

## EDUCATION

---

- **Weill Cornell Graduate School of Medical Sciences** New York, NY  
*Ph.D Candidate in Computational Biology and Medicine* *Jul. 2023 – Present*
- **Northeastern University** Boston, MA  
*B.S. in Computer Science and Mathematics* *Sep. 2016 – May. 2021*  
*Magna Cum Laude*

## RESEARCH EXPERIENCES

---

- **Ph.D Student, Memorial Sloan Kettering Cancer Center** New York, NY  
*Advisor: Caleb Lareau* *Dec 2023 - Present*
  - Computationally designed therapeutically applicable de novo proteins via generative approaches, including diffusion-based (RFDiffusion-ProteinMPNN-AlphaFold2 / BoltzGen) and model inversion-based (BindCraft/Germinal) approaches for protein binder design, with experimental validation in applications including CAR-T therapy, gene editing, radiotherapy and others.
  - Performed large-scale clinical genomic analysis on tumor sequencing data (>100k) such as MSK-IMPACT for biomarker discovery, using tools such as AlphaMissense and AlphaGenome to improve variant interpretation and empower statistical enrichment.
  - Actively participated in experimental design with experimentalist to identify areas where data collection will improve model performance and inform future experiments. Engaged in discussions with clinicians to identify areas of unmet needs and conceptualize new projects leveraging existing resources to maximize clinical impact.
- **Computational Biologist, Dana-Farber Cancer Institute** Boston, MA  
*Advisors: Saud AlDubayan, Eliezer Van Allen* *Jun 2021 - Present*
  - Independently led multiple projects over a range of topics in cancer genomics, including uncovering the germline genetic architecture of breast cancer in understudied populations, understanding the influence of germline elements on disease presentation, identifying predisposition genes in rare cancers, associating tumor's genetic characteristic with immunotherapy response and other clinical outcomes.
  - Developed more than 50 workflows on cloud computing platforms (Terra, Google Cloud) for common genomic tasks including sequence alignment, variant calling (short and structural variants), VCF manipulation, genotype phasing and imputation, genome and transcriptome wide association analysis, polygenic risk score calculation, statistical analysis and data mining. The workflows are extensively documented and are shared amongst team members.
- **Researcher Assistant, Northeastern University** Boston, MA  
*Advisor: Predrag Radivojac* *Mar 2020 - July 2023*
  - Identified population subgroups at high-risk of developing adverse pregnancy outcomes using association rules mining. Discovered non-linear interactions between exercise, genetic predisposition as quantified by polygenic risk scores, and risk for developing gestational diabetes. In addition, racial-ethnic disparities in high-risk subgroups were quantified using Gini coefficient and potential effects of policy interventions were investigated.
  - Constructed protein networks from multiple publicly available databases and utilized graph mining tools to discover protein complexes. Evaluated the performance of various machine learning models such as distance metric, SVM, neural network ensembles on protein-protein interaction prediction using measures such as area under ROC curve, precision-recall curve, and obtained confidence intervals via bootstrapping.
- **Research Experience for Undergraduate, Northeastern University** Boston, MA  
*Advisor: Jonier Autunes* *May 2019 - June 2019*
  - Modeled biological networks with a death-birth update rule using graph libraries and reduced simulation run-time by 40% by implementing a custom data structure and optimizing the existing algorithm

## WORK EXPERIENCES

---

- **National Institute of Allergy and Infectious Diseases** (Remote) Bethesda, MD  
*Data Scientists Fellow* Oct 2020 - Dec 2020
  - Developed web scraping, data storage, data cleaning and visualization pipeline for the data usage of dbGaP. Maintained project repository with standard open-source practices such as version control, documentation, issue tracking and continuous integration
- **Chewy** Boston, MA  
*Software Engineer Co-op* July 2019 - Dec 2019
  - Developed new features on company product display pages using Vue components and maintained the legacy Stripes framework using JSP fragments, Java and Kotlin. Analyzed 2019 e-commerce conversion rate on Google Analytics by retrieving session data using SQL queries, segmenting sessions based on attributes such as device type and traffic channel, and forecasting conversion rate using linear regression
- **BlockTEST** Cambridge, MA  
*Front-end Engineer Co-op* July 2018 - Nov 2018
  - Designed and implemented 20+ webpages with HTML, CSS, Javascript and dynamically populated Bootstrap templates by making REST calls. Organized healthcare themed hackathon with emphasis on blockchain technology and hosted webinars.

## PUBLICATIONS

---

### (CO-) FIRST AUTHOR (\* - Equal Contribution)

1. Arthur Chow\*, **Hoyin Chu\***, *et al.* (2025) Sequence and structural determinants of efficacious de novo chimeric antigen receptors , *Under Review*
2. **Hoyin Chu**, *et al.* (2025) Patient-Partnered Multiomics Reveals the Molecular Architecture of Angiosarcoma, *Under Review*
3. Mohammed Al-Jumaan\*, **Hoyin Chu\***, *et al.* (2023) Interplay of Mendelian and Polygenic Risk Factors in Arab Breast Cancer Patients, *Genome Medicine* <https://doi.org/10.1186/s13073-023-01220-4>
4. Abdullah Alsulaiman\*, **Hoyin Chu\***, *et al.* (2023) Profiling of pharmacogenomic variants in CYP2D6 and DPYD in indigenous Arab breast cancer patients, *Pharmacogenomics*, <https://doi.org/10.2217/pgs-2023-0029>
5. **Hoyin Chu**, *et al.* (2023) Using Association Rules to Understand the Risk of Adverse Pregnancy Outcomes in a Diverse Population, **Pacific Symposium on Biocomputing**, PMID: 36540978
6. Kymberleigh A. Pagel\*, **Hoyin Chu\***, *et al.*, (2022) Association of Genetic Predisposition and Physical Activity With Risk of Gestational Diabetes in Nulliparous Women, *JAMA Network Open*, doi:10.1001/jamanetworkopen.2022.29158

