Machine learning analysis of metabolomics and neuro-imaging data for Parkinson's disease

Enrico Glaab
Luxembourg Centre for Systems Biomedicine

Outline of analyses

1) Cohort overview

2) Neuroimaging data analysis (FDG and F-DOPA PET)

3) Metabolomics analysis (cross-sectional & longitudinal)

4) Machine learning analyses:
   a) FDOPA PET
   b) FDG PET
   c) FDOPA PET + metabolomics
   d) FDG PET + metabolomics
   e) Evaluation (Cross-validation, ROC curves)

Joint analysis of PD neuroimaging & metabolomics data

Motivation:
Metabolomic and PET neuroimaging data capture a wide range of alterations in PD, but synergies of combined analysis not explored

→ **Goal**: Investigate whether the **joint machine learning analysis of blood metabolomics data and PET imaging data**

- may improve discrimination between patients & controls
- provide new pathophysiological insights

Cohort overview

- 60 PD patients and 15 healthy age- and gender-matched controls (University Hospitals Cologne, Giessen and Marburg; Prof. C. Eggers)
- **Medication**: PD patients had been 12 hours off levodopa and 72 hours off dopamine agonists

	PD patients	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

UNIVERSITÉ DU LUXEMBOURG

LCSB

Analyses overview

- **Metabolomics**
  - Gas chromatography coupled to mass spectrometry (GC-MS)
  - Determination of metabolomic profiles for blood plasma samples:
    - Baseline: 60 patients and 15 controls
    - Follow-up exam after 1 year: 18 patients

- **Neuroimaging**

  Positron emission tomography (PET):
    - 3,4-dihydroxy-6-18F-fluoro-L-phenyl-alanine (**FDOPA)**

      44 patients and 14 controls

      → How does dopamine metabolism change?
    - 2-[fluorine-18]fluoro-2-deoxy-D-glucose (**FDG)**

      51 patients and 16 controls

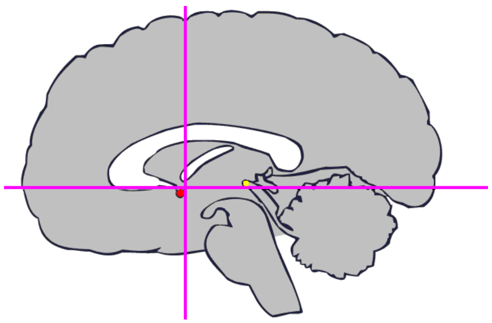
      → How does glucose metabolism change?

UNIVERSITÉ DU LUXEMBOURG

LCSB

PET imaging data pre-processing

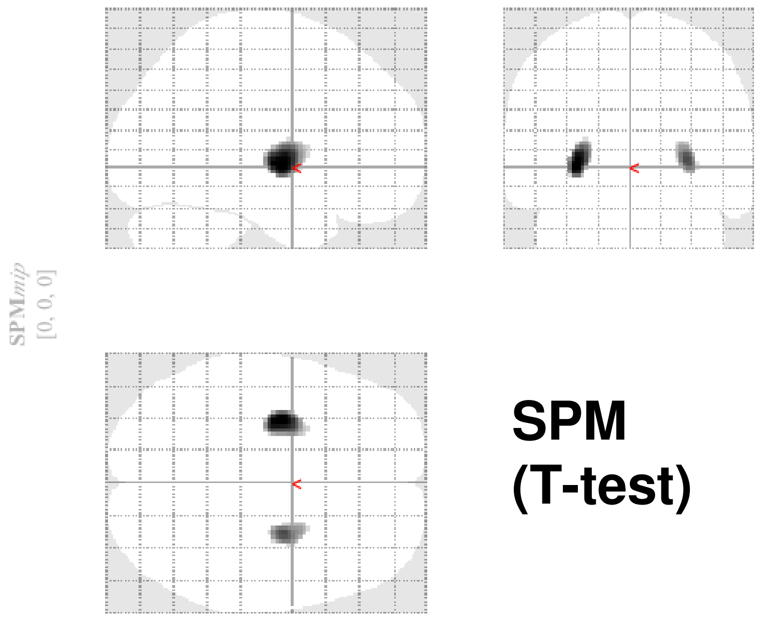
- All pre-processing steps performed in SPM12 (Matlab)
- Co-registration of each subject's averaged FDG and FDOPA images
- Centering on the anterior commissure and horizontal alignment



