

Ginny Xiaohe LI

Computational Biologist

+1 773-383-2781 | ginnyli056@gmail.com | <https://github.com/ginnyintifa>

EXPERIENCE

Department of Pathology, University of Michigan
Research Fellow

MI, USA
07/2021 – present

Lead CPTAC non-Clear Cell Renal Cell Carcinoma (non-ccRCC) proteogenomics analysis

- Collaborate with pathologists to generate research hypothesis for comprehensive molecular characterization of heterogeneous non-ccRCC.
- Design data analysis workflow integrating genomic, transcriptomic, proteomic and metabolomic data.
- Supervise and conduct data analysis to discover diagnostic markers that distinguish RCC subtypes that are histologically similar and novel prognostic markers for rare RCC subtypes.
- Write manuscript and generate data visualization from scratch for high-impact scientific publication.

Conduct statistical analysis in multiple collaborative projects

- Conduct transcriptomic and proteomic data normalization, imputation and batch effect correction in lung cancer and acute myeloid leukemia.
- Evaluate downstream impact of tumor suppressor gene bi-allelic loss in large lung cancer cohort.
- Build statistical models on multi-omic datasets to test hypotheses of biological relevance.
- Refine algorithms for allele-level quantification of polymorphic HLA proteins.
- Test, benchmark and improve open-source software automating downstream proteomics analysis.

Complete You3: A Leadership and Management Program for Postdocs

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

Yong Loo Lin School of Medicine, National University of Singapore
Research Fellow

Singapore
01/2020 – 06/2021

Model phosphorylation signaling in insulin-stimulated human myotubes

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

Annotate and visualize proteome-wide impact caused by kinase inhibitor Staurosporine

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

Saw Swee Hock School of Public Health, National University of Singapore
Research Assistant

Singapore
08/2018 - 08/2019

Multi-omic marker discovery in thyroid and gastric cancer patients

- Aggregate genome variants to protein units and compare the results with respective TCGA cohorts.
- Implement differential expression analysis with transcriptome data and compare with TCGA cohorts.

Teaching Assistant

01/2015 - 05/2017

- Facilitate “Public Health in Action” modules hosting more than 200 undergraduates for 3 semesters.
- Tutor graduate students with R programming 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 RESEARCH

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

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

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

- Collect a total of 173 sequence and structural features to predict generic PTM sites.
- Develop an R package *PTMscape* incorporating *Liblinear* as classifier.
- Draft and revise a manuscript resulting in publication in *Molecular Omics*.

PROFESSIONAL SKILLS

Computational Biology

- **Next generation sequencing (NGS) based genomics, transcriptomics pipeline:** experienced in conducting DNA sequence alignment with Bowtie2, post-processing with SAMtools, RNA-seq pseudo-alignment and quantification with Kallisto, single cell RNA-seq quantification with Kallisto bustools.
- **Mass spectrometry (MS) based proteomics pipeline:** experienced in performing proteomics identification and quantitation with FragPipe.
- **Statistical analysis and machine learning:** proficient in handling high dimensional -omics data such as genomic variants, DNA methylation, RNA-seq (bulk and single cell), proteomics, phosphoproteomics and metabolomics data, performing data preprocessing including normalization, batch effect correction and imputation;
Proficient in conducting differential expression, functional enrichment, network analysis and other advanced analysis with statistical modelling and machine learning algorithms.
- **Large datasets exposure:** familiar with TCGA, COSMIC, CPTAC and other public data resources.

Programming and Data Visualization

- **R:** proficient in scientific programming including and statistical modelling and machine learning; Proficient with *dplyr*, *ggplot2*, *limma*, *biomaRt*, and many other packages on CRAN/Bioconductor; Proficient in writing reproducible reports with R Markdown; Proficient in R package and Shiny App development.
- **Python:** experienced in data analysis with *numpy* and *pandas*; Proficient in calling transcriptomics tools in conda environments.
- **Cloud Computing:** experienced in working on Linux servers with command-line interfaces and collaborating through Git.
- **Cytoscape and Adobe Illustrator:** proficient in publication-quality data visualization.

PUBLICATIONS

- **Li GX***, Lijun Chen*, Hsiao Yi*, Saravana Mohan Dhanasekaran, Nesvizhskii A, CPTAC consortium. Comprehensive proteogenomics characterization of rare kidney tumors. *Submitted*.
- Mumphrey, M. B., **Li GX.**, Hosseini N., Nesvizhskii A., & Cieslik, M. HLAProphet: Personalized allele-level quantification of the HLA proteins. *bioRxiv*, 2023-01.
- **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 2022 | Online |
| Proteogenomics characterization of Pan-Renal Cell Carcinoma (talk). | 10/2022 |
| • US Human Proteome Organization Conference 2022 | Charleston |
| Assessing 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 |