### CO-AUTHOR:

7. Sherry S. Nyeo, ..., **Hoyin Chu**, *et al.* (2025) Population-scale sequencing resolves correlates and determinants of latent Epstein-Barr Virus infection, *Nature (in press)*, <https://doi.org/10.1101/2025.07.18.665549>
8. Maria Caterina Rotiroti, Aidan M. tousley, **Hoyin Chu**, *et al.* (2025) Engineering T cells with a membrane-tethered version of SLP-76 overcomes antigen-low resistance to CAR T cell therapy, *Nature Cancer*, <https://doi.org/10.1038/s43018-025-01056-4>
9. Tudor-Stefan Cotet, ..., **Hoyin Chu** *et al.* (2025) Crowdsourced Protein Design: Lessons From the Adaptyv EGFR Binder Competition, *Under Review*, <https://doi.org/10.1101/2025.04.17.648362>

10. Mark Leick, ..., **Hoyin Chu**, *et al.*, (2025) Genomics correlates of clinical CAR-T cell activity, *Under Review*, <https://doi.org/10.1101/2025.10.08.25337584>
11. Romualdo Barroso-Sousa, ..., **Hoyin Chu**, *et al.* (2025) Nivolumab plus low-dose ipilimumab in hypermutated HER2-negative metastatic breast cancer: a phase II trial (NIMBUS), *Nature Communications*, <https://doi.org/10.1038/s41467-025-59695-1>
12. Alexandre Pellan Cheng, ..., **Hoyin Chu**, *et al.* (2025) Error-corrected flow-based sequencing at whole-genome scale and its application to circulating cell-free DNA profiling, *Nature Methods*, <https://doi.org/10.1038/s41592-025-02648-9>
13. Seung Hun Han, Sabrina Y. Camp, **Hoyin Chu**, *et al.* (2024) Integrative Analysis of Germline Rare Variants in Clear and Non-clear Cell Renal Cell Carcinoma, *European Urology Open Science*, <https://doi.org/10.1016/j.euros.2024.02.006>
14. **The Critical Assessment of Genome Interpretation Consortium**, (2024) CAGI, the Critical Assessment of Genome Interpretation, establishes progress and prospects for computational genetic variant interpretation methods, *Genome Biology*, <https://doi.org/10.1186/s13059-023-03113-6>
15. Riaz Gillani, ..., **Hoyin Chu**, *et al.* (2022) Germline predisposition to pediatric Ewing sarcoma is characterized by inherited pathogenic variants in DNA damage repair genes, *The American Journal of Human Genetics*, <https://doi.org/10.1016/j.ajhg.2022.04.007>

#### ABSTRACTS & PRESENTATIONS

---

1. **Hoyin Chu**, *et al.* (2025) Generative AI enables De novo Protein Design for CAR-T Cells, *Protein Science Symposium*
2. **Hoyin Chu**, *et al.* (2023) Patient-partnered Research Enables Germline Characterization of Angiosarcoma Predisposition Genes, *AACR*
3. **Hoyin Chu**, *et al.* (2022) Multi-modal characterization of ultra-rare germline genetic variants driving breast cancer risk in the indigenous Arab population, *ASHG*
4. Mark Leick, ..., **Hoyin Chu**, *et al.* (2022) Patients with Deleterious Germline Variants in STXBP2 Develop Toxicity after CAR-T Cell Therapy with Axicabtagene Ciloleucel, *ASH*
5. Kiyomi Mashima, ..., **Hoyin Chu**, *et al.* (2022) Characteristics of rare germline ATM variants in chronic lymphocytic leukemia *ASH*

#### COMPETITION PARTICIPATION

---

|  |      |
|--|------|
| <b>Expert Panel Selected Entry</b> - AdaptyvBio Nipah Virus Binder Design Challenge                  | 2025 |
| <b>Best FGFR Loss-of-function Predictor</b> - Critical Assessment of Genome Interpretation 7 (CAGI7) | 2025 |
| <b>Top de novo binder designers</b> - AdaptyvBio EGFR Binder Design Challenge                        | 2024 |

#### HONORS AND AWARDS

---

|  |           |
|--|-----------|
| <b>Dean's List</b> , Khoury College of Computer and Information Sciences | 2016-2021 |
| <b>Civic Digital Fellowship</b> , Coding it Forward                      | 2020      |
| <b>Amelia Peabody Scholarship</b> , Northeastern University              | 2017      |

#### MISCELLANEOUS

---

|                     |           |
|---------------------|-----------|
| <b>Sub-reviewer</b> |           |
| ASM-BCB,            | 2022-2025 |
| ISMB/ECCB           | 2021-2022 |