- Spatial normalization to Montreal Neurological Institute standard space (MNI152) performed using tracer-specific templates
- Spatial smoothing (Gaussian kernel, 5 mm FWHM)

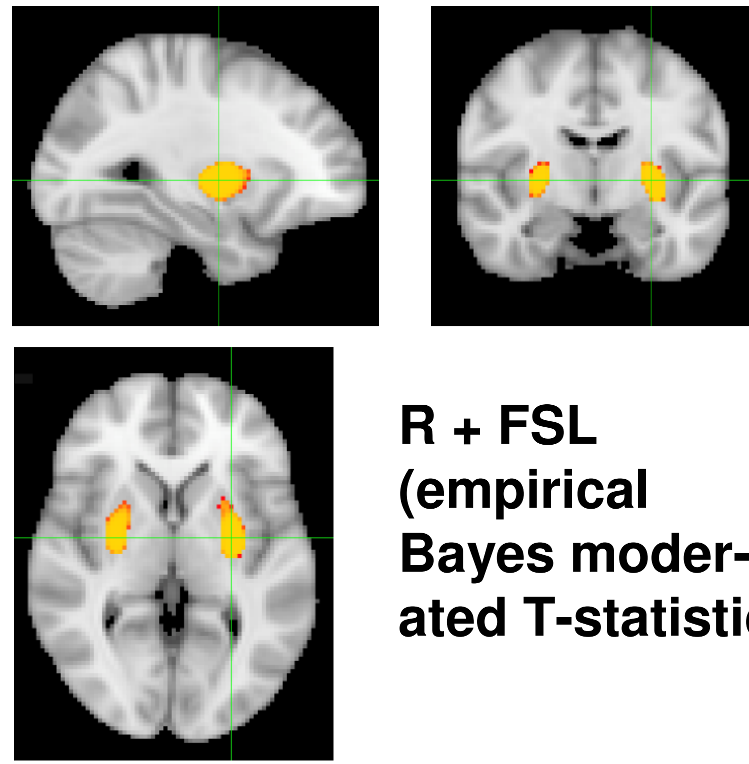
FDOPA PET analyses after global mean normalization

FDOPA PET: Significant changes in putamen/striatum (FDR < 0.05)



SPM (T-test)

SPMresults: ./PD_imaging/FDopa_controls
Height threshold T = 5.112656 {p<0.05 (FWE)}
Extent threshold k = 0 voxels



R + FSL (empirical Bayes moderated T-statistic)

