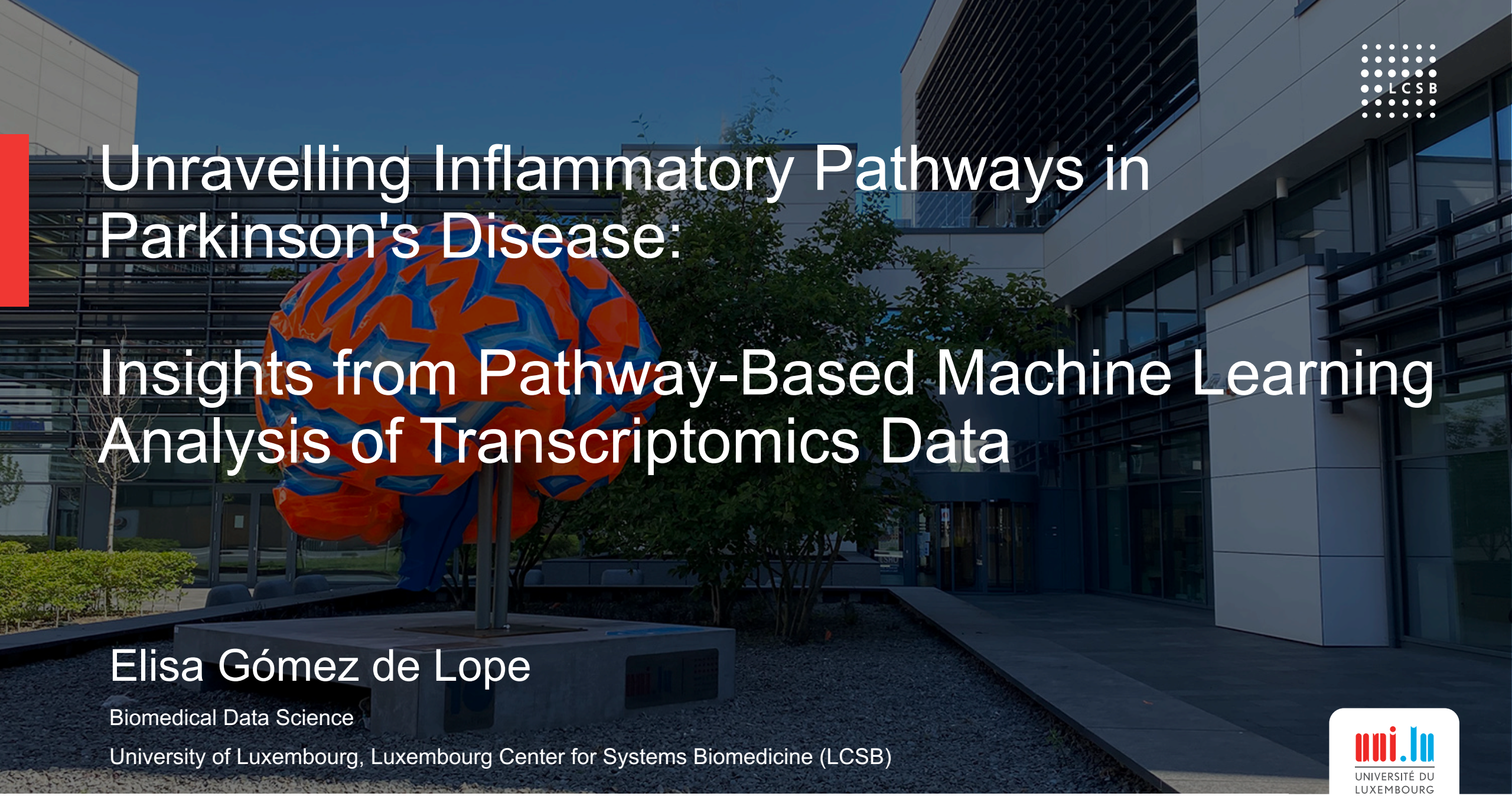


Unravelling Inflammatory Pathways in Parkinson's Disease:

