Patricia Mason | Data Analyst

Hutto, TX • (408) 398-3943
masonpatricia@hotmail.com • www.linkedin.com/in/patricia-mason • Github/masonpatricia
masonpatricia.github.io

A resourceful data scientist focused on genomics, bioinformatics, and health-related DNA analytics. I utilize my academic background to define, develop, and document novel end-to-end computational solutions and software. A loyal employee with a collaborative and bold approach, I successfully manage time-sensitive projects and large-scale deliverables.

Skills

Laboratory | PCR, ELISA, DNA sequencing, gel electrophoresis, blood cell isolation, yeast fermentation, cDNA libraries, Northern and Southern blot, vector cloning
Languages/Environments | Python, SQL, Scala, Google Colab, Git, Slack
Packages | Pandas, NumPy, NLTK, BeautifulSoup, TensorFlow, SciKitLearn
Data Visualization | Seaborn, Plotly, Tableau, Excel
Modeling | Statistical Modeling, Regression Models, Classification Models, Time Series Analysis, Neural Networks, Image Classification
Statistics | Statistical Distributions, Bayesian Analysis, p-Values, Hypothesis Testing, Regression and Categorical models, Metrics

Projects

Differential Gene Expression analysis of Ulcerative Colitis and Crohn's Disease
Differential Gene Analysis and Model Creation using Microarray Data from NCBI GEO 12/2020

- Compared Bioconductor against PCA, kNN, Random Forest, and Logistic Regression to determine which genes distinguish disease and normal samples with microarray RNA expression analysis.
- Created a model that predicted ulcerative colitis 100% accuracy and Crohn's Disease 92% accuracy from differential gene expression.
- Technology used: Jupyter Notebooks, Google Colab, Pandas, RdKit, PaDel, GEO NCBI, KMeans and PCA Statistical Models

Predicting a Biological Response
Created a Classification Model Predicting a Biological Response of a Molecule 10/2020

- Predicted the response of biological molecules given 3700 characteristics with 7 classification models with 75% accuracy on unseen data
- Jupyter Notebooks, Pandas, Numpy, SciKitLearn, Statistics Models and Ensemble Packages

SubReddit Title Prediction Model
Four-classification Model to Predict Differences in Subbreddits 10/2020

- Categorized Subreddit posts by title with 80% accuracy on unseen data using NLP
- Jupyter Notebooks, Pandas, Numpy, SciKitLearn, Natural Language Toolkit, Python Reddit API Wrapper (PRAW), MatPlotLIb

Ames, Iowa Housing Price Model
Linear Regression Model to Predict Home Values 10/2020

- Predicted the price of a home explaining 90% of variability of the price
- Jupyter Notebooks, Pandas, Numpy, SciKitLearn, MatPlotLib

Education

Data Science Immersive Student, September 2020 - December 2020
General Assembly | Remote
12-week full-time immersive educational program strengthening Data Science skills including: Python, SQL, data cleaning, data visualization, regression models, classification models, web-scraping, APIs, NLP, advanced supervised learning, unsupervised learning, time series analysis, and statistics.

Biotechnology Certificate
De Anza College / Cupertino, CA

B.S. Biology
Santa Clara University / Santa Clara, CA

Professional Experience

Round Rock Christian Academy, Round Rock, TX 1/2003 - present
Shadow Mountain Baptist School, Morgan Hill, CA
Educator and Tutor

- Taught elementary and secondary students as a substitute, following instructor notes and leading the classroom.
- Tutor elementary-aged students in phonics, math, reading comprehension, and spelling with a focus on memory retention and subject-matter understanding.
- Tutored ADHD/Asperger's diagnosed student for two years in STEM subjects, increasing his grade point average from 2.5 to 3.5.

Incyte Genomics | Palo Alto, CA
Bioinformatics Associate 02/2001 - 10/2002

- Analyzed cDNA sequences to identify genes from high-throughput DNA sequencing for downstream patenting of over 200 sequences
- Organized gene libraries for further analysis
- Troubleshot difficult gene sequences for completion of full transcript.
- Prepared and categorized sequences for weekly presentation and delivery to legal.

Systemix | Palo Alto, CA
Research Associate 05/1999 - 07/2000

- Performed ELISAs and sequenced DNA of various stem cell lines for stem cell therapy.
- Created CMV transfected cells for in-vivo murine gene-therapy testing.

Stanford University | Palo Alto, CA
Research Associate 09/1996 - 05/1999

- Sequenced MHC region in trout to locate new genes through BLAST searches resulting in a publication.
- Maintained international HLA A, B, C databases in collaboration with transplant immunologists worldwide.

- Isolated WBCs from various humans, apes, and monkeys, EBV transformed them, and preserved them in liquid nitrogen for various projects in the lab.
- Cultured cells for the natural killer cell projects.

Publications

Shum, B.P., Mason, P.M., Magor, K.E. *et al.* Structures of two major histocompatibility complex class I genes of the rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics* 54, 193–199 (2002). https://doi.org/10.1007/s00251-002-0450-z

Here we describe two rainbow trout major histocompatibility complex (MHC) class I genes characterized from λ phage genomic clones prepared from a single fish. An open reading frame is maintained, and thus the gene MhcOnmy-U71 could be expressed in this individual.