Best FDR < 1E-3

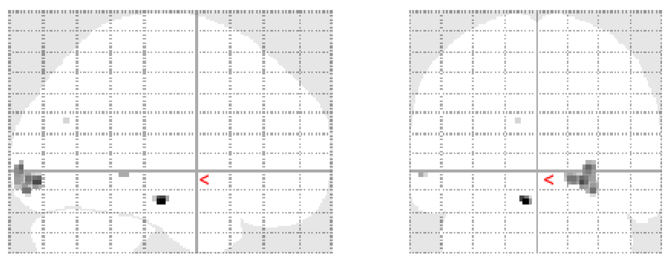
Best FDR < 1.18E-5

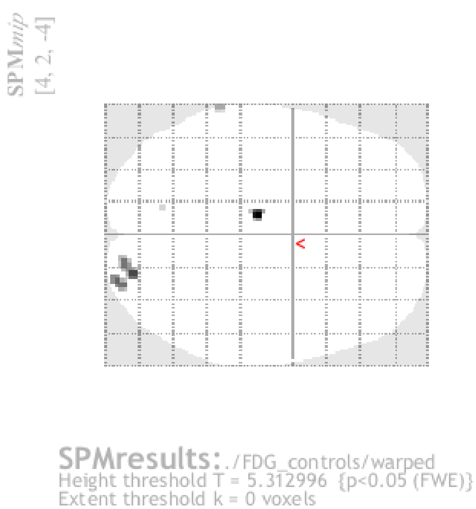
UNIVERSITÉ DU LUXEMBOURG

LCSB

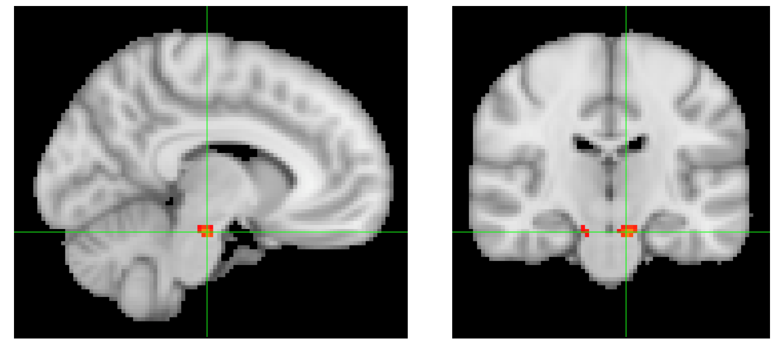
FDG PET analyses after global mean normalization

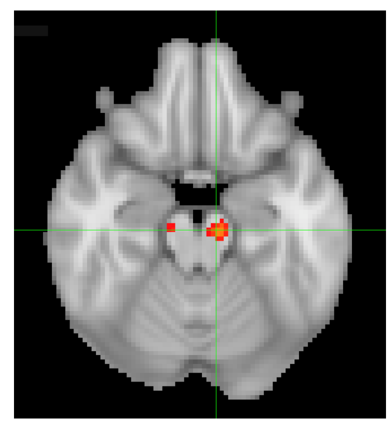
FDG PET: Significant changes in midbrain (FDR < 0.05)





SPM (T-test)





R + FSL (empirical Bayes moderated T-statistic)

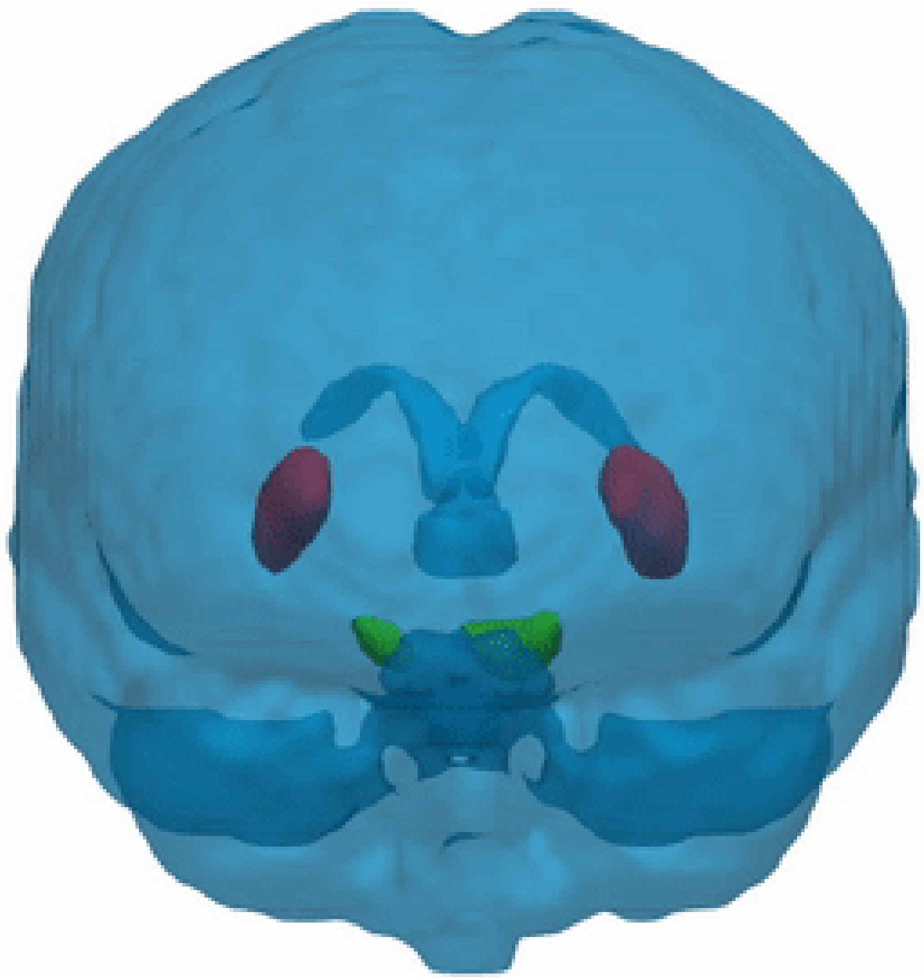
Best FDR = 0.026 (FWE < 1E-3)

