



## HoJoon Lee

Sr Biomedical Data Scientist, Medicine - Med/Oncology

 Curriculum Vitae available Online

### Bio

---

#### CURRENT ROLE AT STANFORD

Senior Data Scientist

#### HONORS AND AWARDS

- 2nd place in Stanford HealthAI hackathon, Stanford University (2019 Jan)
- Travel grants, AACR (2015 April)
- Travel Fellowship, International Society for Computational Biology (ISMB) meeting (2015)
- Best poster award in Oncology retreat, Stanford University (2014 October)
- Travel fellowships of School of Life Sciences, Arizona State University (2009 September)
- GPSA Conference Travel Grant, Arizona State University (2008)
- Travel Fellowship, RECOM Computational Cancer Biology (2007)
- Dr. John and Rose Maher Alumni Scholarship for students involved in cancer research, Arizona State University (2006)

#### EDUCATION AND CERTIFICATIONS

- PhD, Arizona State University , Molecular & Cellular Biology with Bioinformatics (2012)
- Master of Science, Arizona State University , Computational Biology (2005)
- Bachelor of Science, Yonsei University , Biology (2002)

#### PATENTS

- HoJoon Lee (co-inventor). "United States Patent US PTO 62/200,904 High Resolution STR analysis using Next Generation Sequencing", Leland Stanford Junior University

### Professional

---

#### PROFESSIONAL INTERESTS

My primary research interest is "Cancer Treatment and Prevention through Precision Medicine" based on the analysis of genomic data of cancer patients for using genomics information to make proper clinical decisions. During my B.S in biology major, I was captivated by the fact that DNA – digital information - codes life. To gain computational skill, I joined the computational biology program for a master's degree. I learned the principle and techniques of various sequence analyses such as sequence alignment models, molecular phylogenetics, and motif searching from estimating the neural mutation rate by comparing human and mouse genomes. To have an impact on real life, I then applied my existing expertise to cancer sequencing data in order to identify neo-antigens for breast cancer vaccine development during my Ph.D. study. I learned how to analyze RNA-seq data by my own

algorithm and organize/manage large data generated from the project using MySQL. As a post-doc at Stanford, I expanded my research to investigate the clinical implications of genomic features. I applied regularized regression (Elastic-net) to integrate multiple, heterogeneous genomic assays data from the Cancer Genome Atlas (TCGA) project and identify known and novel candidate driver mutations that predict tumor stage and other clinical parameters. As a data scientist, my innovative works include a novel k-mer representation technique for genome sequence analysis without traditional alignment methods, which has been used to characterize the human pangenome reference. Currently, I coordinates interdisciplinary teams combining expertise in genomics, computer science, pathology, and immunology to advance a deep learning model that classifies cell types in hematoxylin and eosin-stained (H&E) images.

## WORK EXPERIENCE

- Senior Research Engineer - Stanford University (3/16/2019 - present)
- Project Leader - Stanford University (3/16/2017 - 3/15/2019)
- Post-doctoral fellow - Stanford University (3/16/2012 - 3/15/2017)
- Research Associate - Arizona State University (8/2005 - 2/2012)

