UC San Diego

JACOBS SCHOOL OF ENGINEERING

"THE FUTURE OF SCIENTIFIC PROGRESS WILL BE AIDED BY BRIDGING THE GAP BETWEEN THE MILLIONS OF PUBLISHED RESEARCH ARTICLES AND MODERN DATABASES SUCH AS THE ALLEN BRAIN ATLAS (ABA)."
- VOYTEK ET AL.

NEURAL CORRESPONDENCE MAPPING

Adita Zeqollari, Arlens Zeqollari, Erik Hoye, Robert Reeves / Advisor: Bradley Voytek
Github: https://github.com/voytek/NCM

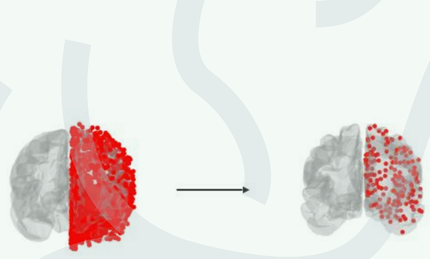
PROBLEM STATEMENT

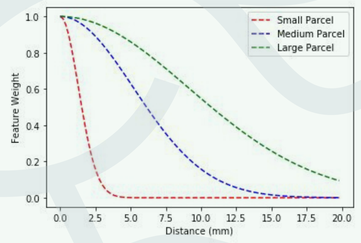
The **analysis & comparison of modern neuroscience data is prohibitively diverse, leading to challenges in aggregation, verification & collaboration** in the neuroscience research community. These challenges make hypothesis generation difficult since findings from one lab are not easily ingested by other labs with different processing techniques, sampling & brain atlases.

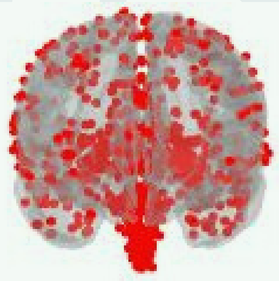
The **NCM Python package was developed to aid neuroscience research discovery by facilitating semi-automatic hypothesis generation** with the creation of a flexible & extendable open-source Python library that merges disparate neuroscience data into a common coordinate system.

