Hassaan Maan

hurmaan99@gmail.com

https://github.com/hsmaan

SUMMARY OF QUALIFICATIONS

Molecular Biology/Genetics	Extensive upper year undergraduate course and lab-work involving molecular biology, biotechnology, human molecular genetics, and immunology. Graduate course and project-work on genetics of various pathologies, including molecular mechanisms of cancer.
Mathematics and Statistics	Strong background in mathematics through physics and physical chemistry courses. Solid understanding of statistical methods and machine learning through computational biology positions, coursework, summer program in epidemiology and biostatistics, and clinical data analysis positions.
Programming	Proficient in Python, R, and Bash scripting. Adept at using R and Python for data analysis, developing machine learning models, and visualization. Accustomed to UNIX environments, Linux operating systems, high-performance computing, and cloud computing. Experienced with analyzing NGS file formats, scraping -omics databases, and using bioinformatics tools.
Machine Learning	Solid understanding of machine learning methods used in bioinformatics workflows. Proficient in employing generalized linear regression models, clustering methods, random forests, and neural networks. Experienced in using TensorFlow and PyTorch frameworks in Python, and various statistical modelling tools in R.

RESEARCH INTERESTS

- Utilizing machine learning methods for accurate and robust integration of multi-modal single-cell datasets.
- Developing methods for out-of-distribution prediction for celltype classification in a single-cell RNA-seq setting.
- Identifying mechanisms of epigenetic dysregulation in the onset and progression of pediatric medulloblastoma.

EDUCATION

Doctor of Philosophy Candidate – Department of Medical Biophysics (Ph.D.)   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 Campbell.

- Completed graduate-level coursework in Machine Learning (CS 2506) and Biostatistics (MBP 1201).
- Awarded the Ontario Graduate Scholarship for 2021-2022 ($15,000).

Master of Bioinformatics (M.Binf.) [4.0]   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]   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.
- 3.7 cumulative GPA – graduated with distinction.

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.

RESEARCH EXPERIENCE

Research 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 manuscripts (1 accepted, 2 under review).

- 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*.

Graduate Research Student – 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 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

- 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. *BioRxiv*, 2021. https://doi.org/10.1101/2021.04.13.439274

- 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

COVID-19 Research Grant - Google Cloud ($14400)	April 2020
Google LLC, Mountain View, CA	
Ontario Graduate Scholarship ($15000)	2021-2022
Province of Ontario, Canada	
University of Guelph Master of Bioinformatics - Research Project Stipend ($5000)	2019
University of Guelph, Guelph, ON	

PRIZES

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	

CONFERENCE POSTERS AND PRESENTATIONS

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

COMPLETED PROJECTS

The COVID-19 Genotyping Tool (CGT)   April 2020-Present
Project Page: https://github.com/hsmaan/CovidGenotyper
Techna Institute, Peter Munk Cardiac Center, University Health Network, Toronto, ON

- Developed an interactive R-Shiny web application for the upload and analysis of SARS-CoV-2 viral genomes.
- Employed machine-learning and network analysis methods including Uniform Manifold Approximation and Projection (UMAP) and Minimum Spanning Trees (MST) of viral genome networks, providing outbreak epidemiological information
- Annotated single-nucleotide polymorphisms in structural protein coding regions of the virus, leading to surveillance of important mutations in the COVID-19 pandemic
- Deployed application using Docker container and elastic load balancing framework on the Google Compute Engine (GCE) platform

Dysregulation of Chromatin Interactions between Long Non-Coding RNAs and Promoter Regions in Cancer Cells   April-December 2019

Ontario Institute for Cancer Research, Toronto, ON

- Integrated several large datasets for analysis, including Hi-C chromatin loop data, RNA-seq data from GTEx and TCGA, ChIP-seq data from ENCODE, and long non-coding RNA data from FANTOM5.
- Developed and programmed reproducible analysis pipeline in Bash and R.
- Discovered potential regulatory interactions between long non-coding RNAs and promoters of cancer-related genes, where many interactions were found to be disrupted in tumor-tissue.
- Visualized analysis results using various packages in R, and created poster and slide-deck for presentation purposes, which included illustrating various novel figures in Inkscape.
- Won best poster prize for presentation at a bioinformatics poster session at Guelph University, and presented work during bioinformatics seminar.

LoopRig– **An R package for Organization and Analysis of Chromatin Loop Data**   April-December 2019
Project Page: https://cran.r-project.org/web/packages/LoopRig/index.html
Ontario Institute for Cancer Research, Toronto, ON

- Developed an R package, accepted in the leading R package repository CRAN, for various analyses involving chromatin loop data, such as determining which elements are linked to each other through loops, given certain adjustable parameters.

- Programmed various functions for typical input and output, such as input of genomic interaction data in the form of BEDPE files.
- Programmed functions for manipulating chromatin loop data, such as determining a set of consensus loops across a series of data inputs, based on specific parameters.
- Documented functions and created package using the roxygen2 and devtools packages.

Predicting Lung Cancer in Smokers Using Airway Epithelial Cell Expression Data   January-April 2019
University of Guelph, Guelph, ON

- For the Bioinformatics Pattern Analysis course, created a machine learning and pattern analysis framework for analyzing microarray data from airway cells available through a previous study.
- Developed a comprehensive framework for optimal classification of smokers likely to develop lung cancer by leveraging many modes of feature selection and classification, such as random forest selection and classification, elastic net regularization, logistic regression, and others.
- Improved on classification accuracy from previous study on holdout set (83% to 85%) through elastic net penalized logistic regression, all while using fewer genes in the biomarker (80 vs 21).

Crotalinae (Pit Viper) Venom Composition and Toxicity Correlation   November-December 2019
Project Page: https://github.com/hsmaan/crotalinae-venom
University of Guelph, Guelph, ON

- For the Bioinformatics Software Tools course, completed a thorough exploratory analysis of the venom composition of the Crotalinae subfamily of snakes.
- Completed analysis in R, employing various phylogenetic, genomics, and statistical modelling packages.
- Developed pipeline for analysis and correlation of three different data types: proteomic venom composition, venom toxicity data, and genomic data for the 12S ribosomal RNA marker for phylogenetic reconstruction.
- Created a R Markdown document for reproducible analysis and future research.

Undergraduate Science Case Competition- 2nd Place (Upper Division)   September 2016-March 2017
University of Western Ontario, London, ON

- Worked in team for Ontario-wide research competition and established effective communication, goal-setting, and delegated activities in a fair and effective manner.
- Developed research proposal to combat Zika virus related microcephaly using iminosugar intervention studies in immunodeficient pregnant mice.
- Established statistical methods to analyze the results, such as two-tailed t-tests for comparison among two groups and chi-squared tests to determine association.
- Presented in finalist round to Western University professors and placed second overall in Ontario.