Best FDR = 0.009

UNIVERSITÉ DU LUXEMBOURG

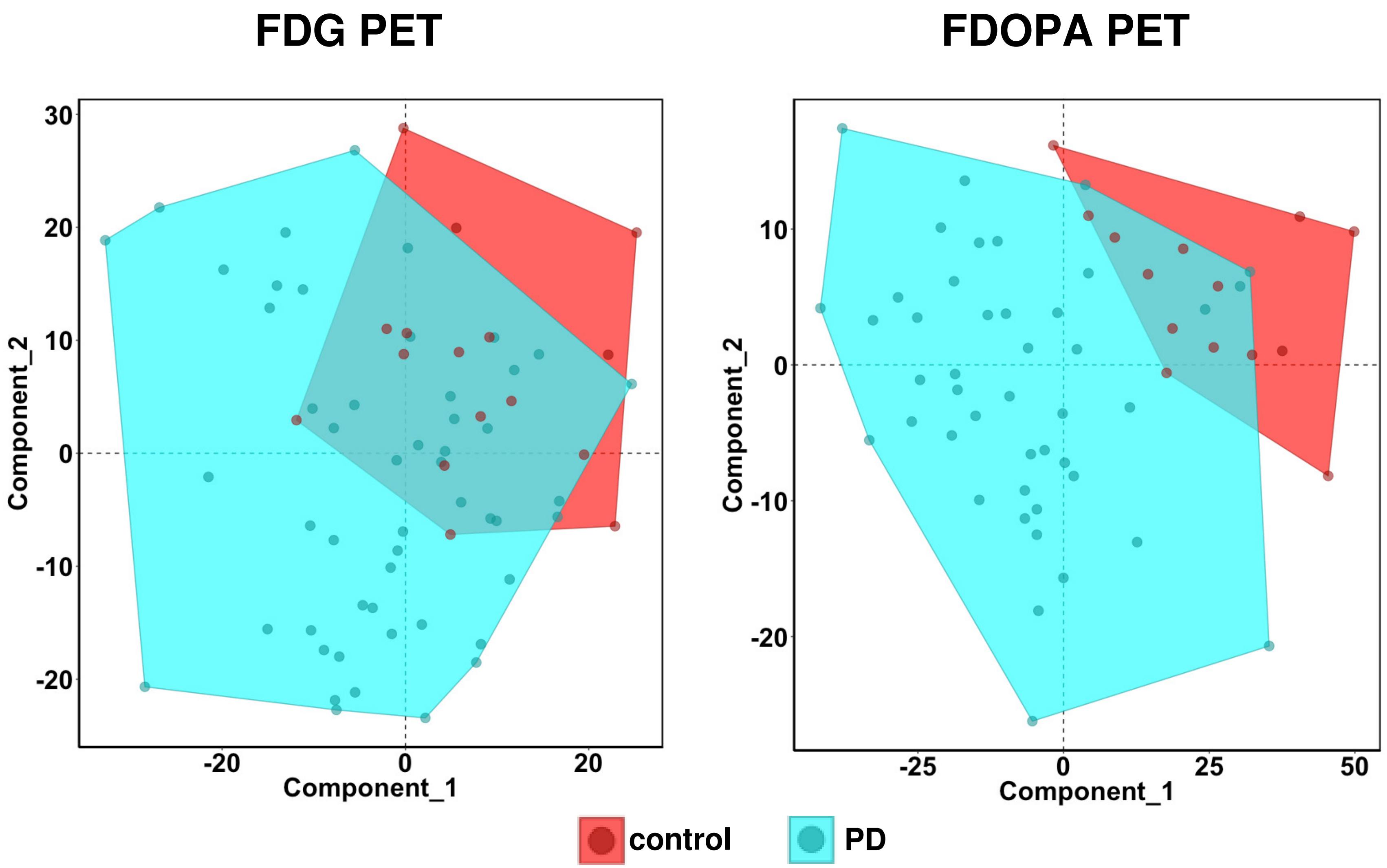
LCSB

FDOPA PET – Visualization of significant clusters



Most significant voxel clusters (eBayes, FDR < 0.05):
- FDOPA
- FDG

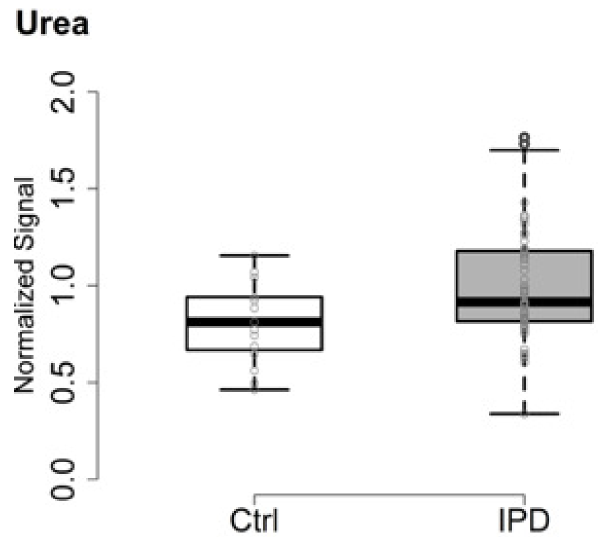
Partial Least Squares Discriminant Analysis



Metabolomics analyses (Baseline, Welch's T-test)

- 1 unknown metabolite (RI 1446) with higher abundance in PD ($FDR < 0.05$) and correlated with top-ranked FDOPA voxel ($p = 0.04$)
- Urea = top-ranked known metabolite → marker of oxidative stress, but $FDR > 0.05$ (see box plot; Zhang et al., 1999)