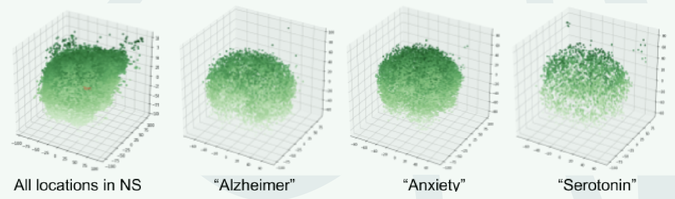
DATA

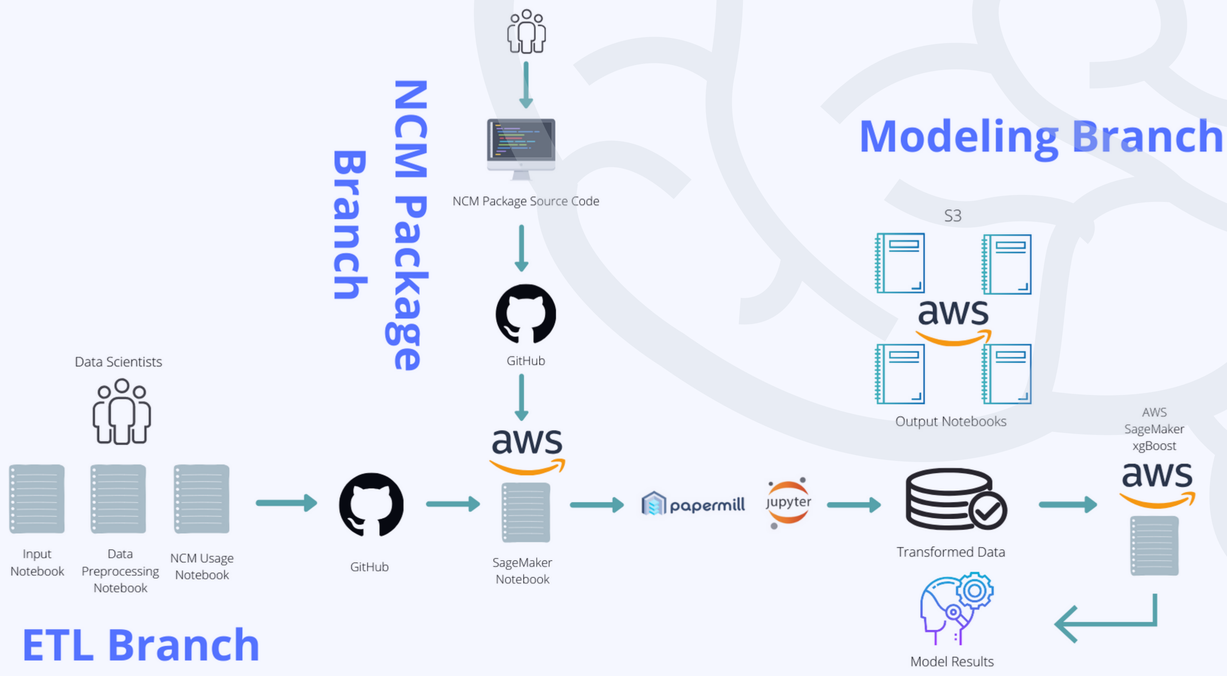
- **Allen Brain Atlas**
  - mni-xyz & mri-voxel
  - Gene expressions (>21,000 genes; 6 donors)
- **NeuroSynth**
  - Word correlations from published research to regions of the brain
- **ECoG**
  - Neural power spectra, characterizing both the aperiodic 'background' component & periodic components as overlying peaks, reflecting putative oscillations









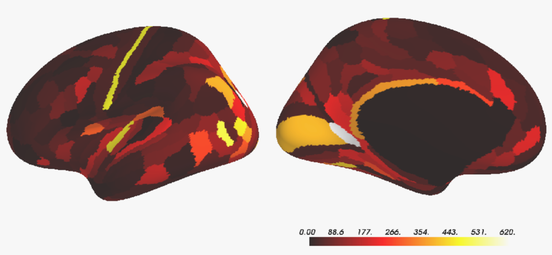


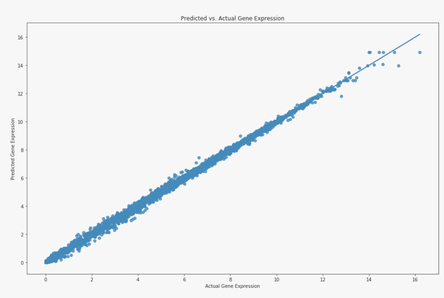
NCM PYTHON PACKAGE

- **Modular, class-based architecture** to support future development and enhancement
- **Interfaces with common coordinates systems** such as MNI-XYZ & MRI-voxel
- **Allows for mapping to a broader parcel scheme** defined by an atlas provided in a Nifti file
- **Custom weighting algorithm** is used to process & map data points to other coordinate systems
- **Single or dual hemisphere analysis** is supported
- **Local or cloud deployment** can be utilized

MODELING OVERVIEW

- **xgBoost modeling** of gene expressions from the ABA data was performed using AWS Sagemaker.
- **Bayesian hyperparameter tuning** jobs were spun up to parallelize & improve model training and tuning accuracy & speed.





Model Performance:
Mean squared error: 0.02

Accuracy was robust across different parcels & distributions of gene expression values.

KEY INSIGHTS & CONCLUSIONS

- A package could be developed in Python to ingest, transform & visualize disparate neuroscience data.
- The package could be utilized to perform modeling of gene expressions in the ABA dataset. This modeling showed that gene expressions can be predicted accurately.
- AWS computing could be leveraged to expand the functionality & efficiency of the package.
- Interesting and unintuitive associations between parcels for a given gene can generate hypotheses about physical or functional interactions.
- Modularity & scalability of the package supports further enhancements & collaborative development.

"...THE WORK YOU'VE DONE IS REALLY AWESOME, AND I HOPE MY LAB CAN PICK IT UP AND RUN WITH IT."
- PROFESSOR BRADLEY VOYTEK