## Publications

---

### PUBLICATIONS

- **Single cell functional characterization of cancer mutations and their cellular phenotype**  
Sun, H., Lee, D., Lee, H., Grimes, S. M., Wood, R., Ji, H. P.  
AMER ASSOC CANCER RESEARCH.2026
- **AI foundation model for single cell annotation from conventional histopathology images of cancer.**  
Bai, X., Lee, H., Tan, X., Li, C., Sathe, A., Wang, Y., Nguyen, Q., Ji, H. P.  
AMER ASSOC CANCER RESEARCH.2026
- **The Gastric Cancer Genome Explorer: A spatial and single cellular registry for identifying novel cellular markers and targets.**  
Ji, H. P., Balusu, N., Cha, J., Wichmann, I., Grimes, S. M., Lee, H., Wang, Y., Sathe, A.  
LIPPINCOTT WILLIAMS & WILKINS.2026: 429
- **CLONEID: A Framework for Longitudinal Integration of Phenotypic and Genotypic Data to Monitor and Steer Subclonal Dynamics.** *bioRxiv : the preprint server for biology*  
Veith, T., Beck, R., Tagal, V., Li, T., Alahmari, S., Cole, J., Hannaby, D., Kyei, J., Yu, X., Maksin, K., Schultz, A., Lee, H., ElNaqa, et al  
2025
- **The single-cell spatial landscape of stage III colorectal cancers.** *NPJ precision oncology*  
Su, A., Lee, H., Tran, M., Dela Cruz, R. C., Sathe, A., Bai, X., Wichmann, I., Pflieger, L., Moulton, B., Barker, T., Haslem, D., Jones, D., Nadauld, et al  
2025; 9 (1): 101
- **The Gastric Cancer Registry: A multi-omic cellular and molecular resource for cancer biomarker and therapeutic discovery.**  
Ji, H. P., Meka, R., Perez, I., Grimes, S. M., Lee, H., Wang, Y., Sathe, A.  
LIPPINCOTT WILLIAMS & WILKINS.2025: 491
- **Detection and analysis of complex structural variation in human genomes across populations and in brains of donors with psychiatric disorders** *Cell*  
Zhou, B., Arthur, J. G., Guo, H., et al  
2024; Published online September 30, 2024
- **Pan-conserved segment tags identify ultra-conserved sequences across assemblies in the human pangenome.** *Cell reports methods*  
Lee, H., Greer, S. U., Pavlichin, D. S., Zhou, B., Urban, A. E., Weissman, T., Ji, H. P.  
2023; 3 (8): 100543
- **Pangenome graph construction from genome alignments with Minigraph-Cactus** *NATURE BIOTECHNOLOGY*  
Hickey, G., Monlong, J., Ebler, J., Novak, A. M., Eizenga, J. M., Gao, Y., Marschall, T., Li, H., Paten, B., Abel, H. J., Antonacci-Fulton, L. L., Asri, M., Baid, et al

2023

- **Single-molecule methylation profiles of cell-free DNA in cancer with nanopore sequencing.** *Genome medicine*  
Lau, B. T., Almeda, A., Schauer, M., McNamara, M., Bai, X., Meng, Q., Partha, M., Grimes, S. M., Lee, H., Heestand, G. M., Ji, H. P.  
2023; 15 (1): 33
- **A draft human pangenome reference.** *Nature*  
Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J., Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, et al  
2023; 617 (7960): 312-324
- **The Gastric Cancer Registry Genome Explorer: A tool for genomic discovery.**  
Almeda, A., Grimes, S. M., Shin, G., Lee, H., Wichmann, I., Greer, S., Ji, H. P.  
LIPPINCOTT WILLIAMS & WILKINS.2023: 434
- **Tumor-associated microbiome features of metastatic colorectal cancer and clinical implications.** *Frontiers in oncology*  
An, H. J., Partha, M. A., Lee, H., Lau, B. T., Pavlichin, D. S., Almeda, A., Hooker, A. C., Shin, G., Ji, H. P.  
2023; 13: 1310054
- **Colorectal cancer metastases in the liver establish immunosuppressive spatial networking between tumor associated SPP1+ macrophages and fibroblasts.** *Clinical cancer research : an official journal of the American Association for Cancer Research*  
Sathe, A., Mason, K., Grimes, S. M., Zhou, Z., Lau, B. T., Bai, X., Su, A., Tan, X., Lee, H., Suarez, C. J., Nguyen, Q., Poultides, G., Zhang, et al  
2022
- **Exploratory genomic analysis of high grade neuroendocrine neoplasms across diverse primary sites.** *Endocrine-related cancer*  
Sun, T. Y., Zhao, L., Van Hummelen, P., Martin, B., Hornbacker, K., Lee, H., Xia, L. C., Padda, S. K., Ji, H. P., Kunz, P.  
2022
- **The Gastric Cancer Registry: A Genomic Translational Resource for Multidisciplinary Research in Gastric Cancer.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*  
Almeda, A. F., Grimes, S. M., Lee, H., Greer, S., Shin, G., McNamara, M., Hooker, A. C., Arce, M. M., Kubit, M., Schauer, M. C., Van Hummelen, P., Ma, C., Mills, et al  
2022
- **KmerKeys: a web resource for searching indexed genome assemblies and variants.** *Nucleic acids research*  
Pavlichin, D. S., Lee, H., Greer, S. U., Grimes, S. M., Weissman, T., Ji, H. P.  
2022
- **A deep learning model for molecular label transfer that enables cancer cell identification from histopathology images.** *NPJ precision oncology*  
Su, A., Lee, H., Tan, X., Suarez, C. J., Andor, N., Nguyen, Q., Ji, H. P.  
2022; 6 (1): 14
- **Analysis of 16S rRNA sequencing in advanced colorectal cancer tissue samples**  
An, H., Partha, M. A., Lee, H., Lau, B., Shin, G., Almeda, A., Ji, H. P.  
LIPPINCOTT WILLIAMS & WILKINS.2022
- **Characterization of the consensus mucosal microbiome of colorectal cancer.** *NAR cancer*  
Zhao, L., Grimes, S. M., Greer, S. U., Kubit, M., Lee, H., Nadauld, L. D., Ji, H. P.  
1800; 3 (4): zcab049
- **Profiling diverse sequence tandem repeats in colorectal cancer reveals co-occurrence of microsatellite and chromosomal instability involving Chromosome 8.** *Genome medicine*  
Shin, G., Greer, S. U., Hopmans, E., Grimes, S. M., Lee, H., Zhao, L., Miotke, L., Suarez, C., Almeda, A. F., Haraldsdottir, S., Ji, H. P.  
2021; 13 (1): 145
- **Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangenome to individual infection quasispecies.** *Genome medicine*  
Lau, B. T., Pavlichin, D., Hooker, A. C., Almeda, A., Shin, G., Chen, J., Sahoo, M. K., Huang, C. H., Pinsky, B. A., Lee, H. J., Ji, H. P.  
2021; 13 (1): 62

