotein granules promote cardiomyopathy in RBM20 gene-edited pigs", Nature Medicine (2020): 1-13.

19. William F. Mueller, Petra Jakob, **Han Sun**, Sandra Clauder-Münster, Sonja Ghidelli-Disse, Diana Ordonez, Markus Boesche, Marcus Bantscheff, Paul Collier, Bettina Haase, Vladimir Benes, Malte Paulsen, Peter Sehr, Joe Lewis, Gerard Drewes, Lars M. Steinmetz. "Loss of N-glycanase 1 Alters Transcriptional and Translational Regulation in K562 Cell Lines." G3: Genes, Genomes, Genetics (2020).

20. Arne H. Smits, Frederik Ziebell, Gerard Joberty Nico Zinn, William F. Mueller, Sandra Clauder-Munster, Paola Grandi, Petra Jakob, Anne-Marie Michon, **Hanice Sun**, Karen Tessmer, Tilmann Burckstummer, Marcus Bantscheff, Lars M. Steinmetz, Gerard Drewes, and Wolfgang Huber. "Biological plasticity rescues target activity in CRISPR knock outs.", Nature Methods 2019 Oct 28:1-7.

21. Suo, Lun, Yu Xiao Zhou, Li Ling Jia, Hai Bo Wu, Jin Zheng, Qi Feng Lyu, Li Hua Sun, **Han Sun**, and Yan Ping Kuang. "Transcriptome profiling of human oocytes experiencing recurrent total fertilization failure." Scientific Reports 8, no. 1 (2018): 17890.

22. **Han Sun**, Chen Chen, Baofeng Lian, Menghuan Zhang, Xiaojing Wang, Bing Zhang, Yixue Li, Pengyuan Yang, and Lu Xie. 2015. "Identification of HPV Integration and Gene Mutation in HeLa Cell Line by Integrated Analysis of RNA-Seq and MS/MS Data." Journal of Proteome Research 14 (4): 1678–86, **co-first author**.

23. **Han Sun**, Chen Chen, Meng Shi, Dandan Wang, Mingwei Liu, Daixi Li, Pengyuan Yang, Yixue Li, and Lu Xie. 2014. "Integration of Mass Spectrometry and RNA-Seq Data to Confirm Human Ab Initio Predicted Genes and lncRNAs." Proteomics 14 (23-24): 2760–68, **co-first author**.

24. **Han Sun**, Xiaobin Xing, Jing Li, Fengli Zhou, Yunqin Chen, Ying He, Wei Li, et al. 2013. "Identification of Gene Fusions from Human Lung Cancer Mass Spectrometry Data." BMC Genomics 14 Suppl 8 (December): S5, **co-first author**.

25. Zhen-Hao Liu, Bao-Feng Lian, Qiong-Zhu Dong, **Han Sun**, Jin-Wang Wei, Yuan-Yuan Sheng, Wei Li, et al. 2018. "Whole-Exome Mutational and Transcriptional Landscapes of Combined Hepatocellular Cholangiocarcinoma and Intrahepatic Cholangiocarcinoma Reveal Molecular Diversity." Biochimica et Biophysica Acta 1864 (6 Pt B): 2360–68.

26. Menghuan Zhang, Hong Li, Ying He, **Han Sun**, Li Xia, Lishun Wang, Bo Sun, et al. 2015. "Construction and Deciphering of Human Phosphorylation-Mediated Signaling Transduction Networks." Journal of Proteome Research 14 (7): 2745–57.

27. Wei Li, Jian Yu, Baofeng Lian, **Han Sun**, Jing Li, Menghuan Zhang, Ling Li, Yixue Li, Qian Liu, and Lu Xie. 2015. "Identifying Prognostic Features by Bottom-up Approach and Correlating to Drug Repositioning." PloS One 10 (3): e0118672.

28. Jing Li, Jia Jia, Hong Li, Jian Yu, **Han Sun**, Ying He, Daqing Lv, et al. 2014. "SysPTM 2.0: An Updated Systematic Resource for Post-Translational Modification." Database: The Journal of Biological Databases and Curation 2014 (April): bau025.

29. Yulin Dai, Shengdi Li, Xiao Dong, **Han Sun**, Chao Li, Zhi Liu, Beili Ying, Guohui Ding, and Yixue Li. 2013. "The de Novo Sequence Origin of Two Long Non-Coding Genes from an Inter-Genic Region." BMC Genomics 14 (Suppl 8): S6.

30. Li Xia, Tong-Dan Wang, Shao-Ming Shen, Meng Zhao, **Han Sun**, Ying He, Lu Xie, et al. 2013. "Phosphoproteomics Study on the Activated PKCδ-Induced Cell Death." Journal of Proteome Research 12 (10): 4280–4301.

31. Jian Yu, Xiaobin Xing, Lingyao Zeng, Jiehuan Sun, Wei Li, **Han Sun**, Ying He, et al. 2012. "SyStemCell: A Database Populated with Multiple Levels of Experimental Data from Stem Cell Differentiation Research." PloS One 7 (7): e35230.

32. Ying He, Menghuan Zhang, Yuanhu Ju, Zhonghao Yu, Daqing Lv, **Han Sun**, Weilan Yuan, et al. 2011. "dbDEPC 2.0: Updated Database of Differentially Expressed Proteins in Human Cancers." Nucleic Acids Research 40 (D1): D964–71.

33. Xiao Chang, Yun Li, Jie Ping, Xiao-Bin Xing, **Han Sun**, Peng Jia, Chuan Wang, Yuan-Yuan Li, and Yi-Xue Li. 2011. "EcoBrowser: A Web-Based Tool for Visualizing Transcriptome Data of Escherichia Coli." BMC Research Notes 4 (October): 405.

34. Xiao-Bin Xing, Qing-Run Li, **Han Sun**, Xing Fu, Fei Zhan, Xiu Huang, Jing Li, et al. 2011. "The Discovery of Novel Protein-Coding Features in Mouse Genome Based on Mass Spectrometry Data." Genomics 98 (5): 343–51.