HARRISON WANG
R&D DATA SCIENTIST

Enthusiastic data scientist with a solid work ethic. Dedicated to upgrading my skills. Committed to making an impact on healthcare through biotech and research.

✉ harrison.c.wang@gmail.com
🌐 https://harrisonized.github.io
in https://www.linkedin.com/in/harrisonized/
⌂ https://www.github.com/harrisonized

PROGRAMMING: Python, SQL, Datalog, Matlab
DATABASES & WEB FRAMEWORKS: Postgresql, Datomic, MongoDB, Flask, Django
CLOUD COMPUTING & DASHBOARDING: AWS, EC2, S3, Sagemaker, Redshift, Athena, Tableau, Power BI
TOOLS & LIBRARIES: Pandas, Numpy, Matplotlib, Plotly, Sklearn, Statsmodels, BeautifulSoup, Selenium
MACHINE LEARNING: Regression, Classification, Clustering, Time Series Analysis, Natural Language Processing
WET LAB: Next-Generation Sequencing, CRISPR/Cas9, Assay Development, ELISA, qPCR, Flow Cytometry, Cloning, Site-Directed Mutagenesis, Tissue Culture, Protein Expression, Protein Purification, Western Blot, Confocal Microscopy, Biolayer Interferometry

Experience

THERMO-FISHER SCIENTIFIC
South San Francisco, CA
Research and Development Data Scientist
Feb. 2021 to Current
- Built a **scalable, end-to-end ETL process** to achieve **real-time monitoring** of **SARS-CoV-2 RT-PCR** test data from labs and universities using **Amazon AWS** and **Power BI**. Listed as co-author on a **symposium abstract** based on this work
- Led the development of the **data pipeline**. Heavily **refactored** and **optimized** the existing data pipeline for **scalability** through clever use of **parallelization** and **batch processing**, reducing the runtime by 40x. Added new KPIs and support for backward compatibility
- **Analyzed** performance of sample retests using **Python** and **SQL**, enabling stakeholders to understand the limit-of-detection of our **SARS-CoV-2 RT-PCR** tests
- Created new **Plotly interactive visualizations** for 96- and 384-well plates, enabling the Data Science team to rapidly **detect anomalies**

INVITAE
San Francisco, CA
Bioinformatics Data Scientist
Jan. 2020 to Sept. 2020
- Discovered the **root cause** for a high-visibility. time-sensitive issue on a new assay that resulted in batch failures impacting 2000+ patients
- Analyzed **trends** using **Python** and **SQL** to monitor **next-generation sequencing (NGS)** production-line data for anomalies
- Efficiently diagnosed **customer issues** by writing a **Python script** to automate generation of standardized **visualizations** and **statistics**
- Wrote maintainable **Datalog** and **SQL** queries for **Flask** apps used to migrate data from **Datomic** to **Amazon Redshift** for consumption by **Tableau**
- Built an 8-page **Tableau dashboard** to enable our operations team to carefully monitor a new product launch

NEXTBEE MEDIA
San Mateo, CA
Data Scientist
Sept. 2019 to Dec. 2019
Led development of the Lighthouse App from inception to deployment. Available at: https://lighthouse.nextbee.com/
- Designed the **relational database model** and wrote the **Python scripts** used to build customer profiles based on orders data
- **Segmented** customers into **tier groups** based on features identified through **ecommerce domain knowledge**
- Used **logistic regression** and **random forest** to predict the likelihood of each customer making another purchase based on their purchase history
- Used **SARIMA time-series forecasting** to predict future revenue and number of new customers

BIOVERATIV, FORMERLY TRUE NORTH THERAPEUTICS
South San Francisco, CA
Research Associate 2
Jan. 2017 to Mar. 2019
- Led **protein-engineering project** of lead antibody drug (TNT009). Handled all **cloning and expression** of antibody variants and most **protein purification** for the lab. **Collected binding and efficacy (KD and IC50) data** using **biolayer interferometry** and **ELISAs.** Performed **regression analysis**, revealing a log-linear relationship between KD and IC50, enabling our team to select variants to use in downstream experiments.
- Led **research project** to make a protein complex of TNT009 and its target (complement C1s) for crystallography study
- Developed **ELISAs, hemolysis assays,** and **flow cytometry assays** to measure classical and alternative complement activation
- Wrote **Python** script to automate design of short DNA oligos, reducing a task that normally took 2 hours/week of manual labor to a 3 minute script.