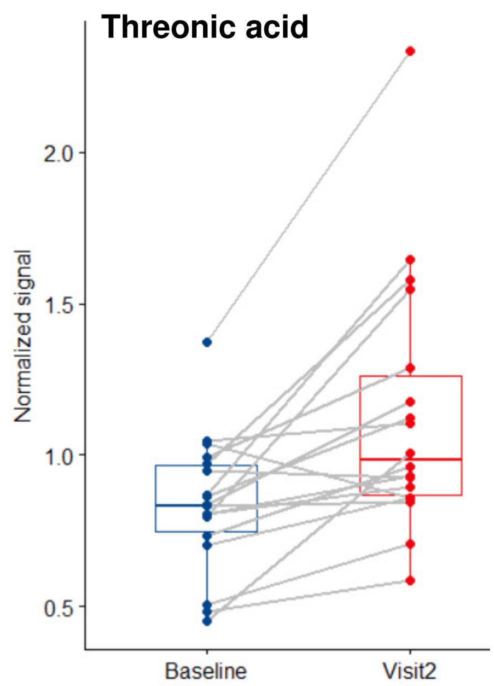
Metabolite	Fold-change	P-value	FDR
RI 1446 (unknown)	1.270	0.001	0.039
Urea	1.262	0.005	0.140
RI 1050 (unknown)	1.324	0.006	0.140
Hexadecanoic acid	1.256	0.030	0.371
Dodecanoic acid	1.403	0.033	0.371



Metabolomics analyses (Longitudinal: Visit 2 vs. Visit 1)

- Threonic and glycolic acid are top-ranked, but FDR > 0.05 (DeNoPa: threonic acid increased in PD vs. controls, FDR = 0.004)
- Most top-ranked metabolites tend to have higher abundance in PD

