Hassaan Maan
PhD Candidate, University of Toronto
hassaan.maan@mail.utoronto.ca
https://hmaan.ca/ (Personal Website)
https://www.linkedin.com/in/hmaan (LinkedIn)
https://github.com/hsmaan (GitHub)

RESEARCH INTERESTS

- Representation learning of genome sequencing data, for salient tasks related to therapeutic and diagnosis development.
- Integration of large-scale multi-modal genomic and healthcare datasets for understanding of disease etiology and progression, and prediction of clinical outcomes.
- Machine learning method development for analysis of multi-modal genome sequencing data, such as multi-modal single-cell data, for discovery and understanding of complex multi-faceted biological processes.

EDUCATION

Doctor of Philosophy Candidate – Department of Medical Biophysics (Ph.D.) [4.0] September 2020-
University of Toronto, Toronto, ON

- Working on machine-learning methods for integrative multi -omics analysis with applications in pediatric medulloblastoma, primarily supervised by Dr. Bo Wang and co-supervised by Dr. Michael D. Taylor and Dr. Kieran R Campbell.
- Published thesis and adjacent work in various journals, including *Nature Methods*, *Nature Biotechnology*, *Cell - Immunity*, and *The Lancet Digital Health*.
- Supervised 3 undergraduate trainees on various projects related to machine learning in computational biology.
- Completed graduate-level coursework in Machine Learning (CS 2506), Statistical Methods for Machine Learning (STA2104), and Topics in Computational Biology: Artificial Intelligence in Medicine (CSC2431).
- Awarded the Natural Sciences and Engineering Research Council of Canada (NSERC) CGS-D Doctoral Scholarship for 2022-2025 (Top 1% of Applicants) ($105,000).
- Accepted at and attended Oxford Machine Learning (OxML) Summer School 2021.

Master of Bioinformatics (M.Binf.) [4.0 - Graduated with Honors] 2018-2019
University of Guelph, Guelph, ON

- Emphasis on learning typical bioinformatics workflows involving key databases; data wrangling, analysis, and visualization using R and Python; learning typical and novel algorithms for tasks such as alignment and clustering; and creating Bash pipelines in UNIX environments to analyze NGS data.
- Statistics and machine learning courses covering a variety of methods employed in bioinformatics workflows, from both a conceptual and practical applications standpoint.
- Seminar based courses covering many bioinformatics related topics, including clustering methods, algorithm design, integration of -omics data, RNA-seq, network and pathway analysis, and more.

Bachelor of Science – Biomedical Sciences (BSc.) [3.7 - Graduated with Honors] 2012-2017
University of Waterloo, Waterloo, ON

- Course and lab work focused on biology, with an elective focus on physics and physical chemistry.
- Relevant coursework: Genetics (Biol239), Principles of Molecular Biology (Biol308), Analytical Methods in Molecular Biology (Biol309), Molecular Biotechnology (Biol342), Human Molecular Genetics (Biol434), Fundamentals of Immunology (Biol341).
- Statistics background through life-science based statistics courses and statistical methods used in physics and physical chemistry courses/lab-work.

Summer Program in Clinical Epidemiology and Biostatistics May-July 2016

Toronto General Hospital, Toronto, ON

- Attended a 10-week series of seminars given by scientists and clinicians on clinical research methodology, clinical research design, epidemiology, and biostatistics.
- Gained knowledge of biostatistics and application in clinical epidemiology through critical appraisal of clinical research papers on infectious diseases, pharmaceutical research, and other biomedical research topics.
- Evaluated clinical research methodology and design, such as types of studies (observational, randomized controlled trials, etc.), regression models, and study biases.

EXPERIENCE

Graduate Research Intern (Immunomics - Antigen Map Project) June-September 2023

Microsoft Research, Redmond, Washington, United States (In-person)

- Mentored and supervised by Dr. Paidamoyo Chapfuwa and Dr. Julia Greissl.
- Worked on extending the AIRIVA deep generative model (https://arxiv.org/abs/2304.13737) to incorporate multiple sources of variation.
- Performed extensive model behavior and sensitivity analyses in an amortized inference setting for T-cell receptor repertoire sequencing data.
- Incorporated various model and training-based changes into production codebase with thorough documentation.

