xiwang@hms.harvard.edu
(858) 203-8221

Xi (Sam) Wang

EDUCATION

Harvard Medical School
Boston, MA

S.M. in Biomedical Informatics, GPA: 4.0/4.0
Aug 2022 – May 2024

- **Concentrations**: AI/Machine-Learning and Cancer Genomics
- **Coursework**: AI/Machine-Learning in Healthcare, Clinical Informatics, Single-cell Analysis, Statistics, Deep Learning for Biomedical Data Science, Animal Ethics, Launching Technology Ventures, Precision Medicine.

University of California – San Diego
La Jolla, CA

B.S. in Bioengineering: Bioinformatics, GPA: 3.84/4.0 (CS GPA: 3.88/4.0), Cum Laude
Sep 2018 – June 2022

- **Concentrations**: Computer Science and Cancer Genomics
- **Coursework**: Bioinformatics Techniques, Organic Chemistry, Molecular Biology, Vector Calculus, Advance Computer Science Engineering (R, Python, Java, C, C++), Differential Equations, Biology Lab.

PUBLICATION

- Marcos Díaz-Gay, Raviteja Vangara, Mark Barnes, **Xi Wang**, S M Ashiqul Islam, Ian Vermes, Stephen Duke, Nithish Bharadhwaj Narasimman, Ting Yang, Zichen Jiang, Sarah Moody, Sergey Senkin, Paul Brennan, Michael R Stratton, Ludmil B Alexandrov, Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment, Bioinformatics, Volume 39, Issue 12, December 2023, btad756, https://doi.org/10.1093/bioinformatics/btad756.

RESEARCH EXPERIENCE

Cristea Lab, Hale Family Center for Pancreatic Cancer Research
Boston, MA

Compuational Biologist
July 2024 – Current

- Independent project on developing novel AI/ML transformer decoder-only large foundation language models for inferring and generating cancer evolution in single-cell level using scRNAseq data under Dr. Simona Cristea, PhD.
  - Data aggregation using publicly available scRNAseq datasets (from CellXGene and 3CA databases) of breast cancer in normal, pre-cancer, primary, and metastatic stages.
  - Pseudo cell evolution progression generations from k-nearest neighbors.
  - Constructing and pre-training custom transformer decoder-only foundation models for generation of cancer cell evolution progression. Given the expression values of a cell in normal stage, the pre-trained decoder will generate and predict expression values or biomarkers of the same cell in pre-cancer, primary, and metastatic stages in a zero-shot manner.
  - Pre-trained model will also be capable of further fine-tuning.
  - Fine-tuning Llama3.1 for the same tasks as performance benchmarks.

Wolpin-Nowak Lab, Dana-Farber Cancer Institute
Boston, MA

Computational Biologist
July 2024 – Current

- Collaborating with Dr. Brian Wolpin, MD, MPH, Dr. Jonathan Nowak, MD, PhD, and Dr. Andrew Aguirre, MD, PhD.
- Break Through Cancer: Collaboration project on biomarkers and tumor micro-environments analysis of pancreatic cancer using spatially resolved single-cell RNA sequencing data.
  - One of the largest spatially-resolved single-cell RNA sequences study on pancreatic cancer with over 130 samples sequenced under 5 different techniques (CosMx, GeoMx, Visium, Xenium, and COMET) that spanned 3 top cancer institutes (Dana-Farber Cancer Institute, MD Anderson Cancer Center, and Memorial Sloan Kettering Cancer Center)
  - Leading major computational analysis and constructing standardized pipelines for cell type annotations and tumor subtyping analysis, such as Basal vs. Classical for tumor cells, and other more granular subtypes. Validating results with spatial information of each cells.

Liu Lab, Dana-Farber Cancer Institute
Boston, MA

Graduate Student Researcher
Nov 2022 – May 2024

- Independent project on developing novel deep neural networks for inferring melanoma tumor ploidy, purity, cancer cell fractions and heterogeneity from tumor-only targeted panel sequencing data under Dr. David Liu, MD, MPH, MS.
  * Data cleaning and sub-sampling of TCGA Melanoma whole exome-sequencing into targeted panel sequences.
  * Variant Allele Frequency (VAF) + mutational status of targeted genes to predict tumor purity.
  * Read depth + mutational status of targeted genes to predict tumor ploidy.
  * Statistical equations for Cancer Cell Fraction (CCF) and heterogeneity inferences.