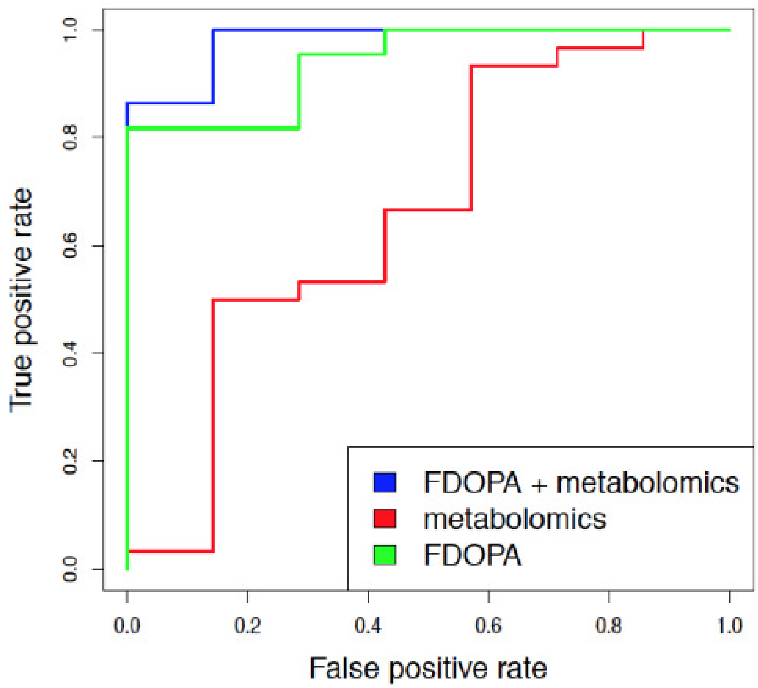
Metabolite	FC	*P*	FDR
Threonic acid	1.353	0.001	0.059
Glycolic acid	1.258	0.002	0.059
Iminodiacetic acid	1.154	0.007	0.140
Glycerol	0.642	0.008	0.140
Succinic acid	1.161	0.029	0.317
Mannose	1.148	0.030	0.317
Glyceric acid	1.229	0.031	0.317
Citric acid	1.144	0.046	0.375
RI 1708	0.876	0.048	0.375



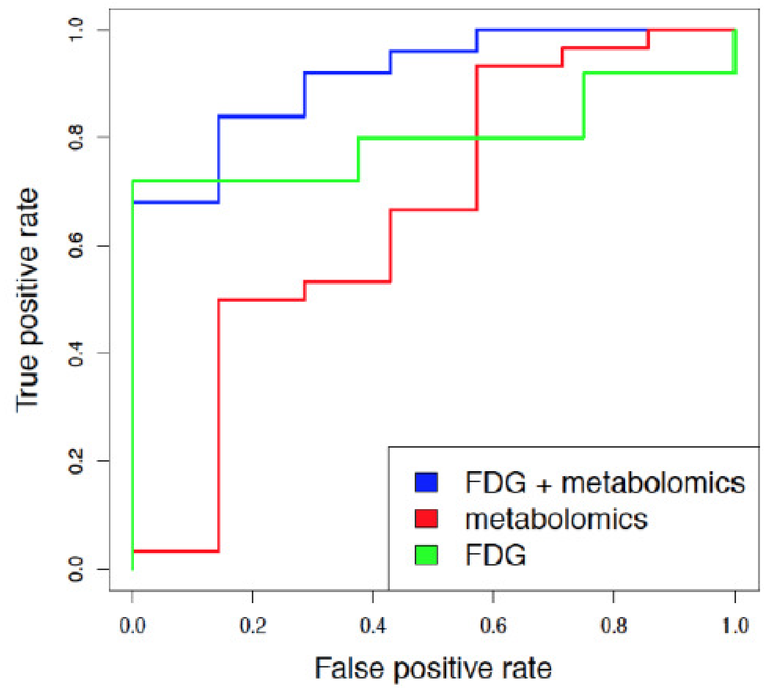
Machine Learning – ROC Curve Analyses

Combine standardized attributes from **FDOPA** and **FDG** PET data with **metabolomics data** → evaluate integrated machine learning models (SVMs)

ROC curves (FDOPA)



ROC curves (FDG)



Summary

- **PET analyses**: Significant changes in both FDOPA (putamen/striatum) and FDG (midbrain) analysis

- **Metabolomics analyses**: Few significant changes; top-ranked metabolites (urea, threonic acid and glycolic acid) have associations with oxidative stress and mitochondrial dysfunction

- **Integrated machine learning**: Combination of standardized PET + metabolomics features provides higher predictive performance than PET or metabolomics only