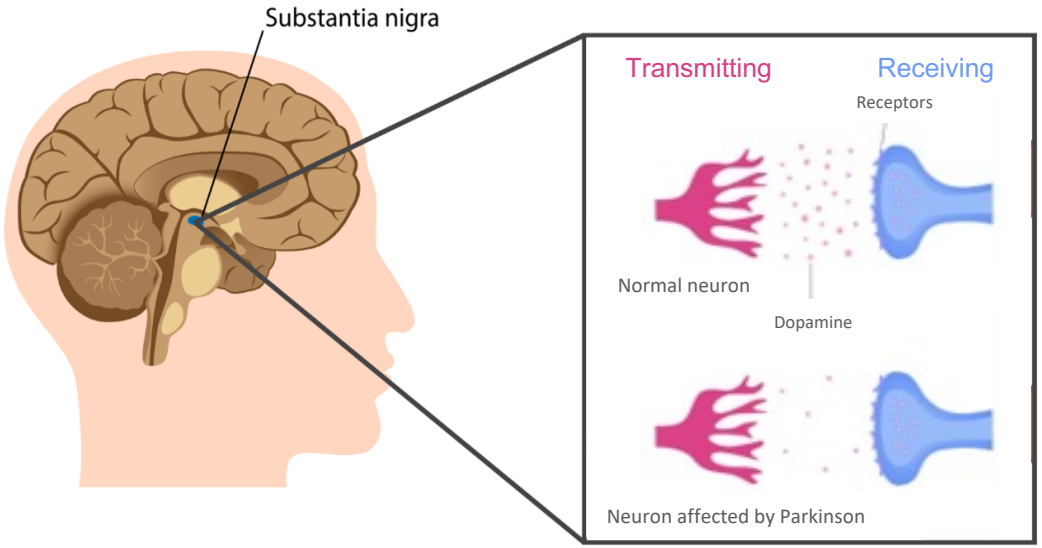
Insights from Pathway-Based Machine Learning Analysis of Transcriptomics Data

Elisa Gómez de Lope

Biomedical Data Science

University of Luxembourg, Luxembourg Center for Systems Biomedicine (LCSB)

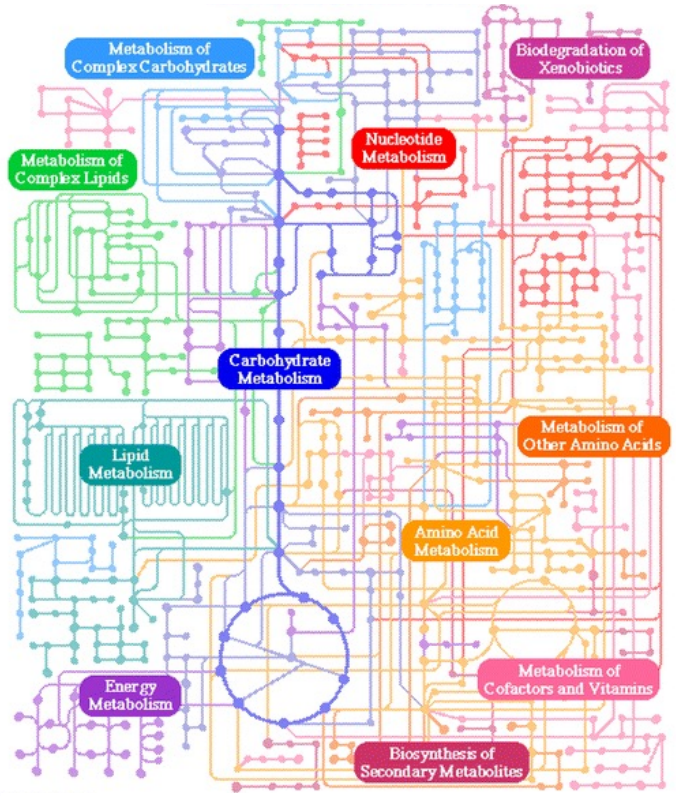
Parkinson's Disease & higher order functional representations



Source: Adapted from shutterstock

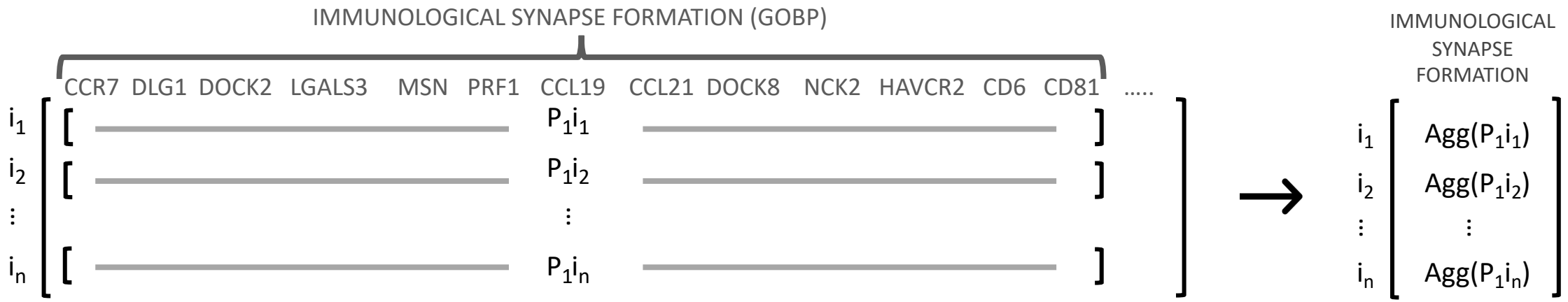
- Single gene mutations?
- Mitochondrial genetics?
- Environmental factors (toxins)?

