Pathway-based machine learning analysis of Parkinson's disease transcriptomics data reveals coordinated alterations in inflammatory pathways

ELISA GÓMEZ DE LOPE, ENRICO GLAAB ON BEHALF OF THE NCER-PD CONSORTIUM
Biomedical Data Science Group, **LCSB, University of Luxembourg**

elisa.gomezdelope@uni.lu
@elisagdelope

UNIVERSITÉ DU LUXEMBOURG

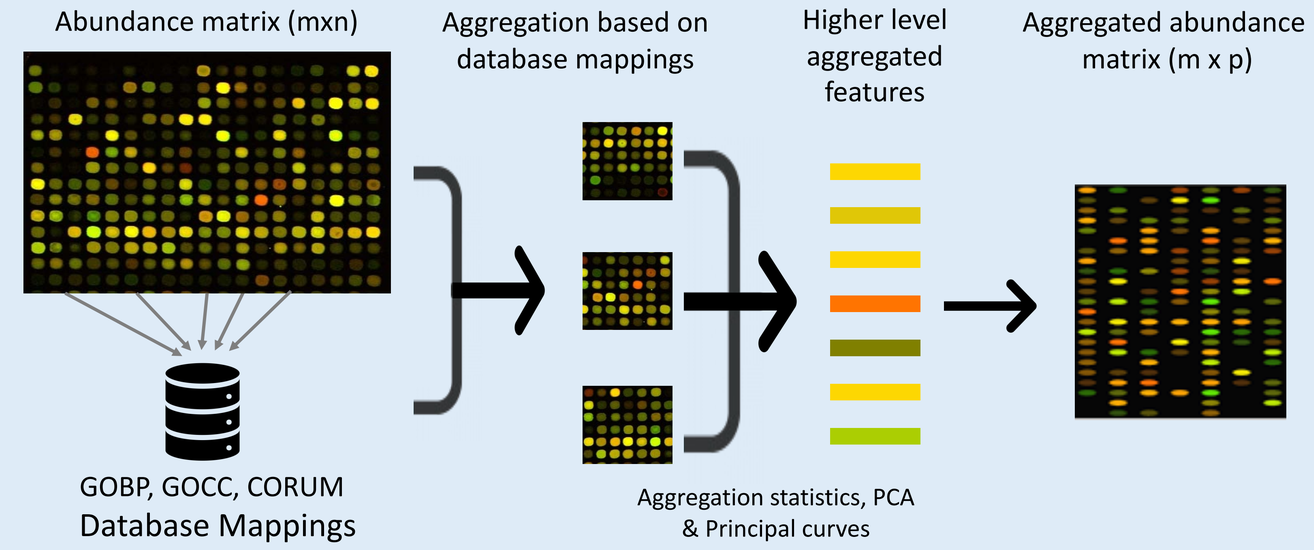
Background

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).

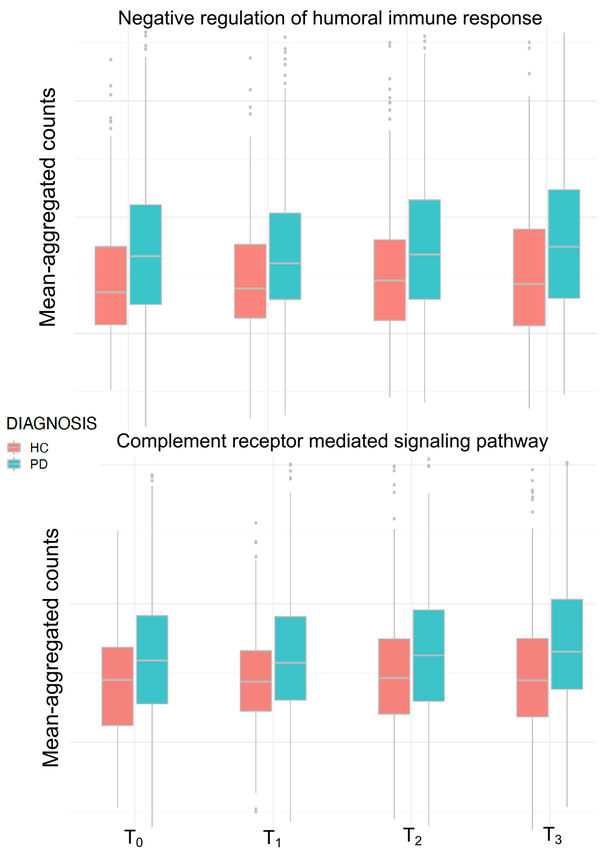
Differential expression analyses accounting for the confounders age and gender, and time correlation analyses (A) were applied to PD case/control data for both gene and functional level representations. Next, we evaluated ML models using a nested cross-validation including the feature selection and parameter 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

