

OMID SHAMS SOLARI

 solari.stat.berkeley.edu
 solari@berkeley.edu
 osolari

 U.S. Green Card Holder
 Piedmont, CA 94611
 (510) 542-7514

AREAS OF EXPERTISE

Generative AI (VAEs, Transformers), (protein) Language Models, Optimization, Interpretable ML/AI, High-Dimensional Statistical Learning, Multi-View Representation Learning, Hierarchical Bayesian Inference, Bio-Marker Discovery.

EDUCATION

University of California, Berkeley

Ph.D., Statistics

Summer 2019

Large Scale Interpretable Multi-View Learning with Application To High-Dimensional Multi-Omics Advisors & Thesis Committee: Prof. Peter J. BICKEL, Dr. James B. BROWN, Prof. Gary KARPEN & Prof. Haiyan HUANG.

- **MuLe**: Formulated and implemented a novel large-scale interpretable and stable multi-view learning method for high-dimensional datasets with low sample size and utilized it to infer co-regulated modules between Metabolomics, Transcriptomics, and Microbiomics data collected on fruitfly samples in a treatment/control experiment.
- **BLOCCS**: Extended **MuLe** to perform block matrix decomposition leading to more interpretable orthogonal canonical covariates.
- **SparkLe**: Formulated and implemented a novel multiple kernel learning approach applied to kernel dimensionality reduction, implemented in.
- Implemented a spark implementation of the three previous packages on **Spark-MLlib**.

University of California, Berkeley

M.A., Mathematics

Fall 2013

Thesis Advisers: Prof. Per-Olof PERSSON & Prof. Alberto GRUNBAUM.

- Numerically efficient implementation and benchmarking of second-order PDE solutions, specifically the wave equation (**Python** and **C++**).
- *Advection-Corrected Correlation Image Velocimetry*: Numerical Computation of the velocity fields of Saturn's surface vortices from Cassini images.

Sharif University of Technology

B.Sc., Computational Physics (B.Sc. in Mechanical Engineering and Aerospace Engineering) Spring 2011

Thesis Adviser: Prof. Saeid SAEIDI.

- Computational particle image velocimetry, modeling the transfer and deposition patterns of aerosols in the upper airways via finite volume analysis in **C++** and **OpenFoam**.

RESEARCH EXPERIENCE

Lead AI Research Scientist

September 2024 - Present

Natera Inc.

Precision Oncotherapeutics AI Modeling Lead

- Implemented a end-to-end **Vision Transformer (ViT)** pipeline for neoantigen prioritization which predicts the HLA binding likelihood of a given neopeptide.

- Developed a Protein **Language Model** to predict post-translational modification likelihoods from short protein sequences.
- Designed and developed a data pipeline and a predictive model to infer patient-specific somatic mutations associated with differential response to platinum-based therapies.

Lead Statistician and Machine Learning Research Scientist March 2022 - September 2024
Natera Inc.
Head of MCED/ECD/tfMRD/COO Biomarker Discovery Algorithms (Ranked “exceeding” (top 10%) for 2 consecutive years)

- Designed and implemented **CpGTools** which is the software infrastructure and the computational pipeline of the newly established Early Cancer Detection (**ECD**) and tissue-free Minimal Residual Disease (**tf-MRD**) program which is now the basis for all computational methylation efforts at Natera.
- Designed and implemented the first completely reference-free methylation biomarker discovery pipeline. (Patent Application Submitted)
- Developed a novel **Hierarchical Variational Bayes** model to infer differentially methylated biomarkers from deep whole genome bisulfite sequencing data. (Patent Application Submitted)
- Implemented a **transformer VAE** to learn the whole genome methylation patterns of healthy cfDNA samples.
- Designed Natera’s CRC **ECD/tfMRD** methylation biomarker panel. (Patent Application Submitted)
- Formed and **directed** the Cell-Type of Origin Pursuit Group (C2pG). A team of 4 dedicated to the cancer cell of origin problem.
- Designed Natera’s first **MCED** biomarker panel

AI Lead Scientist April 2021 - March 2022
Ravel Biotechnology Inc.
Cancer Prediction Model, Seq2Seq modeling and motif discovery, Cell-type Decomposition

- Developed deep **Seq2Seq** models to infer functional events, e.g. histone modifications from accessibility patterns in cfDNA and utilized attribution priors to discover highly predictive motifs from **regularized attributions**.
- Developed a **variational Bayes** model to infer cell-type of origin from functional events e.g. methylation and histone modification.

Senior Machine Learning Research Scientist March 2020 - April 2021
Ravel Biotechnology Inc.
Fragmentomics panel design, Cancer/no-cancer prediction

- Designed Ravel’s fragmentomics targeted panel
- Developed Ravel’s main high-dimensional low sample size cancer/no cancer prediction framework.
- Utilized integrated gradients to find cancer predictive DNA fragmentation patterns

Post-Doctoral Associate (*Chan Zuckerberg BioHub Initiative*) October 2019 - March 2020
University of California, Berkeley Statistics Department
Developed Explainable AI models for Genomic Data
Advisors: Prof. Bin YU, Dr. James B. BROWN & Prof. Michael MAHONEY

Contract/Non-Profit Statistical & Machine Learning Consultant September 2017 - March 2020
Department of Statistics, University of California, Berkeley
Provided consulting services on a variety of topics e.g. Deep Learning, Neuro-science, Linguistics, Public Health, International Development, Genomics, Bio-chemistry etc. to academic and industrial clients.