- **Single Cell Analysis Can Define Distinct Evolution of Tumor Sites in Follicular Lymphoma.** *Blood*  
Haebe, S. E., Shree, T. n., Sathe, A. n., Day, G. n., Czerwinski, D. K., Grimes, S. n., Lee, H. n., Binkley, M. S., Long, S. R., Martin, B. A., Ji, H. P., Levy, R. n.  
2021
- **Unique k-mer sequences for validating cancer-related substitution, insertion and deletion mutations.** *NAR cancer*  
Lee, H., Shuaibi, A., Bell, J. M., Pavlichin, D. S., Ji, H. P.  
2020; 2 (4): zcaa034
- **SPATIAL SINGLE-CELL ANALYSIS OF COLORECTAL CANCER TUMOUR USING MULTIPLEXED IMAGING MASS CYTOMETRY**  
Minh Tran, Su, A., Lee, H., Cruz, R., Pflieger, L., Dean, A., Quan Nguyen, Ji, H., Rhodes, T.  
BMJ PUBLISHING GROUP.2020: A399
- **CRISPRpic: fast and precise analysis for CRISPR-induced mutations via prefixed index counting.** *NAR genomics and bioinformatics*  
Lee, H., Chang, H. Y., Cho, S. W., Ji, H. P.  
2020; 2 (2): lqaa012
- **Entire landscape of epitopes from all possible missense mutations in human coding sequences.**  
Lee, H., Greer, S., Ji, H. P.  
AMER ASSOC CANCER RESEARCH.2020: 118–19
- **Comprehensive genomic sequencing of high-grade neuroendocrine neoplasms**  
Sun, T., Van Hummelen, P., Martin, B., Xia, C., Lee, H., Zhao, L., Hornbacker, K., Ji, H., Kunz, P. L.  
AMER SOC CLINICAL ONCOLOGY.2020
- **Gastric Cancer Registry: A comprehensive patient-reported resource for multidisciplinary and translational genomic approaches to gastric cancer**  
Almeda, A., Hooker, A., Lee, H., Mills, M., Van Hummelen, P., Ford, J. M., Ji, H.  
AMER SOC CLINICAL ONCOLOGY.2020
- **Whole genome analysis identifies the association of TP53 genomic deletions with lower survival in Stage III colorectal cancer.** *Scientific reports*  
Xia, L. C., Van Hummelen, P. n., Kubit, M. n., Lee, H. n., Bell, J. M., Grimes, S. M., Wood-Bouwens, C. n., Greer, S. U., Barker, T. n., Haslem, D. S., Ford, J. M., Fulde, G. n., Ji, et al  
2020; 10 (1): 5009
- **Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangenome to individual infection quasispecies.** *medRxiv : the preprint server for health sciences*  
Lau, B. T., Pavlichin, D. n., Hooker, A. C., Almeda, A. n., Shin, G. n., Chen, J. n., Sahoo, M. K., Huang, C. n., Pinsky, B. A., Lee, H. n., Ji, H. P.  
2020
- **Author Correction: RNA Transcription and Splicing Errors as a Source of Cancer Frameshift Neoantigens for Vaccines.** *Scientific reports*  
Shen, L. n., Zhang, J. n., Lee, H. n., Batista, M. T., Johnston, S. A.  
2020; 10 (1): 6251
- **RNA Transcription and Splicing Errors as a Source of Cancer Frameshift Neoantigens for Vaccines.** *Scientific reports*  
Shen, L., Zhang, J., Lee, H., Batista, M. T., Johnston, S. A.  
2019; 9 (1): 14184
- **Targeted short read sequencing and assembly of re-arrangements and candidate gene loci provide megabase diplotypes.** *Nucleic acids research*  
Shin, G. n., Greer, S. U., Xia, L. C., Lee, H. n., Zhou, J. n., Boles, T. C., Ji, H. P.  
2019
- **Therapeutic Monitoring of Circulating DNA Mutations in Metastatic Cancer with Personalized Digital PCR.** *The Journal of molecular diagnostics : JMD*  
Wood-Bouwens, C. M., Haslem, D. n., Moulton, B. n., Almeda, A. F., Lee, H. n., Heestand, G. M., Nadauld, L. D., Ji, H. P.  
2019
- **Author Correction: RNA Transcription and Splicing Errors as a Source of Cancer Frameshift Neoantigens for Vaccines.** *Scientific reports*  
Shen, L. n., Zhang, J. n., Lee, H. n., Batista, M. T., Johnston, S. A.