- Inflammation pathways displayed PD-specific positive correlation with time (A).
- Features reflecting the variance of aggregated expression at protein complex level (CORUM) provided higher cross-validated performance (AUC) than other types of aggregations and in comparison to the original gene features for PD vs. control sample classification. Logistic regression provided an area under the curve (AUC) of 0.67±0.06 in this setting (B).
- In these ML models, multiple pathways, cellular locations and protein complexes involved in (neuro)inflammation were included among the most relevant predictive features (C).
- A SHAP value analysis for the logistic regression model applied to CORUM-sd aggregated features revealed (neuro)inflammation, mitochondrial and chromatin modification complexes among the feature sets with the highest relevance. Furthermore, the SMN protein complex involved in the survival of motor neurons and neuron degeneration was included among the most predictive feature sets (D).

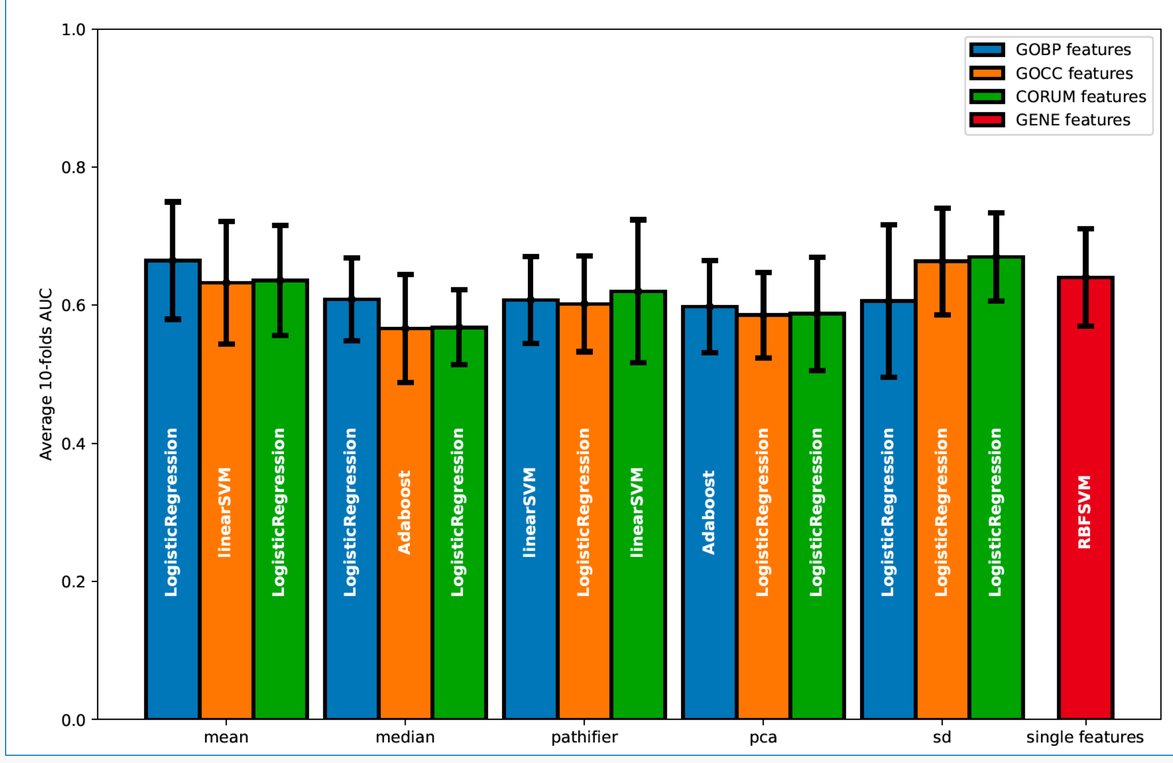
A. Longitudinal mean-aggregated counts at GO BP level



