Han Sun, PhD

hansun@stanford.com • LinkedIn • GitHub • 3165 Porter Drive, Palo Alto, CA 94304

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

PROJECTS

- **Splicing-mediated transcript fusion and oncogene activation in multiple cancers**

With an idea to explore gene fusion events on the transcription level instead of the typical genomic level events in cancer patients, I reanalyzed both whole exome sequencing and RNA-Seq data of over 10,000 patients across 33 cancer types from the TCGA project using a negative binomial-based generalized linear model. I observed significant correlations between somatic mutations and the splicing between adjacent genes (transcript fusion) in subpopulations, including the top hit, mutation within CDKN1A, and the splicing between CDKN1A and RAB44 in bladder, stomach, and skin cancer. I established the causality through collaboration with other labs by introducing the mutation in cell lines using CRISPR/Cas9 and then confirming the appearance of the splicing between CDKN1A and RAB44 using RT-PCR and Sanger sequencing, as well as the activation of RAB44 using qPCR and RNA-Seq. We further performed functional/phenotyping assays to illustrate the connection between RAB44 activation and MDM2 upregulation (TP53 downregulation). I proposed a splicing-mediated readthrough stabilization (SRS) model as one more mechanism for oncogene activation.

- **RNA-binding protein RBM20-deficient dilated cardiomyopathy**

Starting from a family with dilated cardiomyopathy, I analyzed whole exome sequencing (WES) and whole genome sequencing (WGS) data from the family members and identified a suspicious missense mutation in RBM20 through linkage analysis. Considering RBM20 is an RNA-biding protein, I further identified its splicing targets from human and pig RNA-Seq data. We validated this finding by doing CRISPR/Cas9 editing in human induced pluripotent stem (iPS) cells and differentiating them to cardiomyocytes. By integrating gene expression data of thousands of tissues from GTEx, I proposed a model about tissue-specific splicing of a mitochondrial inner membrane protein, IMMT, suggesting a molecular connection between deficiency in energy-supplying and dilated cardiomyopathy.

- **Mutation and regulatory elements underlying neonatal diabetes**

By analyzing whole genome sequencing and array based genotyping data from mice, I identified a suspicious intronic mutation that correlates with neonatal diabetes. We validated that the mutation is indeed diabetes causing using CRISPR/Cas9 editing. I proposed a mechanism that the mutation leads to a novel splice site resulting in longer isoform but truncated protein which has also been verified. We are working on bulk and single cell data (scRNA-Seq and scATAC-Seq) along development to investigate additional regulatory elements to explain why the phenotype is pancreas specific.

- **Deep learning prediction of tissue specific splicing**

After reproducing the famous model, SpliceAI, using both Tensorflow and Pytorch, I am attempting and evaluating models such as attention with GPT backbone, and fourier transform based global convolution (Hyena), to replace the CNN component in SpliceAI, with an application in long-reads sequencing data (PacBio) from multiple tissues including islets.

RESEARCH INTERESTS

- Disease Genetics (Cancer, Cardiomyopathy, Diabetes, …)
- Multiomics (bulk and single cell)
- AI 4 Life Science and Drug Discovery

SCIENTIFIC SKILLS

- Python, R, C, git, Linux, and High Performance Computing
- statsmodels, scikit-learn, Tensorflow, and PyTorch
- Bioinformatics and Biostatistics
- Machine Learning and Deep Learning
- Genomics, Transcriptomics, Proteomics, and Epigenomics
- Linkage analysis, QTL mapping, and GWAS

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.
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. 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.
4. 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.
5. 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.
6. 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.
7. 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).
8. 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.
9. 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.
10. 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

11. 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**.

12. 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.

13. 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).

14. 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.

15. 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.

16. **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**.

17. **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**.

18. **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**.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.