UNIVERSITÉ DU LUXEMBOURG

LCSB

References

1. E. Glaab, J.P. Trezzi, A. Greuel, C. Jäger, Z. Hodak, A. Drzezga, L. Timmermann, M. Tittgemeyer, N. J. Diederich, C. Eggers, Integrative analysis of blood metabolomics and PET brain neuroimaging data for Parkinson's disease, Neurobiology of Disease (2019), Vol. 124, No. 1, pp. 555
2. E. Glaab, *Using prior knowledge from cellular pathways and molecular networks for diagnostic specimen classification*, Briefings in Bioinformatics (2015), 17(3), pp. 440
3. E. Glaab, R. Schneider, *Comparative pathway and network analysis of brain transcriptome changes during adult aging and in Parkinson's disease*, Neurobiology of Disease (2015), 74, 1-13
4. N. Vlassis, E. Glaab, *GenePEN: analysis of network activity alterations in complex diseases via the pairwise elastic net*, Statistical Applications in Genetics and Molecular Biology (2015), 14(2), 221
5. S. Köglsberger, M. L. Cordero-Maldonado, P. Antony, J. I. Forster, P. Garcia, M. Buttini, A. Crawford, E. Glaab, *Gender-specific expression of ubiquitin-specific peptidase 9 modulates tau expression and phosphorylation: possible implications for tauopathies*, Molecular Neurobiology (2016), in press (doi: 10.1007/s12035-016-0299-z)
6. L. Grandbarbe, S. Gabel, E. Koncina, G. Dorban, T. Heurtaux, C. Birck, E. Glaab, A. Michelucci, P. Heuschling, *Inflammation promotes a conversion of astrocytes into neural progenitor cells via NF-kB activation*, Molecular Neurobiology (2016), Vol. 53, No. 8, 5041-5055
7. S. Kleiderman, J. Sá, A. Teixeira, C. Brito, S. Gutbier, L. Evje, M. Hadera, E. Glaab, M. Henry, S. Agapios, P. Alves, U. Sonnewald, M. Leist, *Functional and phenotypic differences of pure populations of stem cell-derived astrocytes and neuronal precursor cells*, Glia (2016), Vol. 64, No. 5, 695-715
8. E. Glaab, R. Schneider, *RepExplore: Addressing technical replicate variance in proteomics and metabolomics data analysis*, Bioinformatics (2015), 31(13), pp. 2235
9. E. Glaab, *Building a virtual ligand screening pipeline using free software: a survey*, Briefings in Bioinformatics (2015), 17(2), pp. 352
10. E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, A. Valencia. *EnrichNet: network-based gene set enrichment analysis*, Bioinformatics, 28(18):i451-i457, 2012
11. E. Glaab, R. Schneider, *PathVar: analysis of gene and protein expression variance in cellular pathways using microarray data*, Bioinformatics, 28(3):446-447, 2012
12. E. Glaab, J. Bacardit, J. M. Garibaldi, N. Krasnogor, *Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data*, PLoS ONE, 7(7):e39932, 2012
13. E. Glaab, A. Baudot, N. Krasnogor, A. Valencia. *TopoGSA: network topological gene set analysis*, Bioinformatics, 26(9):1271-1272, 2010
14. E. Glaab, A. Baudot, N. Krasnogor, A. Valencia. *Extending pathways and processes using molecular interaction networks to analyse cancer genome data*, BMC Bioinformatics, 11(1):597, 2010
15. E. Glaab, J. M. Garibaldi and N. Krasnogor. *ArrayMining: a modular web-application for microarray analysis combining ensemble and consensus methods with cross-study normalization*, BMC Bioinformatics,10:358, 2009
16. E. Glaab, J. M. Garibaldi, N. Krasnogor. *Learning pathway-based decision rules to classify microarray cancer samples*, German Conference on Bioinformatics 2010, Lecture Notes in Informatics (LNI), 173, 123-134
17. E. Glaab, J. M. Garibaldi and N. Krasnogor. *VRMLGen: An R-package for 3D Data Visualization on the Web*, Journal of Statistical Software, 36(8),1-18,2010
18. C. Jaeger, E. Glaab, A. Michelucci, T. M. Binz, S. Koeglsberger, P. Garcia, J. P. Trezzi, J. Ghelfi, R. Balling, M. Buttini, *The Mouse Brain Metabolome: Region-Specific Signatures and Response to Excitotoxic Neuronal Injury*, American Journal of Pathology (2015), Vol. 185, No. 6, pp. 1699