Diagnosis is clinical & difficult

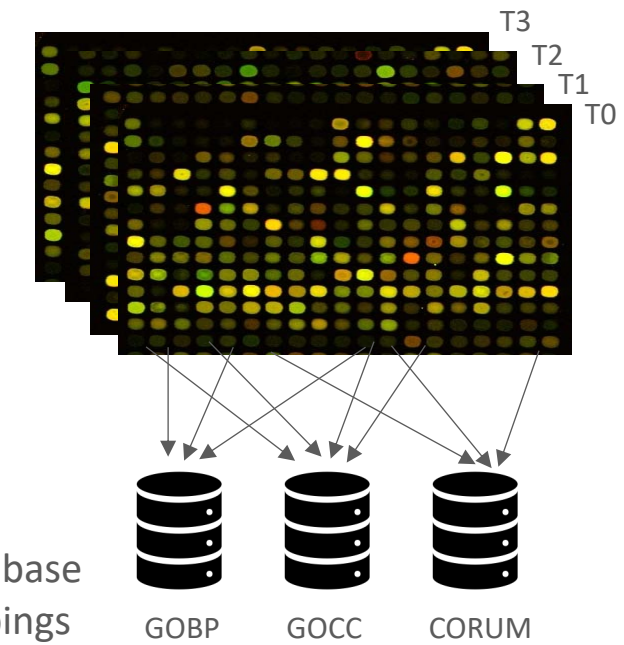


Schematic representation of metabolic networks
Source: The Origin and Evolution of Metabolic Pathways: Why and How did Primordial Cells Construct Metabolic Routes?

Generation of higher order functional features (I)

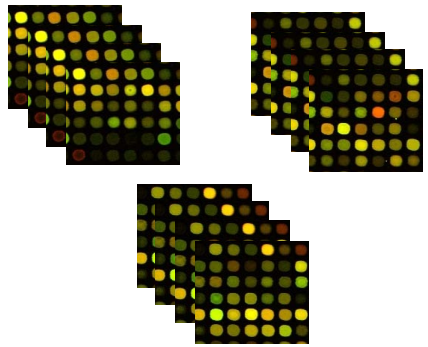


Temporal gene expression matrices (4xmxn)



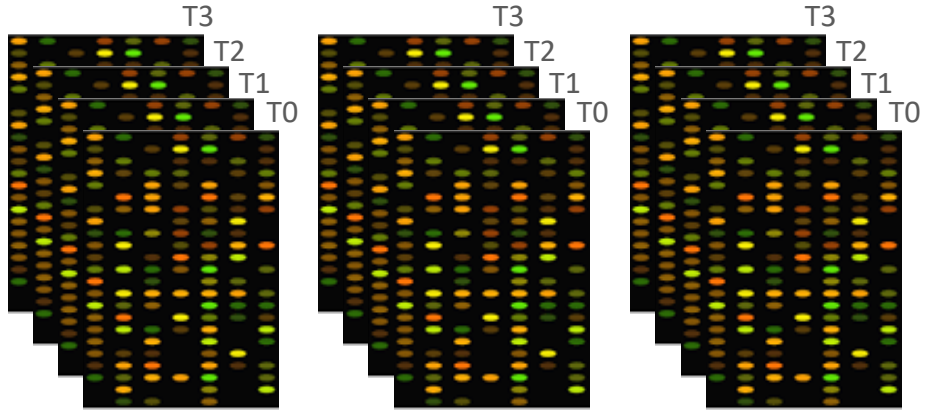
3 Database Mappings

Aggregation based on GOBP, GOCC & CORUM mappings



Aggregation statistics + PCA, Pathifier's dimension reduction-derived deregulation scores

3x Temporal aggregated expression matrices (4xmxp)



