

James V. Talwar

Email: [jtalarwar@ucsd.edu](mailto:jtalwar@ucsd.edu); Website: jvtalwar.github.io

EDUCATION

University of California San Diego, Jacobs School of Engineering Sept. 2018 – Dec. 2024
PhD in Bioinformatics and Systems Biology; GPA: 3.98/4.00; Advisor: Hannah Carter, PhD La Jolla, CA

University of California San Diego, Jacobs School of Engineering Sept. 2021 – Mar. 2023
MS in Computer Science: Artificial Intelligence; GPA: 3.98/4.00 La Jolla, CA

- **Selected Coursework:** Neural Networks, Recommender Systems, Probabilistic Reasoning, Algorithm Design and Analysis, Graduate Networked Systems

University of Pennsylvania, School of Engineering and Applied Science Sept. 2010 – May 2014
Bachelor of Science in Engineering, Bioengineering and Mathematics (Major/Minor) Philadelphia, PA

- **Honors:** Dean's List; Graduation Honor *Cum Laude*

SKILLS

Programming Languages: Python, R, Java, Go, MATLAB

Machine Learning Libraries: PyTorch, Keras, Hugging Face Transformers, scikit-learn, Optuna, LangChain, DGL

SELECTED PROJECTS

Improving Genetic Risk Scores with Transformers Apr. 2020 – Present

- Devised **VADER**, a novel vision transformer-inspired architecture designed to capture disease-specific interactions and genetic heterogeneity in complex diseases.
- Developed **DARTH (Directed Attention Relevance from Transformer Heuristics)**, an interpretability metric that identifies different patterns of complex disease risk in individuals.
- Orchestrated distributed training with PyTorch DDP across 4 A30 GPUs, optimizing resource utilization and reducing training time by 80%.
- Achieved absolute increases of **0.7% in accuracy** (1% relative increase), **1.5% in average precision** (2.1% relative increase), and **0.014 in AUC** (1.96% relative increase) over state-of-the-art statistical genomics approaches to genomic risk scores in prostate cancer.

Self-Supervised Learning for Genomic Sequence Embeddings Jun. 2021 – Nov. 2023

- Implemented **BERT-style masked language model (MLM) pre-training** for genomic sequences, enabling functional stratification in the embedding space.
- Integrated **transformer and attention mechanisms** to the EUGENE package, accelerating the building and evaluation of sequence-based deep learning models in genomics.

Generalized Realignment of Innocuous and Essential Variants Otherwise Utilized as Skewed Jul. 2023 – Aug. 2024

- Created and released **GRIEVOUS**, a command-line tool that addresses **artifactual out-of-distribution performance drop-offs** in genomic risk scores by improving data quality and ensuring variant consistency across multiple datasets.
- **GRIEVOUS automates** the harmonization of variant indexing and allele assignments across multi-cohort datasets, ensuring **high-quality input data** for accurate and reliable genomic risk model training and inference.
- Code and documentation available at: <https://github.com/jvtalwar/GRIEVOUS>

Domain Retrieval-Augmented Generation On Novel Nov. 2023 – Present

Bioinformatics Resources Ensuring Accurate Truthful Help

- Conceptualized and designed **DRAGONBREATH**, an in-house personalized assistant for bioinformatics software, powered by both **closed (OpenAI GPT-4) and open (LLAMA)** models, based on user preference, and enhanced with **retrieval-augmented generation**.
- Demo available upon request.

INDUSTRY EXPERIENCE

Intheon

Jun. 2014 – Aug. 2014

Engineering Intern at startup company specializing in neurotechnology

San Diego, CA

- Investigated mapping real-time EEG signals to spatial brain activities using signal strength and direction.
- Developed interactive three-dimensional brain visualization models and updated GUI libraries.

ResMed

Jun. 2013 – Aug. 2013

Engineering Intern at global medical device company

Dublin, Ireland

- Constructed hardware and wrote software resulting in the development of an automated light sensor test system for non-contact sleep monitor evaluation.

SELECTED PUBLICATIONS AND POSTERS

1. **J.V. Talwar**, D. Laub, M.S. Pagadala, A. Castro, M. Lewis, G.E. Luebeck, B.R. Gorman, C. Pan, F.N. Dong, K. Markianos, C.C. Teerlink, J. Lynch, R. Hauger, S. Pyarajan, P.S. Tsao, G.P. Morris, R.M. Salem, W.K. Thompson, K. Curtius, M. Zanetti, H. Carter. Autoimmune alleles at the major histocompatibility locus modify melanoma susceptibility. *Am. J. Hum. Genet.* (2023).
2. **J.V. Talwar**, A. Klie, M.S. Pagadala, H. Carter. GRIEVOUS: your command-line general for resolving cross-dataset genotype inconsistencies. *Bioinformatics.* (2024)
3. A. Klie, D. Laub, **J.V. Talwar**, H. Stites, T. Jores, J.J. Solvason, E.K. Farley, H. Carter. Predictive analyses of regulatory sequences with EUGENE. *Nat. Comput. Sci.* (2023).
4. M. Pagadala, T.J. Sears, V.H. Wu, E. Pérez-Guijarro, H. Kim, A. Castro, **J.V. Talwar**, C. Gonzalez-Colin, S. Cao, B.J. Schmiedel, S. Goudarzi, D. Kirani, J. Au, T. Zhang, T. Landi, R.M. Salem, G.P. Morris, O. Harismendy, S.P. Patel, L.B. Alexandrov, J.P. Mesirov, M. Zanetti, C. Day, C.C. Fan, W.K. Thompson, G. Merlino, J.S. Gutkind, P. Vijayanand, H. Carter. Germline modifiers of the tumor immune microenvironment implicate drivers of cancer risk and immunotherapy response. *Nat. Commun.* (2023).
5. M.S. Pagadala, J.A. Linscott, **J.V. Talwar**, T.M. Seibert, B. Rose, J. Lynch, M. Panizzon, R. Hauger, M.H. Hansen, J.D. Sammon, M.H. Hayn, K. Kader, H. Carter & S.T. Ryan. PRState: Incorporating genetic ancestry in prostate cancer risk scores for men of African ancestry. *BMC Cancer.* (2022).
6. **J. Talwar** and H. Carter. Assessing cancer drug response prediction from gene expression. Poster presented at: AACR Annual Meeting 2020; April 27-28, 2020 and June 22-24, 2020; Philadelphia, PA