Geneformer fine-tuned to distinguish Parkinson's Disease

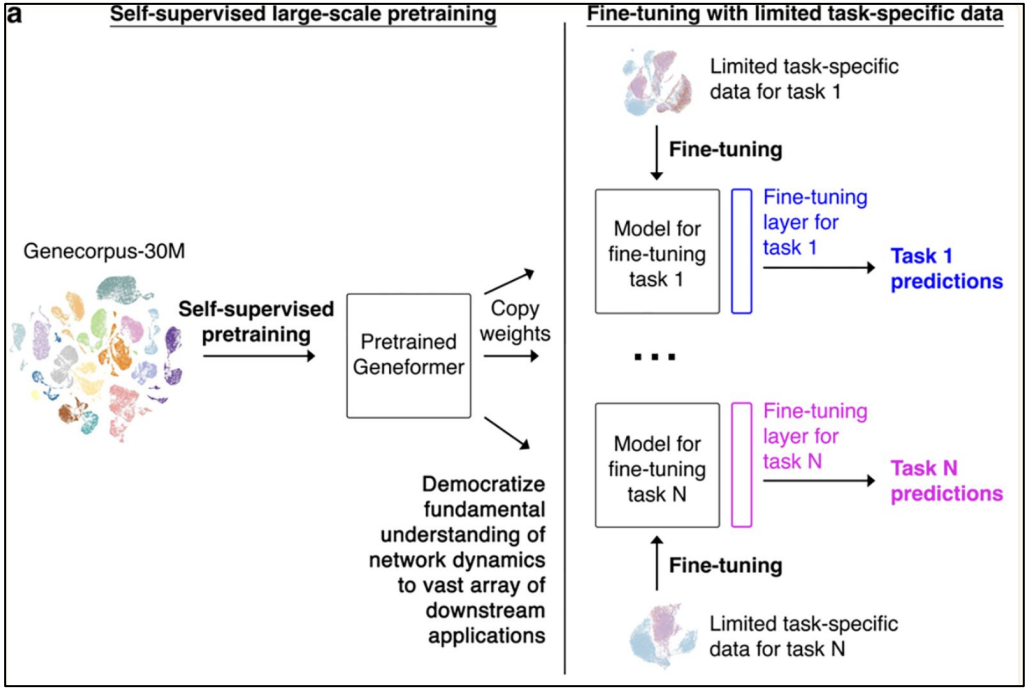
Eryk Kropiwnicki
07/22/2024

Background

- Geneformer is pretrained on Genecorpus-30M, a collection of 30M single cell transcriptomes across various tissue and cell type contexts

- Geneformer learns a latent representation of healthy single cell transcriptomes

- Geneformer can be fine-tuned on downstream tasks using limited task-specific data by leveraging the pre-trained representation of context-specific transcriptomes



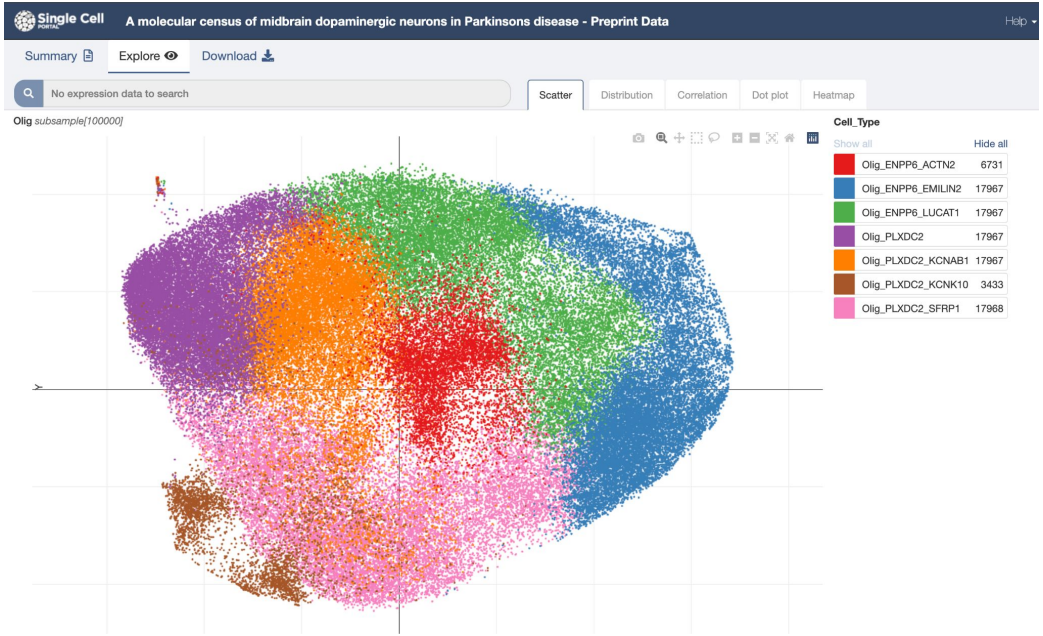
Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, Mantineo H, Brydon EM, Zeng Z, Liu XS, Ellinor PT. Transfer learning enables predictions in network biology. Nature. 2023 Jun;618(7965):616-624. doi: 10.1038/s41586-023-06139-9. Epub 2023 May 31. PMID: 37258680; PMCID: PMC10949956.

