Unravelling Inflammatory Pathways in Parkinson's Disease: Insights from Pathway-Based Machine Learning Analysis of Transcriptomics Data

ELISA GÓMEZ DE LOPE, ENRICO GLAAB
Biomedical Data Science Group, LCSB, University of Luxembourg
elisa.gomezdelope@uni.lu
@elisagdelope
UNIVERSITÉ DU LUXEMBOURG

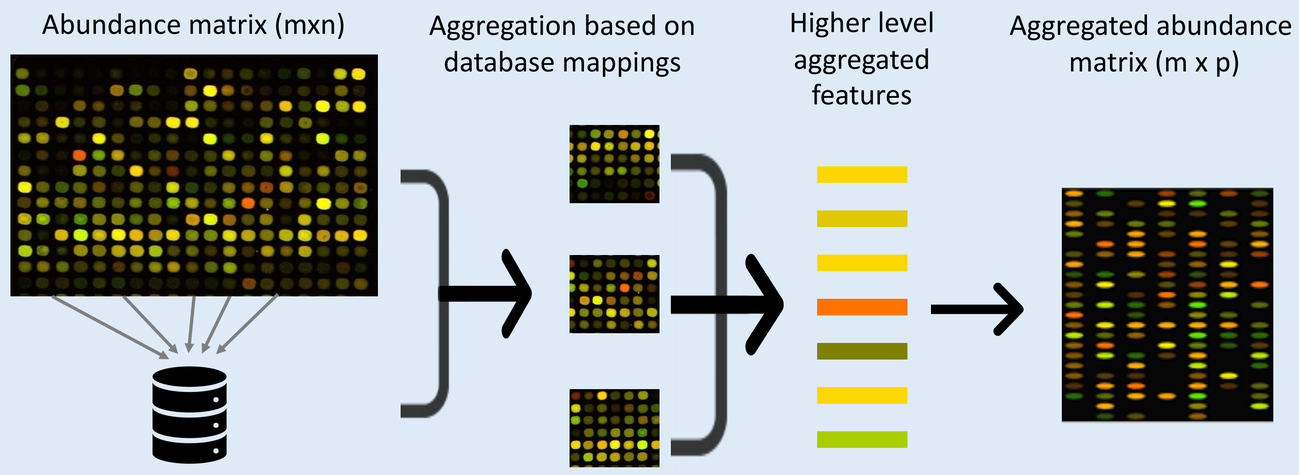
Background

Neuroinflammation has been implicated in the progression of Parkinson's disease (PD) by contributing to dopaminergic neuron loss[1], but the specific molecular pathways involved remain largely unknown. We applied statistical and machine learning (ML) analyses to cross-sectional and longitudinal transcriptomics data from PD patients and controls, examining both gene level changes and aggregated functional representations, such as pathway-, cell compartment- and protein complex-level features.

Methods

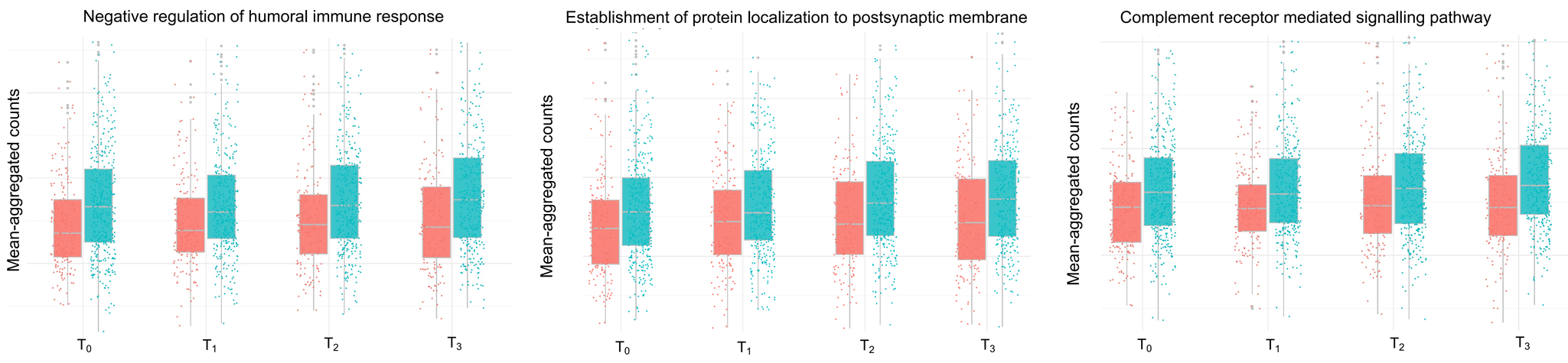
Higher level functional (pathway, cellular location and protein complex) representations of transcriptomic data from the PPMI cohort[2] (whole blood) were generated using aggregation statistics (mean, median, sd) and low-dimensional representations using PCA and principal curves[3] (Pathifier software).

