Combining PET imaging and blood metabolomics data to improve machine learning models for Parkinson's disease

Luxembourg Centre for Systems Biomedicine

Enrico Glaab[1], Andrea Greuel[2], Jean-Pierre Trezzi[1,3], Christian Jäger[1], Zdenka Hodak[1], Lars Timmermann[2], Marc Tittgemeyer[4], Nico J Diederich[5], Carsten Eggers[2]

1. Luxembourg Centre for Systems Biomedicine, Esch-sur-Alzette, Luxembourg; 2. Clinic for Neurology, University Hospital Giessen and Marburg, Marburg, Germany; 3. Integrated Biobank of Luxembourg, Luxembourg City, Luxembourg; 4. Max Planck Institute for Metabolism Research, Cologne, Germany; 5. Centre Hospitalier de Luxembourg, Luxembourg City, Luxembourg

1

Introduction

- Metabolomic profiling and PET neuroimaging capture a wide range of alterations in idiopathic Parkinson's disease (IPD), but have not been jointly used.
- Here, we investigate whether the **joint machine learning analysis of both blood metabolomics and PET imaging data**
  - may enhance diagnostic discrimination
  - provides new pathophysiological insights

2

Cohort & Methods

Cohort

- 60 IPD patients and 15 healthy age- and gender-matched controls (HC) (further characteristics see table)
- IPD patients had been 12 hours off levodopa and 72 hours off dopamine agonists

	IPD patients	Unaffected controls	P-value
N (female/male)	60 (19/41)	15 (8/7)	.14
Age	65.7 ± 9.0	65.1 ± 8.4	.831
UPDRS III	25.1 ± 9.7	2.1 ± 2.6	.000
H&Y stage	2.3 ± 0.4	-	-
BMI	26.8 ± 4.7	24.6 ± 4.1	.101

Methods

- **Metabolomics**
  - Gas chromatography coupled to mass spectrometry (GC-MS)
  - Determination of metabolomic profiles for blood plasma samples:
    - Baseline: entire cohort
    - Follow-up exam after 1 year: 18 IPD patients
- **Neuroimaging**
  Positron emission tomography (PET)
  - 3,4-dihydroxy-6-18F-fluoro-L-phenyl-alanine (**FDOPA):** 44 patients and 14 controls
  - 2-[fluorine-18]fluoro-2-deoxy-D-glucose (**FDG):** 51 patients and 15 controls)
- **Feature selection**
  To compare IPD & control subjects a supervised attribute selection was applied to:
  - pre-processed FDOPA and FDG PET imaging data (empirical Bayes moderated t-statistic [1])
  - GC-MS metabolomics data (Welch's Test)
- **Machine learning**
  Two types of analysis applied to the individual and combined standardized datasets
  - support vector machine [2]
  - random forest [3]

  For validation, both a 50% random training/test set split and a leave-one-out cross-validation was used.

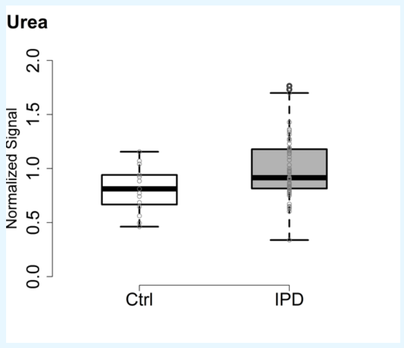
3

Results

1. Metabolomics analyses

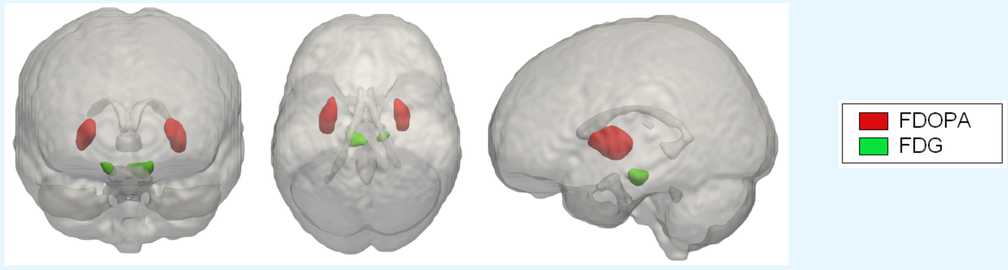
- 1 unknown metabolite (RI 1446) with higher abundance in PD (FDR < 0.05),
- 4 metabolites with indicative changes (p < 0.05, FDR < 0.5)
- Urea = top-ranked known metabolite (see box plot).