2019; 9 (1): 17815

- **SVEngine: an efficient and versatile simulator of genome structural variations with features of cancer clonal evolution.** *GigaScience*  
Xia, L. C., Ai, D., Lee, H., Andor, N., Li, C., Zhang, N. R., Ji, H. P.  
2018
- **Mapping the comprehensive landscape of missense-mutation neoantigens across the human genome**  
Lee, H., Greer, S. U., Ji, H. P.  
AMER ASSOC CANCER RESEARCH.2018
- **Improved detection and identification of microsatellite instability features in colorectal cancer: Implications for immunotherapy**  
Shin, G., Lee, H., Grimes, S. M., Kubit, M. A., Ji, H. P.  
AMER ASSOC CANCER RESEARCH.2018
- **High-quality CNV segments from low-coverage whole genome sequencing from FFPE cancer biopsies based on an evaluation of multiple CNV tools**  
Lee, H., Xia, L., Greer, S., Bell, J., Grimes, S. M., Bouwens, C., Shin, G., Lau, B. T. C., Johnson, L., Andor, N., Day, K., Miller, M., Escobar, et al  
AMER ASSOC CANCER RESEARCH.2018
- **Single-Color Digital PCR Provides High-Performance Detection of Cancer Mutations from Circulating DNA.** *The Journal of molecular diagnostics : JMD*  
Wood-Bouwens, C., Lau, B. T., Handy, C. M., Lee, H., Ji, H. P.  
2017; 19 (5): 697-710
- **CRISPR-Cas9-targeted fragmentation and selective sequencing enable massively parallel microsatellite analysis** *NATURE COMMUNICATIONS*  
Shin, G., Grimes, S. M., Lee, H., Lau, B. T., Xia, L. C., Ji, H. P.  
2017; 8
- **The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations** *GENOME MEDICINE*  
Lee, H., Palm, J., Grimes, S. M., Ji, H. P.  
2015; 7
- **The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations.** *Genome medicine*  
Lee, H., Palm, J., Grimes, S. M., Ji, H. P.  
2015; 7 (1): 112-?
- **Systematic genomic identification of colorectal cancer genes delineating advanced from early clinical stage and metastasis** *BMC MEDICAL GENOMICS*  
Lee, H., Flaherty, P., Ji, H. P.  
2013; 6
- **Systematic genomic identification of colorectal cancer genes delineating advanced from early clinical stage and metastasis.** *BMC medical genomics*  
Lee, H., Flaherty, P., Ji, H. P.  
2013; 6: 54-?