# CALVIN CHI


**CONTACT**

calvin.chi@berkeley.edu
(626) 203-1829
https://github.com/CalvinTChi
https://calvintchi.github.io


**EDUCATION**

**University of California, Berkeley** 2015 - present
Ph.D., Computational Biology (anticipated 2020)
Coursework: ML, computer vision, NLP, deep reinforcement learning, optimization, linear models

**Case Western Reserve University** 2011 - 2014
B.S., Biochemistry
*summa cum laude* (GPA 4.0/4.0)

**INDUSTRY**

**Applied Scientist Intern at Amazon** May 2019 - Aug 2019
- Developed LSTM-based deep learning model for credit abuse classification for Amazon Business, estimated to reduce credit write-off loss by 20%.
- Shipped $\sim$ 1,000 lines of python and SQL code for data and model production.

**Data Scientist Intern at Amazon** June 2018 - Aug 2018
- Built random forest models for customer scoring of $\sim$ 17MM Twitch Prime sign-ups.
- Achieved an AUC of 0.80 on fraud detection and 70% accuracy on churn prediction, estimated to save $5 MM for first half of 2018.

**RESEARCH**

**Chi C**, Ye Y, Huang H. Hierarchical Clustering using Sparse Canonical Correlation Analysis for Pharmacogenomic Cluster Discovery. Bioinformatics. [In Preparation]

**Chi C**. HLA Allele Imputation with Deep Convolutional Neural Network. Bioinformatics. [In Preparation]

**Chi C**, Quach D, Quach D, Taylor K, Barcellos L, Criswell L. Epigenetic stratification reveals hypomethylation of immune genes between severe and mild Sjögren's Syndrome patient subgroups. Arthritis and rheumatism [In Preparation]

**Chi C**, Quach D, Quach D, Taylor K, Barcellos L, Criswell L. Hypomethylation of Immune Genes Mediates Methylation Quantitative Trait Loci at the Major Histocompatibility Complex in Sjögren's Syndrome. Ann. Rheum. Dis [In Preparation]

**Chi C**, Shao X, Rhead B, Gonzales E, Smith JB, Xiang AH, et al. (2019) Admixture mapping reveals evidence of differential multiple sclerosis risk by genetic ancestry. PLoS Genet 15(1): e1007808. https://doi.org/10.1371/journal.pgen.1007808

**PROJECTS**

**Embedding-Augmented Deep CNN for PubMed Journal Recommendation** Dec 2018
- Journal detection from PubMed abstract with 415,381 programmatically-collected abstracts.
- Compared multitask and embedding-augmented CNNs with output space of 1,548 journals.
- Best performance when CNN input augmented with topic and impact factor embeddings, with accuracy 23.7% and 90% of true journals in top 60 recommendations.

**Data Augmentation using GAN for Breast Cancer Classification** May 2018
- Synthetic data augmentation using DCGAN to improve histology breast cancer classification with Resnet-18 re-trained on 5,547 breast histology images.
- Augmentation with 400 DCGAN images improved prediction accuracy and precision by 5% and 12% respectively, but decreased recall by 15%.

**SKILLS**

**Programming**: Python R, SQL, Java, Bash, Matlab, HTML, CSS, Javascript, C
**Libraries**: Scikit-Learn, H2O, Keras, TensorFlow

**AWARDS**

NSF Graduate Research Fellowship Mar 2017