Longitudinal analyses

Transcriptomics aggregated temporal profiles

Feature selection strategies on transcriptomics data for a Random forest predictive model of PD diagnosis

			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

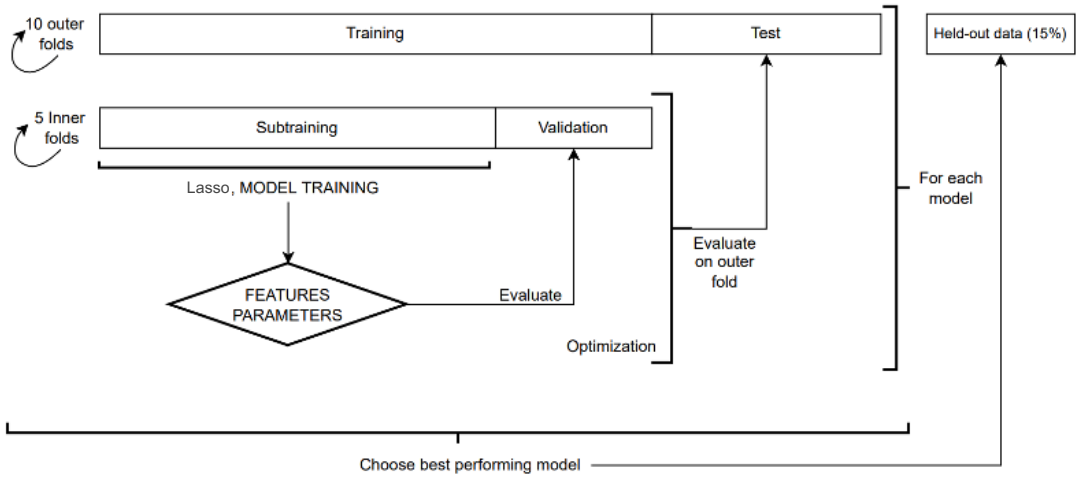
Table 4.1 Predictive performance for PD diagnosis measured as average cross-validated AUC ± standard deviation from gene expression data at gene and aggregated levels (mean and standard deviation) using DEA filter, Lasso penalty, a Pearson correlation filter (0.85) and no feature selection at all on a random forest model.

Lasso penalty obtained highest cross-validated AUC in a Random Forest model across most transcriptomics-based datasets.

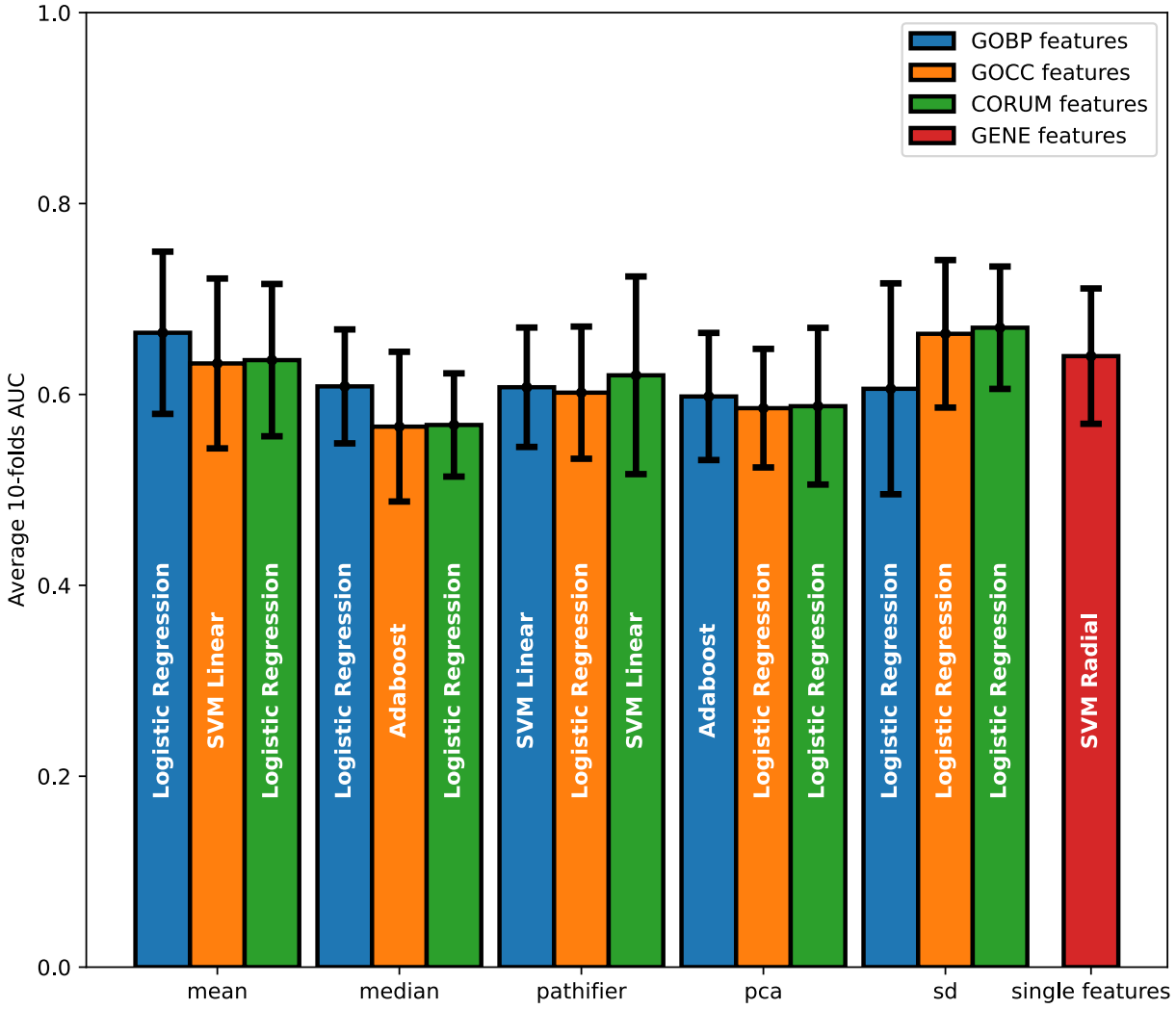
Predictive PD diagnosis with ML models on transcriptomics data