Approach

- We explored available single-cell RNA sequencing datasets profiling Parkinson's disease

- We fine-tuned Geneformer on a dataset of ~143,000 oligodendrocyte cells to classify Parkinson's single cell transcriptomes.

- We evaluated the performance of fine-tuned Geneformer against Logistic Regression and Neural Network baselines

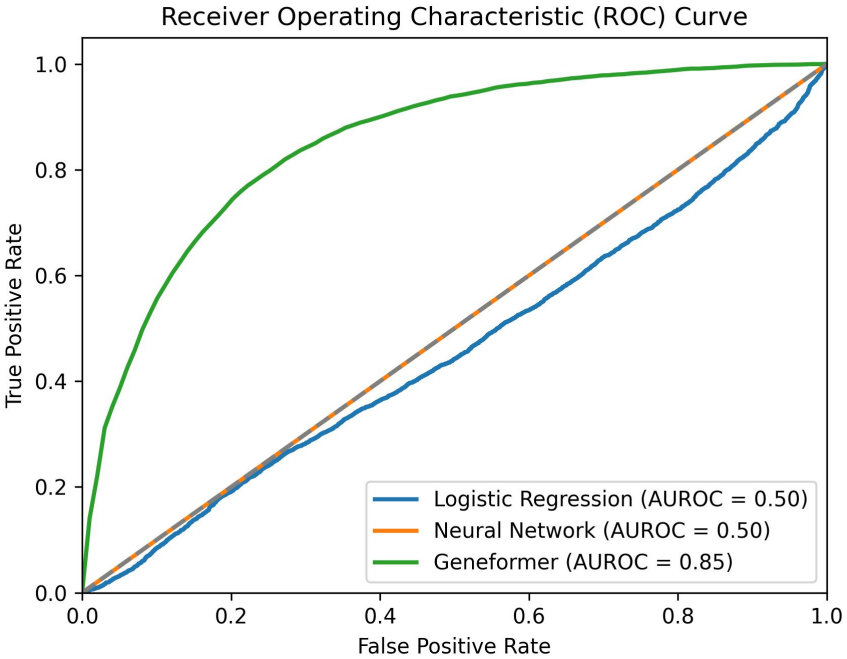


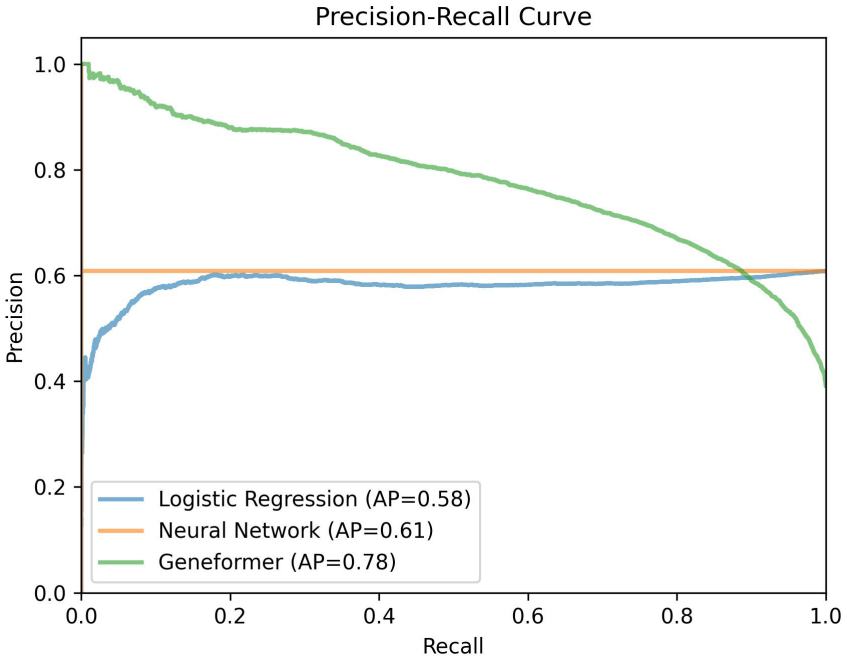
Tushar Kamath, Abdulraouf Abdulraouf, SJ Burris, Vahid Gazestani, Naeem Nadaf, Charles Vanderburg, Evan Z Macosko

bioRxiv 2021.06.16.448661; doi: https://doi.org/10.1101/2021.06.16.448661

Results