Statistical Learning Research Associate

October 2015 - March 2020

Lawrence Berkeley National Laboratory

Collaborated with the Molecular Eco-Systems Biology Division and the Daphnia Consortium in mining, modeling and interpretation of genomic data.

Ensemble Machine Learning Research Intern

Summer 2017

Illumina Inc.

Designed and implemented DEnsLe, a deep variational ensemble learner, in PyTorch, to construct an ensemble learner that models the dependence structure between multiple different SNV callers.

Lead Data Analyst

January - July 2014

Genapsys Inc.

- Implemented a MATLAB pipeline to perform QC of raw output signals of Genius[®] short-read genome sequencers.
- Statistical analysis of the sensor array network output, e.g. clustering, correlation analysis, and base-calling.

PUBLICATIONS/POSTERS

- “Comprehensive analysis of differentially methylated regions in colorectal cancer (CRC).” **Solari, Omid** et al. *American Society of Clinical Oncology*, 2024
- “Sa1175 Comparison of Colorectal Cancer (CRC) Characteristics Across Genetic Ancestries: Implications for Early Cancer Detection.” Myer, Parvathi A; ... ; **Solari, Omid Shams** et al. *Gastroenterology*, 2024
- “Circulating differential methylation allele fraction (DMAF) strongly correlates with circulating tumor DNA (ctDNA) variant allele fraction (VAF)”. Srinivasan, Preethi, ... , **Solari, Omid Shams** et al. *Cancer Research*, 2024
- “Tail Probability Estimation of Factor Models with Regularly-Varying Tails: Asymptotics and Efficient Estimation.”, **Shams Solari, Omid** and Pourbabae, Farzad (co-first authors) *SSRN*, 2022 <https://arxiv.org/abs/1903.12299>
- “Cell-free DNA fragments inform epigenomic mechanisms for early detection of breast cancer.” Gafni, E. and Harvey, A. and Jaroszewicz, A. and **Solari, Omid Shams** et al. (co-first author) *Cancer Research*, 2021
- “Sparse Canonical Correlation Analysis via Concave Minimization”, **Omid Shams Solari**, James B. Brown, Peter J. Bickel, <https://arxiv.org/abs/1909.07947>
- “BLOCCS: Block Sparse Canonical Correlation Analysis With Application To Interpretable Omics Integration”, **Omid Shams Solari**, Rojin Safavi, James B. Brown, <https://arxiv.org/abs/1909.07944>
- “Exploiting regulatory heterogeneity to systematically identify enhancers with high accuracy”, Hamutal Arbel, ..., **Omid Shams Solari**, et al. *Proceedings of the National Academy of Sciences* 116, no. 3 (2019): 900-908.
- “Patterns of DNA Methylation in Daphnia: An Evolutionary Perspective”, Jouni Kvist, Camila G. Athanàsio, **Omid Shams Solari**, et al. *Genome biology and evolution* (2018)
- “Early transcriptional response pathways in Daphnia magna are coordinated in networks of crustacean-specific genes”, Luisa Orsini , James B. Brown, **Omid Shams Solari** et al. *Molecular Ecology* 2017 Jul 26. doi: 10.1111/mec.14261.
- “Daphnia magna transcriptome by RNA-Seq across 12 environmental stressors”, Luisa Orsini, Donald Gilbert, Ram Podicheti, Mieke Jansen, James B Brown, **Omid Shams Solari** et al. *Nature Scientific Data*. 2016; 3: 160030

CURRENT PROJECTS

- “Cell-Type of Origin Inference from cfChIP Data via Tensor Regression”, **Omid Shams Solari**
- “Tissue of Origin Inference via Hierarchical Bayesian Regression”, **Omid Shams Solari**
- “Detecting Epithelial Ovarian Cancer and Patient Prognosis from Shallow Whole-Genome Sequencing of Cell-Free DNA (cfDNA) via Bias-Reduced Generalized Linear Models”, **Omid Shams Solari**
- “Sparse Multiple Kernel Learning With Application To Interpretable Kernel Dimensionality Reduction”, **Omid Shams Solari**, James P. Duncan, James B. Brown, Peter J. Bickel
Code: <https://github.com/osolari/sMKL>
- “Large-Scale Multi-View Learning in PySpark”, James P. Duncan, **Omid Shams Solari**, James B. Brown, Fernando Perez
Package: <https://spark-packages.org/package/jpdunc23/sparkle>
JOSS report: <https://github.com/jpdunc23/sparkle-JOSS>

AWARDS

Chan Zuckerberg Biohub Initiative Award
Sally & Terry Speed Award, UC Berkeley
UC Berkeley Excellence Award
UC Berkeley Fellowship for Graduate Studies
UC San Diego Fellowship
University of British Columbia Scholarship