Models:
Logistic regression
Random Forest
SVM (Linear kernel)
SVM (Radial kernel)
Adaboost
Gradient boosting

Feature selection: Lasso penalty regression

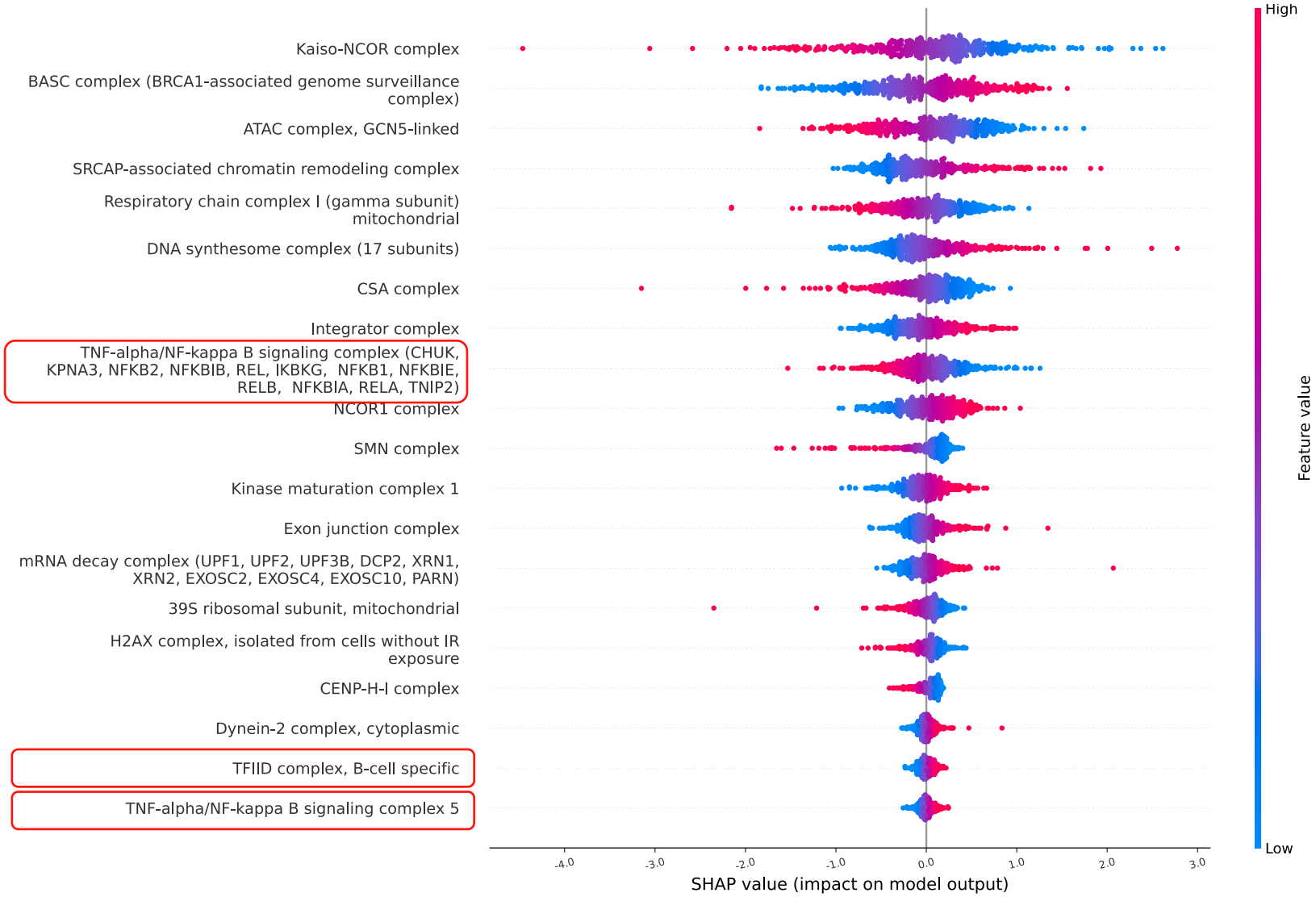


Crossvalidated AUC scores from best performing model on transcriptomics-based datasets



Relevant features from predictive PD diagnosis

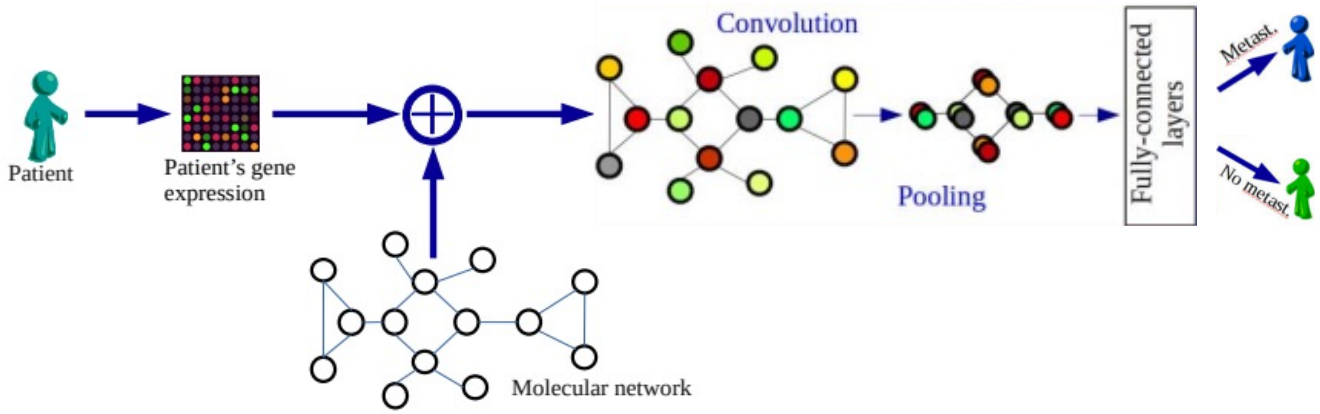
Shap values of (SD -aggregated) protein complexes (CORUM) predictors on Logistic Regression model



Limitations & outlook for future analyses

- ✗ Unknown confounders
- ✗ Large variability among PD patients makes identifying common trends difficult
- ✗ Data represents late stages of the disease

- ➡ Modelling other PD prognostic outcomes (e.g. motor dysfunction scores)
- ➡ Use a graph representation of the data via protein-protein interactions network



Gene expression profile as a graph signal of the molecular network

Source: Chereda, H., 2022. Explaining decisions of graph convolutional neural networks for analyses of molecular subnetworks in cancer [Doctoral thesis, Georg-August-Universität Göttingen]

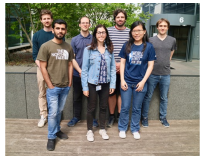
Unravelling Inflammatory Pathways in Parkinson's Disease:

Insights from Pathway-Based Machine Learning Analysis of Transcriptomics Data

Elisa Gómez de Lope

Biomedical Data Science

University of Luxembourg, Luxembourg Center for Systems Biomedicine (LCSB)



Enrico Glaab

A. Rauschenberger M. Ali R. Loo F. Nasta M. Ledda L.-C. Tranchevent Q. Klopfenstein