

BETH BAUMANN PHD



Baumann.Bethany@gmail.com

847-987-5433

Chicago, IL 60622

/baumannbethany/

Beth526

Skills

PROGRAMMING LANGUAGES

Python

R

SQL

Command Line

LIBRARIES

Scikit-learn

Matplotlib

Seaborn

BeautifulSoup

Selenium

ggplot2

Pandas

Spacy

Gensim

Keras

TOOLS

Git

Google Cloud Platform

Dask

MACHINE LEARNING

Regression

Classification

Clustering

Natural Language Processing

Neural Networks

BIOINFORMATICS

Alignment Pipelines

Pathway Analysis

Differential Expression

Bioconductor

Single-cell Transcriptomics

Education

University of Illinois at Chicago Fall 2018 to Fall 2019
Certificate Bioinformatics
2019

Northwestern University Sept. 2009 to May 2016
PhD Biological Sciences
2016

University of Pennsylvania May 2005 to May 2009
BSE Bioengineering 2009
Minor Mathematics 2009

Experience

Metis

Data Scientist

Remote

Sept. 2020 to Current

- Metis is an ACCET-accredited 12-week immersive data science bootcamp focused on project-oriented learning
- The curriculum focuses on Python, statistics, and supervised and unsupervised machine learning
- Completed five self-designed data science projects from conception to presentation, including data collection, data management, exploratory data analysis, modeling, and visualizations

University of Illinois at Chicago

Postdoctoral Researcher

Chicago, IL

June 2016 to Current

- Published 5 research papers (3 first author)
- Awarded 2 competitive research grants from Urology Care Foundation and Dept. of Defense worth \$230k total
- Developed expertise in most standard molecular biology techniques and library preparation
- Bioinformatically analyzed RNA-seq data, CHIP-seq data, microarray data, and single-cell transcriptomic data
- Mentored and managed projects of undergraduate students

Cactus Communications

Freelance editor

Remote

Nov. 2018 to Jan. 2020

- Edited a variety of medical and life sciences manuscripts with high customer satisfaction ratings

Data Science Projects

Detecting Tumor Mutational Signatures with a CNN

Dec. 2020

Generated a dataset of >3 million synthetic but realistic 100bp normal or tumor sequencing alignments by inserting tumor mutations from The Cancer Genome Atlas and normal human variation from Gnomad into the human reference sequence. The sequence alignments were used to train a convolutional neural network for classification as normal or coming from one of 12 tumor types. This model had recall between 17 and 59% for 9 of the tumor types and provided insights into the mutational patterns and subsequences most useful in classification of each type.

NLP of Cell Biology Abstracts

Nov. 2020

Collected 300k cell biology abstracts from PubMed for topic modeling of signaling pathways. Spacy and SciSpacy were used for tokenization and named-entity tagging of genes/gene products, and Gensim and Scikit-learn were used for latent topic detection and clustering. The best results were achieved with a small Word2Vec model.

Overall Survival of Clinical Trials

Oct. 2020

Using trial data from clinicaltrials.gov, optimized several classifier algorithms in scikit-learn to obtain the highest possible F1 score in predicting early trial termination. Created a Streamlit app for clinical trials designers to see how trial features affect termination risk.

Predicting Vinyl Record Prices

Sept. 2020

Scraped data from Discogs and Spotify to produce a regularized regression model in sklearn that could predict the price of first-release vinyl albums.

Activities

University of Illinois at Chicago Postdoctoral Association · Executive Member

Jan. 2019 to July 2020

- Moderated the Data Science panel and helped plan annual Careers Symposium
- Planned and presented webinars with interactive notebooks on Python and Pandas for Chicago-area postdocs

Big Data Training for Translational Omics Research - Purdue Statistics

June 2017

- Completed a two-week long intensive bootcamp in R and genomics research analysis

Science Club at Boys and Girls Club · Mentor

2013 to 2016

- Lead small group STEM projects and encouraged students toward STEM careers

Publications

Baumann B, Lugli G, Gao S, Zenner M, Nonn L. High levels of PIWI-interacting RNAs are present in the small RNA landscape of prostate epithelium from vitamin D clinical trial specimens. *The Prostate*. 2019; 79(8):840-855.

McCray T, Moline D, Baumann B, Vander Griend DJ, Nonn L. Single-cell RNA-Seq analysis identifies a putative epithelial stem cell population in human primary prostate cells in monolayer and organoid culture conditions. *American Journal of Clinical and Experimental Urology*. 2019; 7(3):123-138.

Baumann B, Acosta AM, Richards Z, Deaton R, Sapatynska A, Murphy A, Kajdacsy-Balla A, Gann PH, Nonn L. High miR-182 levels associate with low-risk prostate cancer. *American Journal of Pathology*. 2019; 189(4):911-923.

Additional publications: <https://bit.ly/2IAEnXB>