

# JACK ETHEREDGE, PHD DATA SCIENTIST

✉ jack.etheredge@gmail.com

📧 jacketheredge.com

📞 8033671152

📍 New York City, 10016

in jack-etheredge-phd-

84532585/

🔗 Jack-Etheredge

Data scientist with experience predominantly in bioinformatics - genomic and video data. Research scientist with 11 years of experience in developmental biology of neuronal cells. Experience with transgenic Drosophila and mice as model organisms. 7 years of experience with Next-Gen sequencing and RNA-Seq - both sample preparation and data analysis.

## Skills

### PROGRAMMING

R  
Python  
bash  
SQL  
Machine Learning  
Deep Learning  
Big data analysis

### BIOINFORMATICS

Video tracking  
Computer vision  
RNA-Seq  
Genomics  
Neuroscience  
Developmental Biology

## Awards

Coursera -  
Using Python to Access Web Data  
Charles Severance, University of Michigan

Coursera - Machine Learning  
Andrew Ng, Stanford University

Barbara Houtz, STEM Education Solutions  
· Scientists Teaching Science  
Pedagogy course for science educators

Admitted to Insight Health Data Science

Admitted to NYC Data Science Academy

## Education

University of Cambridge  
PhD Physiology, Development, and Neuroscience 2018

Emory University  
BS/MS Biology 2011

## Experience

### Metis Data Science Bootcamp

Data scientist

Full-time immersive data science bootcamp including five full-cycle projects from inception to execution, including data acquisition to analysis. Machine learning and statistical modeling methods: clustering, classification, natural language processing, supervised and unsupervised learning including linear and logistic regression, deep learning. Python tools: numpy, pandas, SciPy, scikit-learn, NLTK, Word2Vec, Keras, TensorFlow, Pytorch, matplotlib, D3.js, seaborn, Beautiful Soup, Selenium, Scrapy, lxml. Big data tools: AWS, SQL (Postgres and SQLAlchemy), MongoDB, Hadoop, PySpark.

Completed several end-to-end data science projects working primarily in Python:

- 1) **Exploratory data analysis of MTA turnstile data:** <https://github.com/Jack-Etheredge/MTA-Turnstile-Analysis>
- 2) **Linear regression of web-scraped Steam video game data:** Predicting number of users and determining the most important predictive features. <https://github.com/Jack-Etheredge/Steam-Web scraping-Linear-Regression>
- 3) **Predicting early hospital readmissions among diabetic patients:** Many different supervised machine learning models were evaluated and a cost function was used to optimize a random forest predictor as a web app. <https://github.com/Jack-Etheredge/Predicting-early-hospital-readmissions>
- 4) **Project Guten-bag-of-words:** Automatic genre grouping, similar book recommendation, and visualization of narrative patterns in Project Gutenberg books using unsupervised machine learning and natural language processing (NLP).  
\* Web app: <http://flask-env.svfjqyqgqf.us-east-1.elasticbeanstalk.com/>  
\* Code: <https://github.com/Jack-Etheredge/Project-guten-bag-of-words>
- 5) **Radiology-CNN:** Using convolutional neural networks to locate and classify brain tumors

### Howard Hughes Medical Institute

Research Specialist

Performed data analysis for high-throughput behavioral assays and RNA-Sequencing data. Updated and revived deprecated analysis pipeline code in bash, R, and Matlab. Created documentation for behavioral assays and analysis pipelines in bash, R, and Matlab. Created new analysis pipelines for genomic data in bash and R. Trained new junior members of the team (scientific research technicians). Supervisor: Gudrun Ihrke, PhD

### Howard Hughes Medical Institute and University of Cambridge

PhD Graduate student

Project: Transcriptional identity of neuronal stem cell lineages in fruit fly nerve cord - How are diverse neurons generated from similar stem cells? I performed RNA-Seq analysis on data I generated for 64 samples in an organism with ~16000 genes (Drosophila). This involved bash, SQLite, and R. Data was visualized with principal component analysis and heatmaps with hierarchical clustering. I performed genomic alignments with command line software using a local compute cluster through ssh and LSF. I used multiple analysis pipelines to determine the concordance in statistically significant differential gene expression. University of Cambridge Advisor: Andrea Brand, PhD  
HHMI Advisor: Jim Truman, PhD

### Howard Hughes Medical Institute - Jim Truman, PhD

Howard Hughes Medical Institute Janelia Undergraduate Scholar

Project: Neuronal lineage developmental independence - Do co-fasciculating neuronal lineages in the fruit fly nerve cord require one another for proper development?  
Advisor: Jim Truman, PhD

### Emory University - Ping Chen, PhD

Master's student and undergraduate laboratory technician

Project: Planar cell polarity in the developing mouse utricle and saccule - subcellular localization of PCP proteins do not follow the line of reversal in utricle and saccule.  
Advisor: Ping Chen, PhD - Departments of Cellular and Developmental Biology & Neuroscience

## Projects

Genetic tools to study Juvenile Hormone Action in Drosophila  
Baumann, A.A., Texada, M.J., Chen, H.M., Etheredge, J.N., Miller, D.L., Picard, S., Warner, R., Truman, J.W., Riddiford, L.M. (2017) "Genetic tools to study Juvenile Hormone Action in Drosophila." Scientific Reports

Generating adult neuron diversity in Drosophila: Transcriptional diversity in ventral nervous system neuronal lineages  
Etheredge, J.N., Baumann, A.A., Brand, A., Truman, J.W. (in-prep) "Generating adult neuron diversity in Drosophila: Transcriptional diversity in ventral nervous system neuronal lineages"  
Summary: I generated RNA-seq transcriptomes of 11 of 33 neuronal stem cell lineages in the Drosophila CNS analogous to the spinal cord. I used bash and R to determine unique and combinatorial expression between stem cell lineages. Broader developmental classes of neurons were also assayed for differences arising from separate cell fate commitment and maturation processes.

Transcriptional changes in a neuronal lineage in a Drosophila extended third instar  
Etheredge, J.N., Miller, D., Brand, A., Truman, J.W. (in-prep) "Transcriptional changes in a neuronal lineage in a Drosophila extended third instar"  
Summary: I generated larvae that did not undergo pupation and took an RNA-Seq transcriptome time series of neurons that displayed aberrant morphological changes. I used bash and R to compare gene expression between time points.

Drosophila melanogaster CA-LP1 neurons control developmental dynamics.  
Baumann, A.A., Etheredge, J.N., Wong, C.D., Spokony, R. (in-prep) "Drosophila melanogaster CA-LP1 neurons control developmental dynamics."  
Summary: Activation of neurons projecting to the corpora allata produce pupation timing differences when they do not express a receptor for juvenile hormone.