Graduate Research Intern (Immunomics - Antigen Map Project) May-September 2022

Microsoft Research, Redmond, Washington, United States (Virtual Internship)

- Mentored and supervised by Dr. Julia Greissl.
- Worked on incorporating Human Leukocyte Antigen (HLA) information into current production models for detection of disease using T-cell receptor (TCR) repertoire sequencing data.
- Utilized probabilistic frameworks for determining HLA and TCR relationships in an end-to-end manner.
- Incorporated metadata through Graph Convolutional Neural Networks (GCNs) for better population stratification and prediction in confounded contexts.
- Incorporated HLA-aware-factors into production codebase and current production models.

Applied Machine Learning Intern (AI/Engineering Team) September-December 2021

Vector Institute, Toronto, ON

- Led software development of COVID-19 research application to be used by biomedical researchers worldwide.
- Helped develop Graph Convolutional Network (GCN) model for determining variants of concern from COVID-19 genome sequencing data.
- Led team meetings, organized software development sprints, and liaised with collaborators on different research areas.

- Deployed python dash application with user-interactive platform for COVID-19 viral genome sequence analysis.
- Application and GCN work was presented to DNAStack-led collaborative consortium on multi-million dollar SuperCluster Grant, which funded the project.

Machine Learning Intern – Machine Learning and Computational Biology (Wang Lab)
January-September 2020

Techna Institute, Peter Munk Cardiac Center, University Health Network, Toronto, ON

- Led and managed 3 collaborative projects related to single-cell RNA-seq data analysis for cancer and immunology research.
- Developed and performed proof-of-concept analysis for machine-learning based subtyping analysis of cardiovascular disease patients using clinical and genomic data.
- Liaised with various investigators at the University Health Network, Princess Margaret Cancer Center, Sickkids Research Institute, McMaster University, Sunnybrook Health Sciences Center, and The University of Waterloo.
- Analyzed various datasets related to SARS-CoV-2 using pathway, RNA-seq, and viral whole genome sequencing data – part of three peer-reviewed and published manuscripts.
- Developed a state-of-the-art user-interactive platform (CGT) for the analysis of public SARS-CoV-2 viral genome sequencing data, and published a paper accepted in *The Lancet Digital Health*.

Computational Biology Intern – Computational Biology (Reimand Lab) April-December 2019

Ontario Institute for Cancer Research, Toronto, ON

- Drafted a thorough research proposal for the study of enhancers and long non-coding RNAs and their interactions with promoter regions in tumor-tissue.
- Developed an R package *(Loopkit)* for the organization, manipulation, and analysis of chromatin loop data, to be used in subsequent analyses and made public once the project is complete.
- Developed an R and shell pipeline for the integration and analysis of chromatin loop data, long non-coding RNA data, ChIP-seq data and RNA-seq data, to analyze disrupted regulatory interactions between long non-coding RNAs and promoter regions in tumor-tissue.
- Analyzed, summarized, and visualized results of various computational and statistical experiments.
- Presented project at a bioinformatics poster session at the University of Guelph and was awarded a best poster prize.

Clinical Research Assistant- Joint Department of Medical Imaging 2016-2017

Princess Margaret Hospital and Toronto General Hospital, Toronto, ON

- Co-authored research paper "*Correlation of PIRADSv2 with transrectal ultrasound (TRUS)-guided fusion biopsy*" published by the Abdominal Radiology journal.
- Co-authored research paper - *"Evaluation of the Sonographic Features for Thyroid Nodules Biopsy Suggested by the 2015 American Thyroid Association Guidelines"* published by the European Journal of Radiology – Open Access.
- Aided in study conceptualization, methodology, data acquisition/analysis, literature review, investigation, and writing.
- Acted as a liaison between the primary investigators and consulting doctors in the Prostate Center and at Princess Margaret Hospital.

- Assessed inter-reader discrepancies in thyroid biopsies based on the American Thyroid Association guidelines and used rudimentary Excel statistical methods to analyze results for the study.

