Ginny Xiaohe LI

Computational Biologist

+1 9055809809| ginnyli056@gmail.com| ginny.li@roche.com| https://github.com/ginnyintifa

EXPERIENCE

Data and Analytics Chapter, Hoffmann-La Roche Limited **ON, Canada**
Computational Biologist 05/2023 – *present*

Colorectal Cancer (CRC) Metastasis Dissection – Multi-Omic Integration

- Led multi-omic data analysis integrating RNA-seq, ATAC-seq, and scRNA-seq from preclinical models, and engineered a suite of tools to streamline interpretation and insights generation.
- Collaborated with Research Oncology wet-lab scientists to decipher complex signaling dynamics driving CRC metastasis and validated therapeutic potential of key gene inhibitors.
- Co-authored original manuscript with custom data visualizations for publication in *Nature Cancer*.

Non-Invasive Biomarker Development in ER+ Breast Cancer – 5hmC ctDNA Epigenetics

- Pioneered the organization's first 5hmC processing pipeline, establishing a reusable computational asset for non-invasive epigenetic profiling.
- Validated ctDNA as a high-fidelity proxy for solid tumor RNA-seq, enabling longitudinal disease monitoring without the need for invasive biopsies.
- Modelled novel gene signatures for estrogen receptor activity and prognosis, providing the foundational framework for future biomarker development.

Immune Annotation in Multiple Myeloma Patients – Single-Cell Transcriptomics

- Validated differential abundance signatures using multi-method statistical testing, confirming primary therapeutic targets.
- Designed a robust cell-type annotation framework to resolve fine-grained immune cell heterogeneity within the cellular landscape of a novel patient cohort.

Cellular Repair characterization in Liver Injury models – Single-Cell Multi-Omics

- Mapped hepatocyte zonation and injury-responsive states using scRNA-seq and scATAC-seq, identifying key epigenetic regulators.
- Evaluated doublet detection algorithm and trajectory inference methods in transitional cell states to ensure high-resolution cell-state mapping.

Department of Pathology, University of Michigan **MI, USA**
Research Fellow 07/2021 – *05/2023*

Non-Clear Cell Renal Cell Carcinoma (non-ccRCC) Characterization– Proteogenomics

- Led comprehensive multi-omic integration of genomic, transcriptomic, proteomic, and metabolomic data to reveal novel rare kidney tumor biology in CPTAC patient cohorts.
- Collaborated with pathologists in research hypothesis generation towards comprehensive molecular characterization of heterogenous non-ccRCC subtypes.
- Supervised and conducted data analysis leading to discovery of diagnostic markers distinguishing RCC subtypes that are histologically similar and novel prognostic markers for rare RCC subtypes.
- Authored manuscript and generated data visualization for publication in *Cell Reports Medicine*.

You3: A Leadership and Management Program for Postdocs

- Completed the 8-week structured program developing project and team leadership and management skills.

Yong Loo Lin School of Medicine, National University of Singapore **Singapore**
Research Fellow 01/2020 – 06/2021

Insulin-stimulated Human Myotubes Signal Modelling – Phosphoproteomics

- Built an empirical Bayesian statistical framework to detect significant abundance changes in co-varying pairs of kinases and substrates based on phosphoproteomic profiling.
- Identified uniquely activated signaling transduction cascades in insulin-stimulated myotubes.
- Developed an R package *KSA2D* to achieve the analysis and draft the full manuscript.

Staurosporine kinase inhibition Impact Delineation – Phosphoproteomics

- Collaborated with molecular biologists from Institute of Molecular and Cellular Biology in research hypothesis generation towards delineating multiple response patterns.
- Designed statistical analysis and developed an R package and Shiny App *KOPI* for protein changes annotation and visualization under various experimental settings.
- Drafted and revised a manuscript with collaborators resulting in publication in *Scientific Reports*.

Saw Swee Hock School of Public Health, National University of Singapore **Singapore**
Research Assistant 08/2018 - 08/2019

Multi-omic marker discovery in thyroid and gastric cancer patients

- Aggregated genome variants to protein units and compare the results with respective TCGA cohorts.
- Implemented differential expression analysis with transcriptome data and compare with TCGA cohorts for biomarker discovery.

Teaching Assistant 01/2015 - 05/2017

- Facilitated “Public Health in Action” modules hosting more than 200 undergraduates for 3 semesters
- Tutored graduate students with R programing in the module “Quantitative Epidemiology Methods” and “Advanced Biostatistics”.

EDUCATION

Saw Swee Hock School of Public Health, National University of Singapore **Singapore**
Doctor of Philosophy 08/2014 - 07/2019

- Thesis title: Cancer genome biomarker discovery with integrated molecular data and system biology-oriented approaches.

School of Mathematics, Shandong University **Jinan, China**
Bachelor of Science 09/2010 - 06/2014

- Mathematical training in calculus, linear algebra, numerical analysis, probability, and statistics.

DOCTORAL PROJECTS

Prognostic exome variants discovery in protein-centric approach 01/2018 - 12/2019

- Mapped exome variants across 33 cancer types in TCGA data to domains and PTM sites on proteins.
- Constructed a survival analysis model to detect prognostic protein units, resulting in an R package *GPD*.
- Drafted and revised a manuscript resulting in publication in *Human Mutation*.