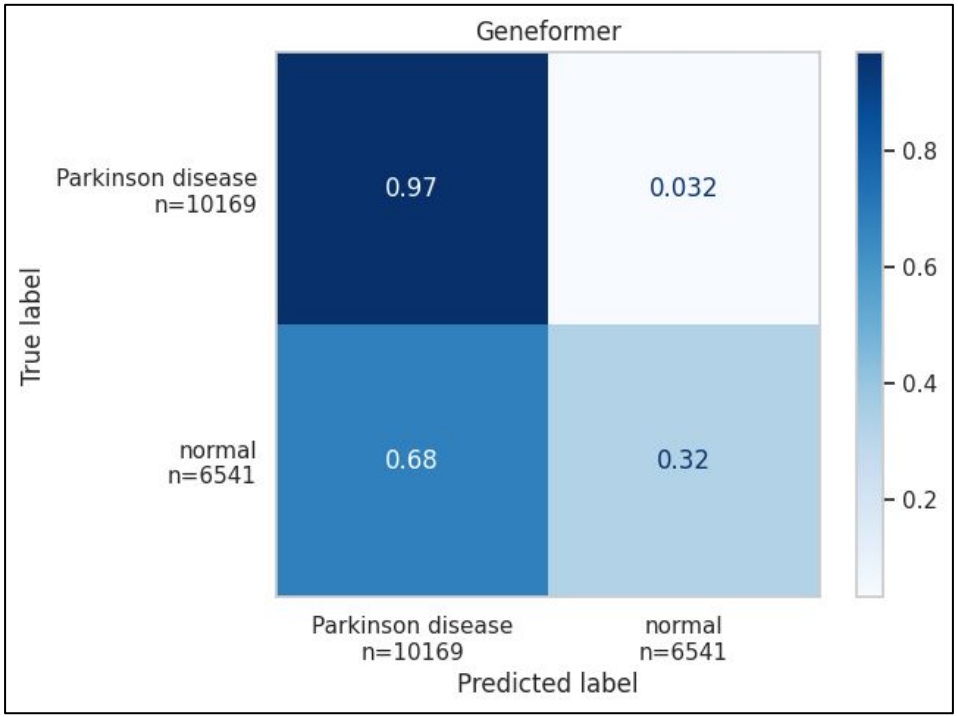
Geneformer outperformed simple baselines on the basis of AUROC and average precision score.





Results

While the model is adequate at identifying true positives for instances of Parkinson's afflicted cells, it struggles with differentiating between Parkinson's disease and normal control states, leading to a high rate of false positives.



Future Directions

- Model improvements and running more in-depth hyperparameter optimization for fine-tuning
- Extending this approach to multiple cell types instead of just oligodendrocytes
- Running ablation/overexpression experiments