Clinical Research Assistant- Transplant Infectious Diseases February – September 2016

Toronto General Hospital, Toronto, ON

- Gained knowledge of translational research and application of statistical methods in clinical research through preparing data for analysis in various infectious disease related studies.
- Recruited post-transplant patients and collected samples for study involving invasive aspergillosis and maintained a database of patients enrolled in study to ensure all parties involved in research are aware of changes in patient recruitment.
- Created and implemented data collection standard operating procedure (SOP) for a study involving cytomegalovirus infection in post-transplant patients.
- Liaised between primary investigators, clinical coordinators, and research technicians. Effectively communicated study data-related issues to primary investigators to ensure accuracy of data and a more rigorous collection process.

PUBLICATIONS

*** - indicates equal contribution**

- Cui, H.*, Wang, C.*, **Maan, H.,** Pang, K., Luo, F., Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods*. https://doi.org/10.1038/s41592-024-02201-0

- **Maan, H.**, Zhang, L., Yu, C., Geuenich, M., Campbell, K., Wang, B. (2024). Characterizing the impacts of dataset imbalance on single-cell data integration. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-02097-9

- Cui, H.*, **Maan, H.*,** Taylor, M.D., Wang, B. (2024). DeepVelo: deep learning extends RNA velocity to multi-lineage systems with cell-specific kinetics. *Genome Biology***.** https://doi.org/10.1186/s13059-023-03148-9

- Cui, H., Wang, C., **Maan, H.,** Duan, N., Wang, B. (2022). scFormer: A Universal Representation Learning Approach for Single-Cell Data Using Transformers. *bioRxiv*. https://doi.org/10.1101/2022.11.20.517285

- Emoto, T.*, Lu, J.*, Sivasubramaniyam, T.*, **Maan, H.,** … Robbins, C. (2022). Colony stimulating factor-1 producing endothelial cells and mesenchymal stromal cells maintain monocytes within a perivascular bone marrow niche. *Immunity - Cell Press* 55(5), 862-878. https://doi.org/10.1016/j.immuni.2022.04.005

- Sugiyama, M., Cui, H., Redka, D. S., Karimzadeh, M., Rujas, E., **Maan, H.,** … Antonescu, C. N. (2021). Multiscale interactome analysis coupled with off-target drug predictions reveals drug repurposing candidates for human coronavirus disease. *Nature Sci Rep* 11, 23315. https://doi.org/10.1038/s41598-021-02432-7

- Banerjee, A., El-Sayes, N., Budylowski, P., Jacob, R. A., Richard, D., **Maan, H.**,, … Mossman, K. (2021). Experimental and natural evidence of SARS-CoV-2-infection-induced activation of type I interferon responses. *IScience*, *24*(5), 102477. https://doi.org/10.1016/j.isci.2021.102477

- Nasir, J.A., Kozak, R.A., Aftanas, P., Raphenya, A.R., Smith, K.M., Maguire, F., **Maan, H.,** … Mubareka, S. (2020). A Comparison of Whole Genome Sequencing of SARS-CoV-2 Using Amplicon-Based Sequencing, Random Hexamers, and Bait Capture. *Viruses*. 2020, 12(8), 895. https://doi.org/10.3390/v12080895

- **Maan, H**., Mbareche, H., Raphenya, A. R., Banerjee, A., Nasir, J. A., Kozak, R. A., … Wang, B. (2020). Genotyping SARS-CoV-2 through an interactive web application. *The Lancet Digital Health*. 7500(20), 19–20. https://doi.org/10.1016/s2589-7500(20)30140-0

- Pang, Z., Margolis, M., Menezes, R. J., **Maan, H**., & Ghai, S. (2019). Diagnostic performance of 2015 American Thyroid Association guidelines and inter-observer variability in assigning risk category. *European Journal of Radiology Open*. https://doi.org/10.1016/j.ejro.2019.03.002

- Mathur, S., O'Malley, M. E., Ghai, S., Jhaveri, K., Sreeharsha, B., Margolis, M., Zhong L., **Maan H**., Toi A. (2018). Correlation of 3T multiparametric prostate MRI using prostate imaging reporting and data system (PIRADS) version 2 with biopsy as reference standard. *Abdominal Radiology.* https://doi.org/10.1007/s00261-018-1696-8