GENE YEO LAB, UCSD
La Jolla, CA
Staff Research Associate 1
Aug. 2015 to Nov. 2016
- Used **CRISPR/Cas9** to create, validate, and maintain 36 stable HEK293XT cell lines with end-tagged RNA-Binding Proteins (RBPs) for CLIP-seq as an extension of the Encyclopedia of DNA Elements (ENCODE) project
- Constructed and tested catalytically dead Cas9 (dCas9) fused to split proteins (GFP, Venus, thymidine kinase) as tools to detect RNA abundance in **cultured HEK293XT and U2OS cells**. Performed **flow cytometry** and **Western blot** to quantify results.
- Designed and **Gibson assembled** plasmids encoding human codon-optimized Cas-protein orthologs
- **Co-authored a Cell paper** on using CRISPR/Cas9 to track RNA in live cells

HARRISON WANG
R&D DATA SCIENTIST

Experience (Continued)

GENE YEO LAB, UCSD
La Jolla, CA
Undergraduate Research Assistant
May 2013 to Aug. 2015

- Optimized **fluorescence-coupled activity assay** to measure D- and L-serine levels with up to 5 nM sensitivity on in-vitro synthesized DAO and in-vivo DAO-overexpression HEK293 cell lines
- Used **Flp-In system** to create, validate, and maintain stable D-amino Acid Oxidase (DAO) overexpression HEK293 cell lines
- Took high-resolution **confocal microscopy** images of stained iPSC-derived motor neurons
- Validated via **qPCR** splicing changes of 36 genes related to familial amyotrophic lateral sclerosis (fALS) found via RNA-seq
- **Co-authored a Neuron paper** that included my experiments on investigating the mechanism for how a mutation in an RNA-binding protein (hnRNPA2B1) can cause ALS

Teaching Experience

OFFICE OF ACADEMIC SUPPORT AND INSTRUCTIONAL SERVICES, UCSD
La Jolla, CA
Physics Tutor
Apr. 2013 to Mar. 2015

- Independently led 2-hour workshops of 10 to 20 students twice a week
- Created 18 original physics lectures with 4 practice problems each
- Fostered a collaborative learning environment free of discrimination and judgment

Publications

EPIDEMIOLOGICAL AND BIOLOGICAL ASSOCIATIONS OF SARS-COV-2 VARIANTS BASED ON REAL WORLD OBSERVATIONAL DATA FROM TAQPATH COVID-19 COMBO KIT
Nov. 2021

Imran Mujawar, Mani Manivannan, **Harrison Wang**, David Woo, Teddy Proctor, Jared Auclair, Manoj Gandhi
Conference Abstract. Multiple submissions in review.

ELIMINATION OF TOXIC MICROSATELLITE REPEAT EXPANSION RNA BY RNA-TARGETING CAS9
Aug. 2017

David A. Nelles, Elaine Pirie, Steven M. Blue, Ryan J. Marina, **Harrison Wang**, Isaac A. Chaim, James D. Thomas, Nigel Zhang, Vu Nguyen, Stefan Aigner, Sebastian Markmiller, Guangbin Xia, Kevin D. Corbett, Maurice S. Swanson, and Gene W. Yeo. **Cell**. 2017 Aug 24;170(5):899-912.e10. doi: 10.1016/j.cell.2017.07.010. Epub 2017 Aug 10. PMID: 28803727; PMCID: PMC5873302.

PROTEIN-RNA NETWORKS REGULATED BY NORMAL AND ALS-ASSOCIATED MUTANT HNRNPA2B1 IN THE NERVOUS SYSTEM
Oct. 2016

Fernando J. Martinez, Gabriel A. Pratt, Eric Van Nostrand, Ron Batra, Stephanie C., Huelga, Katannya Kapeli, Peter Freese, Seung J. Chun, Karen Ling, Chelsea Gelboin-Burkhart, Layla Fijany, **Harrison Wang**, Julia K. Nussbacher, Hong-joo Kim, Rea Lardelli, Balaji Sundararaman, John P. Donohue, Jens Lykke-Andersen, Frank Bennett, Manuel Ares Jr., Christopher B. Burge, J. Paul Taylor, Frank Rigo, Gene W. Yeo. **Neuron**. 2016 Nov 23;92(4):780-795. doi: 10.1016/j.neuron.2016.09.050. Epub 2016 Oct 20. PMID: 27773581; PMCID: PMC5123850.

Data Projects

CLIMBING LOG WEB APPLICATION

Live app at: https://harrisonized-climbing-app.herokuapp.com
Blog post at: https://harrisonized.github.io/2020/11/05/climbing-dashboard.html
Built a **Flask app** to function as a **live dashboard** that keeps track of my climbing records. On the back-end, executes **SQL queries** to extract data from **Heroku Postgres**, transform the data, and generate **Plotly interactive visualizations**. On the front-end, displays the **visualizations** in logically organized web pages. Visualizations are **cached** to streamline data access and improve user experience. Users can **upload custom data** to generate their own figures.

Education

Metis Data Science Bootcamp
2019

University of California, San Diego
2015
Double Major: B.S. Physics, B.S. Physiology & Neuroscience