Jeremy Li

Bellevue, WA 98006 · h.jeremy.li@gmail.com

Summary

Statistical geneticist and engineer. I have extensive experience in fast-paced startup environments, hiring and rapidly growing R&D teams, and interfacing with executive, lab, and product teams.

Experience

Staff Research Engineer 12/2024 – present
Stealth Biotech Startup, USA

- Unifying disparate multi-TB genomic datasets for novel clinical genomics applications and developing methods for family-based genomic analyses
- Developing imputation methods for reconstructing offspring genotypes from ultra-low coverage sequencing data

Director of Data Science 02/2022 – 12/2024
Head of Data Science 10/2021 – 02/2022
Gencove, New York, NY

- Led research team (7 PhD-level scientists) in developing methods and algorithms for analysis of low-coverage WGS data
- Set company-level R&D strategy alongside executive and product teams and led the execution thereof
- Spearheaded deep learning initiatives (PyTorch) to replace expert-driven models, reducing manual overhead by 30%+ while accelerating analysis pipelines
- Designed and maintained genomics pipelines processing petabytes of genomic data in a high-throughput AWS environment, decreasing compute costs by >25%

Senior Data Scientist – Statistical Genetics 03/2021 – 10/2021
Data Scientist – Statistical Genetics (employee #5) 02/2019 – 02/2021
Gencove, New York, NY

- Owned dozens of statistical genetics projects end-to-end, involving DNA-seq analysis, statistical modeling, CNV/SV analysis, genomic prediction, and GWAS
- Built out the initial production workflows for genomic data, achieving stable parallel processing of tens of thousands of low-coverage WGS samples
- Authored three first-author publications and presented at major genomics conferences

Bioinformatics and Data Scientist Lead (employee #4) 07/2017 – 02/2019
Genomic Prediction, North Brunswick, NJ

- Main contributor to the core bioinformatics pipeline for aneuploidy and polygenic score screening, resulting in hundreds of screened pre-implantation embryo samples

Education

University of Washington 2013 – 2017
B.S. in Physics and B.A. in Chemistry

Selected Skills

Languages: Python, R, bash, C++, Rust, Nextflow, Snakemake

Tools: Docker, AWS (EC2, S3, ECR, Batch), git, GNU/Linux, cluster computing

Bioinformatics: Sequence analysis, variant calling, copy number variation, genotype imputation, ancestry inference, etc.

Statistics: Hidden Markov Models, likelihood-based models, Bayesian inference

Selected Datasets: 1000 Genomes, UK Biobank, HRC

Posters & Presentations

Poster; American Society for Human Genetics Conference, 2022
Poster; Association for Medical Pathology Conference, 2022
Talk; Korean Genome Organization Annual Conference, 2021
Poster; American Society for Human Genetics Conference, 2020
Talk; International Quantitative Genetics Conference, 2020
Poster; 8th Human Genetics in NYC Conference, 2019
Poster; Complex Traits Consortium/Rat Genomics 17th Annual Meeting, 2019

Selected Publications

J.H. Li, ..., S. Kruglyak. (2024). "Low-pass sequencing plus imputation using avidity sequencing displays comparable imputation accuracy to sequencing by synthesis while reducing duplicates." *G3: Genes—Genomes—Genetics*. doi.org/10.1093/g3journal/jkad276

J.H. Li, ..., L. Flagel. (2023). "The effects of reference panel perturbations on the accuracy of genotype imputation." *bioRxiv*. doi.org/10.1101/2023.08.10.552684

J.H. Li, ..., J.K. Pickrell. (2021). "Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays." *Genome Research*. doi.org/10.1101/gr.266486.120

J.H. Li, ..., S. Huang. (2019). "Time-dependent saddle node bifurcation: Breaking time and the point of no return in a non-autonomous model of critical transitions." *Physica D: Nonlinear Phenomena*. doi.org/10.1016/j.physd.2019.02.005

B. Wenz, ..., **J.H. Li**, et al. (2024) "Genotype inference from aggregated chromatin accessibility data reveals genetic regulatory mechanisms." *bioRxiv*. doi.org/10.1101/2024.09.04.610850

C. Kaelin, ..., **J.H. Li**, et al. (2024). "Ancestry dynamics and trait selection in a designer cat breed." *Current Biology*. doi.org/10.1016/j.cub.2024.02.075

B. Noyvert, ..., **J.H. Li**, et al. (2023). "Imputation of structural variants using a multi-ancestry long-read sequencing panel enables identification of disease associations." *medRxiv*. doi.org/10.1101/2023.12.20.23300308

W. Snelling, ..., **J.H. Li**, et al. (2020). "Assessment of imputation from low-pass sequencing to predict merit of beef steers." *Genes, 11*(11), 1312. doi.org/10.3390/genes11111312

E. Petter, ..., **J.H. Li**, et al. (2020). "Relative matching using low coverage sequencing." *bioRxiv*. doi.org/10.1101/2020.09.09.289322

K. Wasik, ..., **J.H. Li**, et al. (2019). "Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics." *bioRxiv, 632141*. doi.org/10.1101/632141

N.R. Treff, ..., **J.H. Li**, et al. (2019). "Validation of concurrent preimplantation genetic testing for polygenic and monogenic disorders, structural rearrangements, and whole and segmental chromosome aneuploidy with a universal testing platform." *European Journal of Medical Genetics*. doi.org/10.1016/j.ejmg.2019.04.004