Proteome-wide post-translational modification prediction 01/2015 - 05/2018

- Collected a total of 173 sequence and structural features to predict generic PTM sites.
- Developed an R package *PTMscape* incorporating *Liblinear* as the classifier.
- Drafted and revised a manuscript resulting in publication in *Molecular Omics*.

PROFESSIONAL SKILLS

Computational Biology

- **NGS based genomics, epigenomics, transcriptomics pipeline:** experienced in conducting DNA sequence alignment with Bowtie2, post-processing with SAMtools, RNA-seq pseudo-alignment and quantification with Kallisto, peak calling with MACS2.
- **MS based proteomics pipeline:** experienced in performing proteomics identification and quantitation with FragPipe.
- **Downstream data analysis:** proficient in handling high dimensional -omics data such as RNA-seq (bulk and single cell level), ATAC-seq (bulk and single cell level), proteomics, phosphoproteomics and metabolomics data, conducting differential expression, functional enrichment, network analysis with statistical modelling and machine learning algorithms.
- **Large datasets exposure**: familiar with TCGA, COSMIC, CPTAC and many other public data resources.

Programming and Data Visualization

- **R**: proficient in scientific programming including data wrangling and statistical modelling; Proficient with *dplyr*, *ggplot2*, *limma*, *biomaRt*, *Seurat* and many other packages on Bioconductor; Proficient in writing reproducible reports with R Markdown; Proficient in R package and Shiny App development.
- **Conda/Python**: proficient in conducting single cell analysis using Scanpy; Proficient in conducting machine learning in Python.
- **Cytoscape and Adobe Illustrator**: proficient in publication-quality data visualization.
- **Other**: experienced in working on HPC clusters and collaborating through Git.

Interdisciplinary Collaboration

- Experienced in cross-functional collaboration with wet-lab scientists, pathologists, and clinicians, translating complex computational findings into actionable biological insights.

PUBLICATIONS

- Heinlein M**, Li GX,**…Piskol R, De Sauvage, F. J. A high MAPK cell state drives metastatic dissemination in colorectal cancer. *Accepted in Nature Cancer.*
- **Li GX***, Chen L*, Hsiao Y*, Nesvizhskii, A.I. CPTAC consortium. Comprehensive proteogenomics characterization of rare kidney tumors. *Cell Reports Medicine. 5.5 (2024).*
- Hsiao Y, Zhang H, **Li GX,** …, Nesvizhskii, A.I. Analysis and visualization of quantitative proteomics data using FragPipe-Analyst. *Journal of Proteome Research. 2024 4303-4315.*
- **Li GX***, Zhao T*, Choi H, Lim YT*, Sobota R. KOPI: Kinase inhibitOr Proteome Impact analysis. *Scientific Reports. 2022 12(1),1-8*.
- Ko S, **Li GX**, Choi H, Won JH. Computationally scalable regression modeling for large-scale clinical omics data with ParProx. *Briefings in bioinformatics. 2021 22 (6).*
- Han B***, Li GX***, Chan E*, Liew W, Huang S, Liu MH, Choi H, Kim KP, Tai ES. Proteome and phosphoproteome dynamics in insulin-stimulated myotubes of a multi-ethnic Asian cohort. *In preparation.*
- **Li GX**, Munro D, Fermin D, Vogel C, Choi H. A protein-centric approach for exome variant aggregation enables sensitive association with clinical outcomes. *Human Mutation. 2020 41(5): 934-945*.
- **Li GX**, Vogel C, Choi H. PTMscape: an open source tool to predict generic post-translational modifications and map modification crosstalk in protein domains and biological processes. *Molecular Omics. 2018 14.(3),* 197-209.

SOFTWARE PORTFOLIO

- **PTMscape:** proteome-scale PTM prediction using SVM with a comprehensive set of features. https://github.com/ginnyintifa/PTMscape
- **GPD:** protein-centric exome variants aggregation for cancer biomarker discovery. https://github.com/ginnyintifa/GPD
- **KSA2D**: differential analysis for discovery of co-varied kinase-substrate pairs. https://github.com/ginnyintifa/KSA2D
- **KOPI**: tracking and visualizing protein abundance changes in various experiment settings. https://ginnyintifa.shinyapps.io/ProteomeNodesShiny/

CONFERENCE PRESENTATIONS

- **CPTAC Annual Consortium Meeting 2023** **Bethesda**
  Comprehensive proteogenomic characterization of rare kidney tumors (talk). 03/2023
- **US Human Proteome Organization Conference 2022** **Charleston**
  Assessment of discordance between peptide and protein level differential expression analysis (poster). 03/2022
- **CPTAC Annual Consortium Meeting 2021** **Online**
  Proteogenomics characterization of non-Clear Cell Renal Cell Carcinoma (poster). 10/2021
- **8th Annual Biomedical Scientific Congress** **Singapore**
  GPD: a segmentation-based approach for exome association analysis (talk). 10/2019
- **25th Intelligent Systems for Molecular Biology Conference** **Prague**
  Protein-centric Exome Association (PREXA) analysis for prognostic coding variants discovery (poster).07/2017
- **15th Human Proteome Organization World Congress** **Taipei**
  PTMtopographer: an interface for proteome-scale PTM prediction (poster). 09/2016

AWARDS

- **Best Visualization Award**, Yale-NUS Data1.0 Hackathon. 05/2017
- **Outstanding Graduate of Class 2014**, Shandong University. 06/2014
- **First Class Scholarship**, Shandong University. 06/2013