Emanuel Flores

Computational Biology & Applied Math
Ph.D. candidate @ Caltech
Matt Thomson Lab
Dept. of Biology and Bioengineering

manuflores
manuflores@caltech.edu
www.linkedin.com/in/eflobau/

Research experience

Topological exploration of developmental processes (Caltech)
I'm currently developing analysis algorithms using concepts from algebraic topology to find global and local insights in the structure of data. Recently, I applied such approaches to single-cell transcriptome spaces. I discovered that topological signatures represent convergent and self-renewing differentiation trajectories in gene expression space. Broadly, we were the pioneers in formally establishing the topology of developmental single-cell manifolds.

Transcriptomic data integration across time (CZ Biohub)
Used a geometric model to understand technical effects in single cell transcriptome data. Using this model, we built an analysis pipeline that enabled to faithfully integrate datasets from different timepoints that were possibly subject to the studied technical effects.

Representation learning of cells and molecules for drug discovery (Caltech)
Led a project for the discovery of immunomodulatory drugs using deep learning. We were able to automatically extract molecules with anti-inflammatory effect using a classifier trained the model embedding space. We found molecules such as the tyrosine kinase inhibitor Dasatinib classified as anti-inflammatory using our deep model. Furthermore, we found that the Dasatinib-perturbed cell population was enriched with M2 macrophages, a class of cells that drives anti-inflammatory responses, suggesting that our approach was effective at extracting global phenotypic responses caused by small molecule perturbations.

Integrative multiomics analysis of mTOR using CCA (Caltech / National Cancer Inst. NIH)
Led a bioinformatics team in multiomics data integration project to understand non-canonical mTORC signaling. We used the Canonical Correlation analysis (CCA) linear model to elucidate molecular programs common and exclusive between different mTORC pathways. Using this model, we were able to find a link between the metabolism and genetic networks controlling cell growth affecting primarily upon mTORC3 and mTORC1 knockout cell lines exclusively.

Internships

(2022) Data Science Internship at Chan Zuckerberg Biohub. Hosted by Angela Pisco.

(2018) Summer Undergraduate Research Fellowship (SURF), Caltech. Hosted by Richard Murray.

Skills

Computational mathematics: Computational Linear Algebra, Topology and Differential Geometry, Deep Learning, Multimodal Data Analysis, High-dimensional Statistics, Convex Optimization, Clustering and Segmentation.

Programming: Python Scientific Computing (numpy, scipy, pytorch, pandas), Functional Programming, Bash scripting, Streaming, R, Matlab.

Presentations

(2023) Poster at Topological Data Analysis Week (TDAWeek) conference. Kyoto, JP.

(2019) Poster at NeurIPS Workshop - Learning Meaningful Representations of Life. Vancouver, CA.

(2019) Workshop at PyData Los Angeles. Title: Analyzing genetic networks using neural networks. Los Angeles, CA, USA.

(2019) Machine Learning and Data Science Workshop at Avis Car Rental Mexico Corporation. Merida, MX.

Publications

E. Flores-Bautista, M. Thomson (2023) Unraveling cell differentiation mechanisms through topological exploration of single-cell developmental trajectories. *bioRxiv.*

W. Ireland , S. Beeler, **E. Flores-Bautista**, *et al.* (2020) Deciphering the regulatory genome of Escherichia coli, one hundred promoters at a time. *Elife.*

A. Halleran, **E. Flores-Bautista**, R. Murray (2019) Quantitative characterization of random partitioning in the evolution of plasmid-encoded traits. *bioRxiv.*

E. Flores-Bautista, *et al.* (2018) Functional prediction of hypothetical transcription factors in E. coli K-12. *Computational and Structural Biotechnology Journal.*