Differential expression analyses accounting for the confounders age and gender, and time correlation analyses (A) were applied to PD case/control data for both gene and functional level representations. Next, we evaluated ML models using a nested cross-validation including the feature selection and parameter optimization for PD versus control sample classification, assessing the performance of pathway-level aggregation statistics and single-level features. A SHAP value analysis was conducted to identify the most informative predictive features.



Results

A. Longitudinal mean-aggregated counts at GO BP level

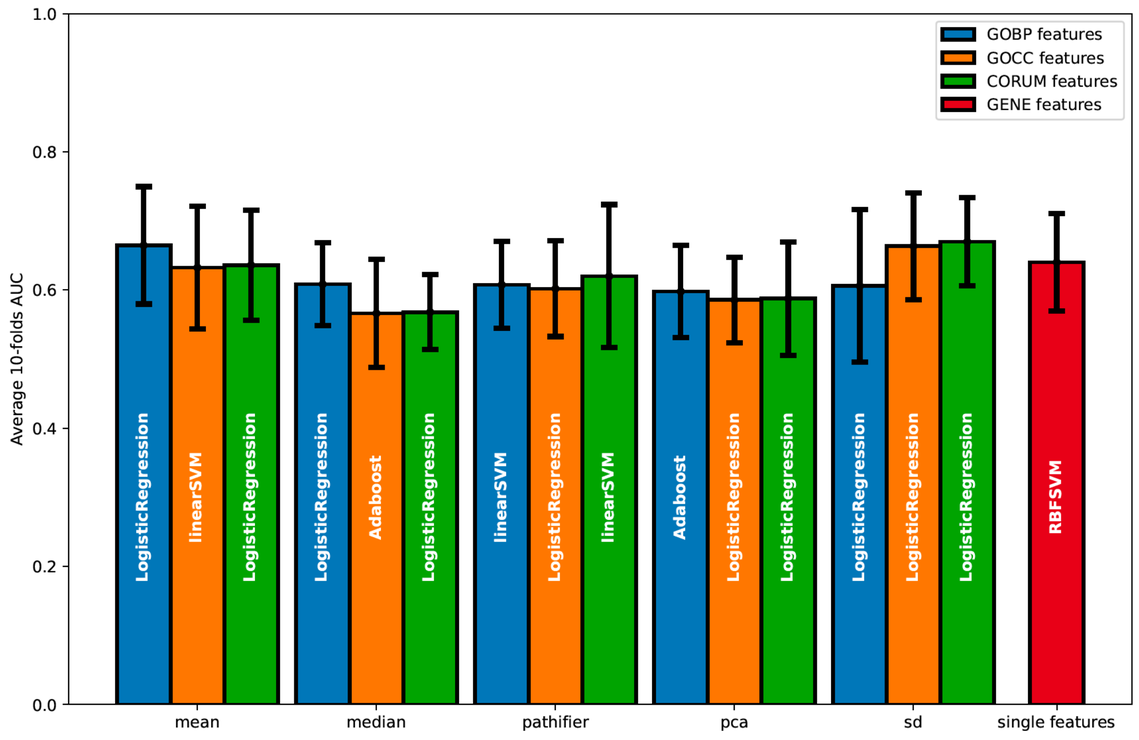


- Inflammation and neurological pathways displayed PD-specific association with time (A)
- Lasso penalty obtained highest cross-validated AUC in a Random Forest model across most transcriptomics-based datasets (B)
- No large differences in performance are observed between single-features and aggregated features. Yet some aggregations outperform genes as predictors (C)
- TNF-alpha/NF-kappa B signaling complexes among the most relevant features for the model's outcome (D)