FUNDING/AWARDS

Natural Sciences and Engineering Research Council of Canada ($105,000) 2022-2025

Government of Canada

Ontario Graduate Scholarship ($15,000) 2021-2022

Province of Ontario, Canada

COVID-19 Research Grant - Google Cloud ($14,400) April 2020

Google LLC, Mountain View, CA

University of Guelph Master of Bioinformatics - Research Project Stipend ($5000) 2019

University of Guelph, Guelph, ON

CONFERENCE POSTERS AND PRESENTATIONS

Wellcome Connecting Science: Single-cell Biology 2024 - Conference June 2024

Hinxton Hall, South Cambridgeshire, England

- Presented first author published paper titled *Characterizing the impacts of dataset imbalance on single-cell data integration* , which quantifies biologically-relevant effects of dataset imbalance in integration scenarios and introduces guidelines and novel metrics for integration of disparate datasets

Single-cell Genomics 2022 - Conference October 2022

TivoliVredenburg, Utrecht, Netherlands

- Presented first author paper titled *The differential impacts of dataset imbalance in single-cell data integration*, which quantifies biologically-relevant effects of dataset imbalance in integration scenarios and introduces guidelines and novel metrics for integration of disparate datasets

International Conference on Machine Learning (ICML) - Compbio Workshop July 2022
Baltimore Convention Center, Baltimore, Maryland, United States

- Presented co-first author paper titled *DeepVelo: Deep Learning extends RNA velocity to multi-lineage systems with cell-specific kinetics*, which utilizes Graph Neural Networks (GCN) to extend RNA velocity to complex systems typically analyzed through single-cell RNA sequencing

Oxford Machine Learning Summer School August 2021
AI for Global Goals, London, England

- Presented on PhD research - *Modelling complex biological systems with machine learning and integrative single-cell genomics* - in the Unconference Track of the program

Terry Fox Research Institute - Ontario Node (Poster) December 2019
MaRS Collaboration Centre, Toronto, ON

- Submitted abstract, *Disrupted chromatin insulated interactions of long non-coding RNAs and protein-coding genes indicate transcriptional rewiring of oncogenic and tumor-suppressive pathways*, which was accepted for a poster presentation
- Presented poster in a session comprising of innovative experimental and computational cancer research in Ontario

SUPERVISION

Elliott Sicheri (Incoming 4th year Undergraduate - University of Toronto) June 2024-
Vector Institute, Toronto, ON

- Ongoing work related to representation learning of Visium spatial transcriptomics data

Dmytro Kalitin and Roman Ferenets (3rd year Undergraduate and Masters Student - Ukrainian Catholic University) September-December 2022
University of Toronto, Toronto, ON

- Supervised two students for a ‘global classroom’ course between the University of Toronto and two leading Ukrainian universities for Computer Science - the National University of Kyiv-Mohyla Academy (NaUKMA) and the Ukrainian Catholic University (UCU)
- Drafted project outline and provided supervision for project titled *Multi-modal and uncertainty-aware integration of single-cell sequencing data*
- Provided brainstorming sessions, helped troubleshoot computational problems, developed tutorial sessions for computational infrastructure and deep learning

Sarvagya Agrawal (University of Toronto - Statistical Sciences/Molecular Biology Undergraduate Program) 2021-2022
University Health Network, Toronto, ON

- Supervised research intern in the WangLab through the University Health Network for project on cross-species transfer from mouse to human in pediatric cancer research
- Drafted project outline, provided tutorials on single-cell sequencing, deep learning, and pediatric medulloblastoma as primers to engage in the project
- Provided brainstorming sessions, helped troubleshoot computational problems, developed tutorial sessions for computational infrastructure and deep learning
- Project demonstrated poor label transfer from mouse to human in developmental data, with strong implications for cancer and developmental mouse models

PRIZES

International Conference on Machine Learning Computational Biology Workshop - Best Poster Award July 2022

Baltimore Convention Center, Baltimore, Maryland, United States

University of Guelph Bioinformatics Symposium - Best Poster Award August 2019

University of Guelph, Guelph, ON

Ontario Undergraduate Science Case Competition - 2nd Place Overall April 2017

University of Waterloo, Waterloo, ON