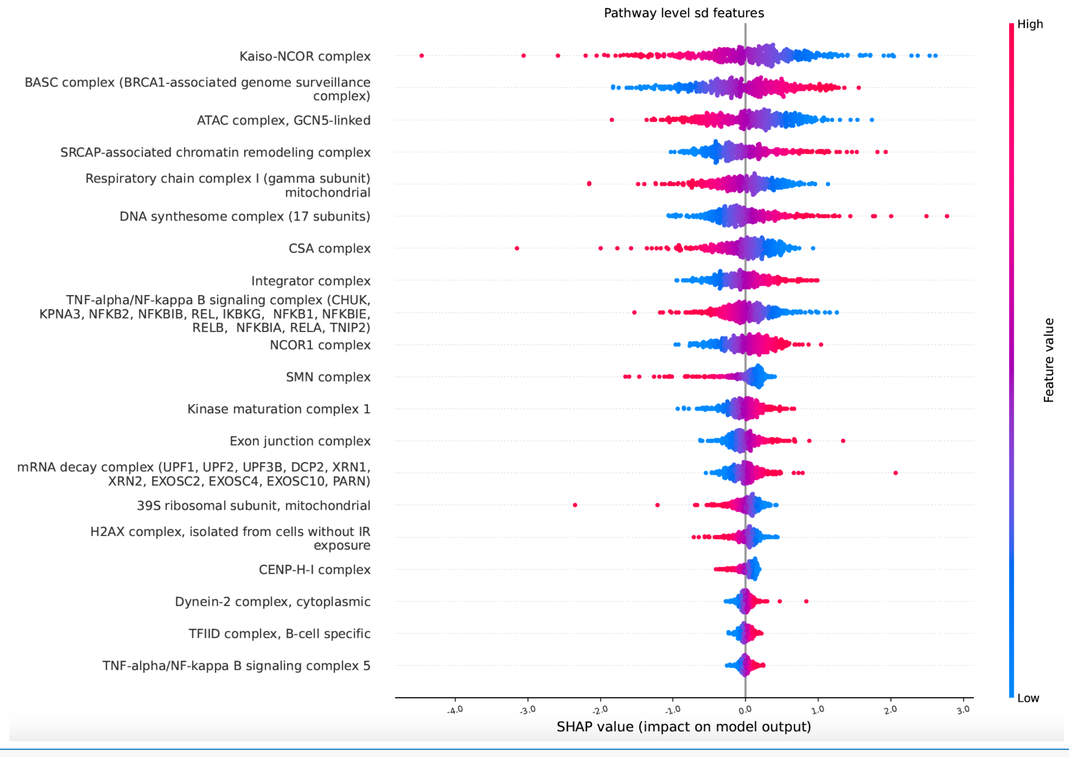
C. Aggregated functional representations associated with inflammation and neuro-inflammation among the top-20 most relevant features for corresponding ML models according to a SHAP value analysis of feature relevance.

	GO BP (Pathway level)	GOCC (Cellular compartment)	CORUM (Protein complex)
MEAN	Innate immune response in mucosa Negative regulation of intrinsic apoptotic signaling in response to DNA damage	Chitosome Postsynaptic endosome Platelet dense tubular network Platelet alpha granule membrane Endocytic vesicle lumen	TNF-alpha/NF-kappa B signaling complex (CHUK, KPNA3, NFKB2, NFKBIB, REL, IKBKG, NFKB1, NFKBIE, RELB, NFKBIA, RELA, TNIP2) HRD1 complex
PATHIFIER	Negative regulation of mitochondrial membrane permeability involved in apoptotic process Regulation of antigen processing and presentation of peptide antigen Granzyme mediated programmed cell death signaling Diapedesis Cellular response to UV-B Natural killer cell degranulation Apoptotic process in bone marrow cell Sulfation	Phagocytic vesicle membrane Autophagosome Cytolytic granule Late endosome lumen	Ubiquitin E3 ligase (DET1, DDB1, CUL4A, RBX1, COP1) ITGAV-ITGB3-NOV complex CDK1-CCNB1-CCNF complex SMN complex (GEMIN5,4,3), SMN-independent intermediate JUND-FOSB-SMAD3-SMAD4 complex
PCA	Defense response to fungus Response to interferon beta Complement activation alternative pathway Regulation of complement activation Eosinophil migration Cytolysis B cell receptor signaling pathway Innate immune response in mucosa Cellular response to nitrogen starvation	Mhc class ii protein complex Secondary lysosome Inflammasome complex Autolysosome	TNF-alpha/NF-kappa B signaling complex (CHUK, KPNA3, NFKB2, NFKBIB, REL, IKBKG, NFKB1, NFKBIE, RELB, NFKBIA, RELA, TNIP2) TRBP containing complex (DICER, RPL7A, EIF6, MOV10 and subunits of the 60S ribosomal particle)
SD	Innate immune response in mucosa Negative regulation of intrinsic apoptotic signaling in response to DNA damage	Chitosome Protein phosphatase type 2a complex Platelet dense tubular network Cul4 ring e3 ubiquitin ligase complex	TNF-alpha/NF-kappa B signaling complex (CHUK, KPNA3, NFKB2, NFKBIB, REL, IKBKG, NFKB1, NFKBIE, RELB, NFKBIA, RELA, TNIP2) TFIID complex, B-cell specific TNF-alpha/NF-kappa B signaling complex 5

B. 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