Alexandrov Lab, University of California, San Diego
La Jolla, CA
Undergraduate Bioinformatics Researcher
May 2021 – Aug 2022
- Independent project Comprehensive Benchmarking of Mutational Signature Assignment Tools under Dr. Ludmil B. Alexandrov, PhD.
  * Installation and execution of 20+ different mutational signature assignment tools based on different programming languages and mathematical algorithms (extensive use of Conda environments and Git version control).
  * Data analysis and benchmarking of the tools considered using synthetic datasets and various qualitative and quantitative metrics (extensive experience in AWS).
  * Profiling, upgrade, and testing of SigProfilerSingleSample, the assignment tool developed by the Alexandrov lab.

Masuda Lab, University of California, San Diego
La Jolla, CA
Staff Research Associate
May 2022 – Aug 2022
- Project on intervertebral disc restoration using the rabbit model with Dr. Koichi Masuda, MD.
  * Perform animal pre-operational, surgical, and post-operational procedures on over 100 rabbits).
  * Execute 2D x-rays analysis and 3D micro-CT post-harvesting analysis.

Cartilage Tissue Engineering Lab, University of California, San Diego
La Jolla, CA
Undergraduate Researcher
Apr 2019 – Aug 2022
- Senior Design project on Surgical Restoration of the Anterior Cruciate Ligament and Knee Biomechanics.
  * Improving the surgical techniques used for ACL reconstruction surgery using rabbit model.
  * Testing different cutting equipment, ligaments and bones grabbing methods, and engineering special tools for improvements.
- Quantification of Marine Animal Teeth Enamel Structure with micro-CT.
  * Finished an industrial project on quantification of marine animal teeth structure.
  * 2D analysis through sagittal and transaxial multi-slices using fitted-curve calculated in MATLAB.
  * 3D analysis through thresholds using gray-scale histogram. Utilized 3D models for 3D volume analysis.

PROJECTS

- **RxPlorer, Harvard University**: RxPlorer provides specialized pre-trained Large Language Model that performs meta-analysis of current research in the biomedical and clinical field with information that is dynamically updated with sanity and citation checks. Fine-tuned Llama2 (by Meta) with QLoRA, utilized Chat GPT3.5 (OpenAI), implemented quantization of LLM, created web-frontend and server-backend with monitoring system through WhyLab.
- **Yogurt World (product: yogurtworlducsd.com)**: Using HTML, design a professional website for UCSD students' favorite yogurt place on campus. Design the whole website using Figma and then implement the site with HTML on GitHub.
- **Android App: Zoo Seeker**: Designed an Android Java-based app that allows users to search animal exhibits and show the directions to exhibits with the shortest combined distances, using the Traveling Sales Man algorithm. Intergraded auto-correction and voice-recognition for users to search for animals.
- **IOS App: UCSD Dining Hall Review (https://github.com/xiw020-Sam/UCSDDiningHallReview)**: Designed an IOS Swift-based app that lets users to log-in, post reviews, and comment on reviews for various UCSD dining halls.

SKILLS

AI/ML: Experienced with various ML architectures, custom Neural-Network/Transformer constructions and pre-training with Torch and/or Tensorflow, LLM fine-tuning, and deep neural network fine-tuning.

Programming Languages: Advance knowledge in data structure in R, Java, C, C++, Python, Bash.

Programming Technologies: Git, Node.js, React, MySQL, MongoDB, Flask, AWS, Docker, Google Cloud Service, Linux, HTML/CSS

Bioinformatics Techniques: Experienced with bioinformatics pipeline development and analysis using whole-genome, whole-exome, and targeted panel sequences. Expeirneced in single-cell and spatial transcriptomics analysis.

Languages: Native proficiency in English and Mandarin Chinese, conversational proficiency in Cantonese.