Metabolite	Fold-change	P-value	FDR
RI 1446 (unknown)	1.270	0.001	0.039
Urea	1.262	0.005	0.140
RI 1050 (unknown)	1.324	0.006	0.140
Hexadecanoic acid	1.256	0.030	0.371
Dodecanoic acid	1.403	0.033	0.371



2. PET imaging analyses

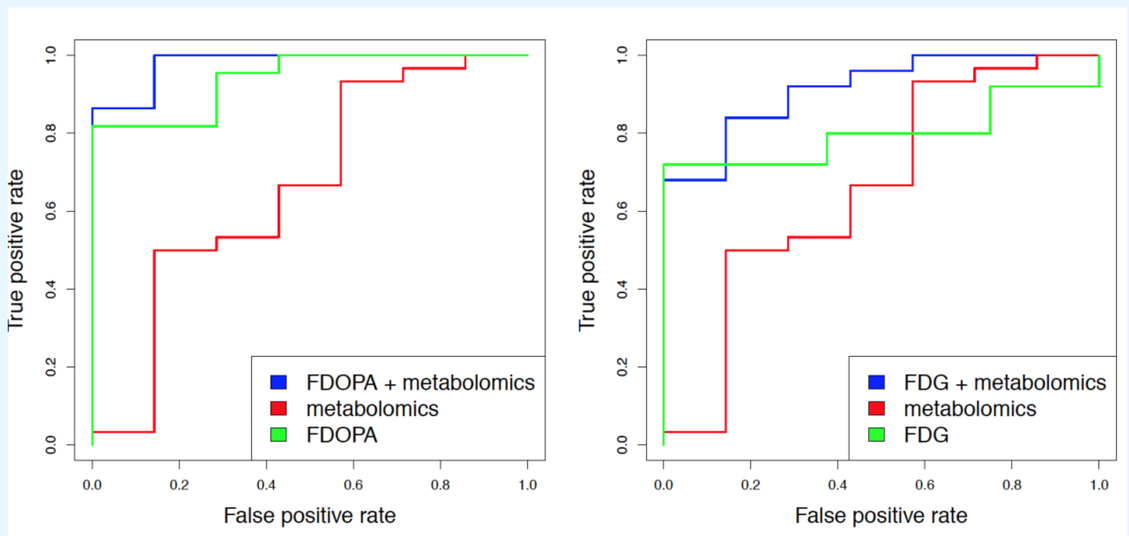
- FDOPA: significant differences in 2 voxel clusters in posterior part of the putamen (FDR < 0.05, see red areas in the figure below).
- FDG: significant differences in cerebral peduncles of the lower midbrain (including SNc, see green areas in the figure below).



3. Combining both: Discrimination improved

Analysis of area under the Receiver Operating Characteristic curve (AUROC):

- FDOPA: higher predictive performances (left fig.) compared to FDG (right fig.)
- Combining PET features with metabolomics data → highest predictive performance
  - FDOPA + metabolomics: 0.98 AUROC
  - FDG + metabolomics: 0.91 AUROC.



4

Conclusions

Combining blood metabolomics and brain PET imaging data provides a unique setting and means to investigate IPD-associated changes. Our results suggest:

- Combined analysis increases the predictive performance of sample classification models
- Metabolite alterations point to interesting changes of oxidative stress & inflammation pathways
- Most stable and population-independent predictive features still need confirmation in larger and independent cohorts

References

[1] Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, 3(1), 1-25

[2] Vapnik V (1995) Support vector machine. *Mach Learn* 20:273–297

[3] Breiman L (1999) Random forest. *Mach Learn* 45:1–35

UKGM Universitätsklinikum Giessen und Marburg · IBBL · CHL Centre Hospitalier de Luxembourg · Max Planck Institute for Metabolism Research · Fonds National de la Recherche Luxembourg