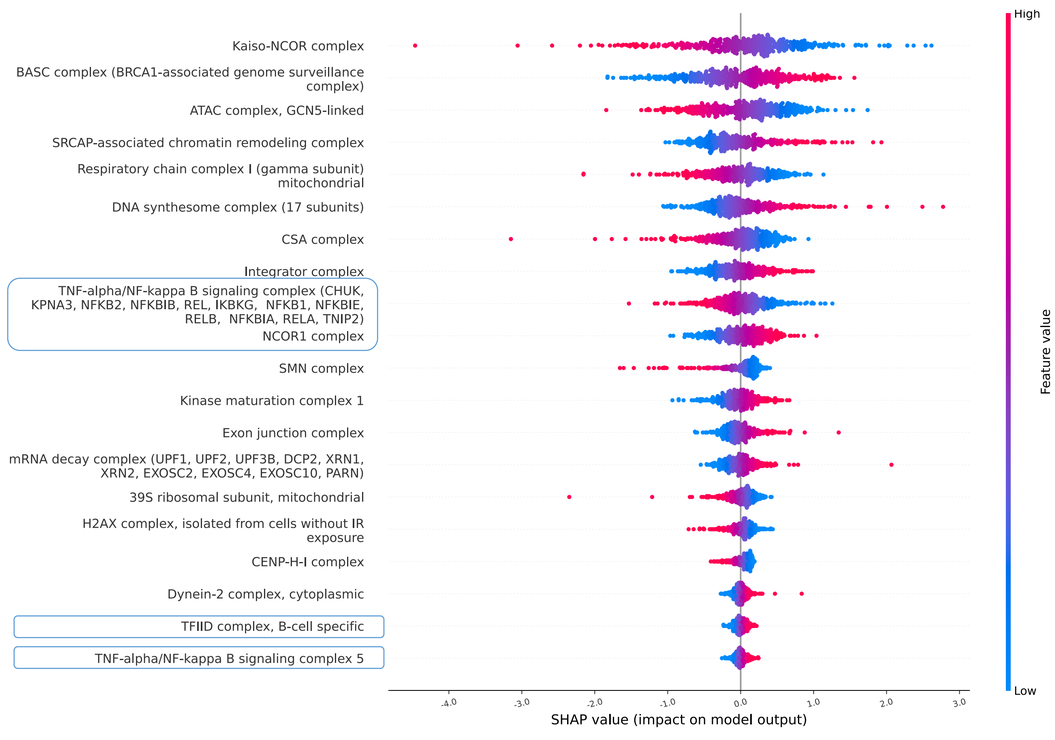
B. Predictive performance for PD diagnosis measured as average cross-validated AUC ± standard deviation from gene expression data at gene and aggregated levels (mean and standard deviation) using DEA filter, Lasso penalty, a Pearson correlation filter (0.85) and no feature selection at all on a random forest model.

			GOBP		GOCC		CORUM	
Ft. selection	**N ft. cutoff**	**GENES**	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**
DEA	10	0.58 ± 0.05	0.55 ± 0.06	0.54 ± 0.03	0.54 ± 0.05	0.57 ± 0.05	0.55 ± 0.04	0.55 ± 0.06
DEA	100	0.59 ± 0.06	0.58 ± 0.06	0.57 ± 0.06	0.54 ± 0.07	0.56 ± 0.07	0.55 ± 0.05	0.55 ± 0.06
DEA	1000	0.56 ± 0.07	0.54 ± 0.09	0.56 ± 0.06	0.55 ± 0.06	0.55 ± 0.07	-	-
Lasso	-	**0.63 ± 0.08**	**0.64 ± 0.09**	**0.60 ± 0.11**	**0.59 ± 0.05**	**0.63 ± 0.1**	0.54 ± 0.1	**0.66 ± 0.07**
Pearson corr.	-	0.59 ± 0.06	0.57 ± 0.06	0.56 ± 0.06	0.56 ± 0.05	0.56 ± 0.06	**0.56 ± 0.04**	0.56 ± 0.05
No selection	-	0.58 ± 0.06	0.57 ± 0.05	0.57 ± 0.05	0.56 ± 0.05	0.57 ± 0.05	0.55 ± 0.05	0.56 ± 0.04

C. Cross-validated performance of different aggregators for GO BP, GO CC, CORUM and single genes.



D. Top 20 most relevant features for the logistic regression model on CORUM-sd features (0.67±0.06 AUC).



Conclusions

- ML analyses of aggregated features enriched with biological information can reveal robust and interpretable global coordinated alterations in cellular processes.
- (Neuro)inflammatory processes display shared cross-sectional and longitudinal alterations with potential relevance for PD vs. control sample classification.
- Variation (sd) in feature sets representing protein complexes and cellular compartments and average expression changes in biological pathways provide significant predictive information for PD vs. control sample classification.

1. Grotemeyer A, McFleder RL, Wu J, Wischhusen J and Ip CW. Neuroinflammation in Parkinson's Disease – Putative Pathomechanisms and Targets for Disease-Modification. *Front. Immunol.* 13, 878771 (2022).
2. Marek, K. et al. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol 95, 629 (2011).
3. Drier, Y., Sheffer, M. & Domany, E. Pathway-based personalized analysis of cancer. Proc Natl Acad Sci U S A 110, 6388–6393 (2013).

LCSB
FONDATION DE LUXEMBOURG
Fondation du Pélican