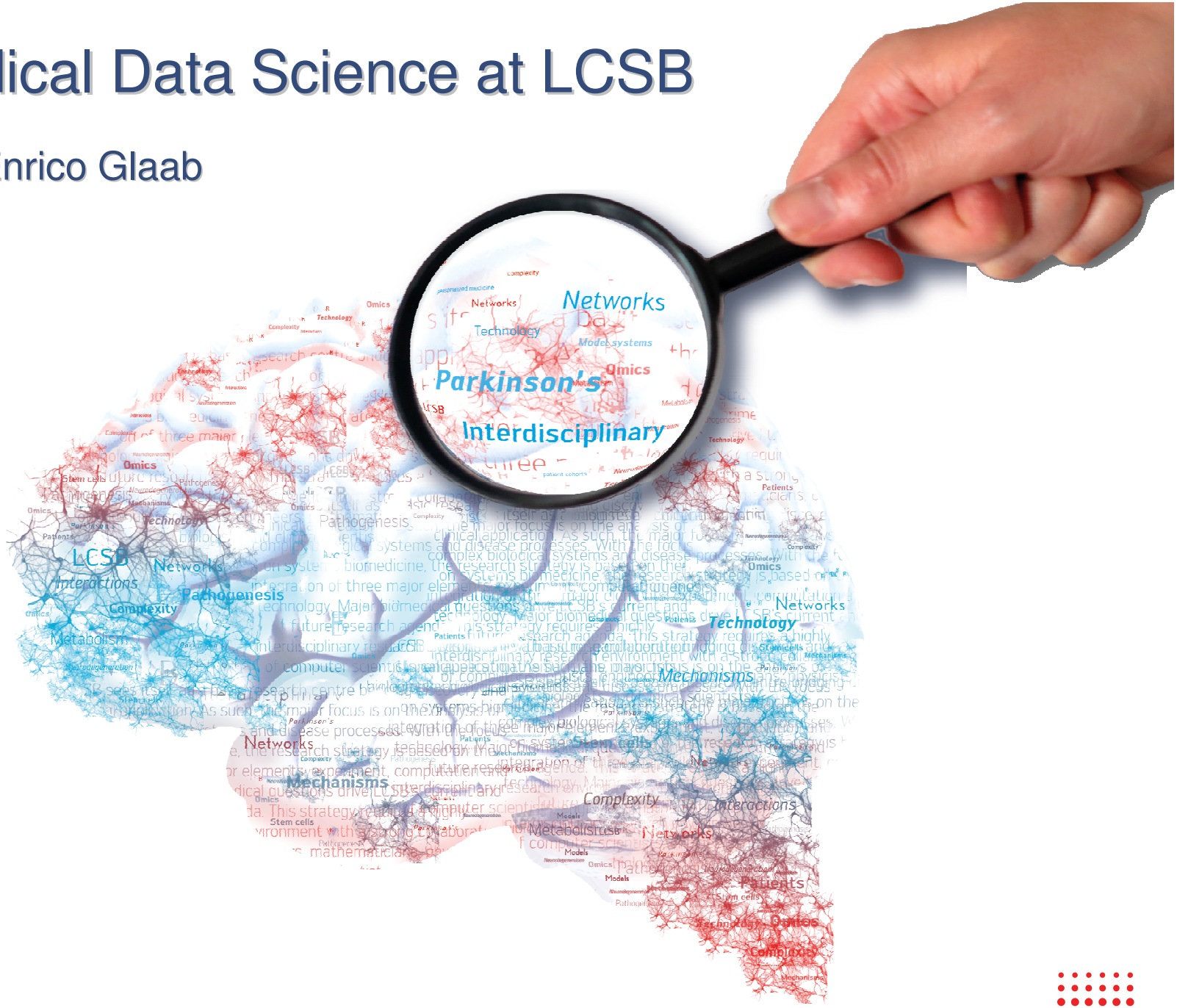
Biomedical Data Science at LCSB

Speaker: Enrico Glaab

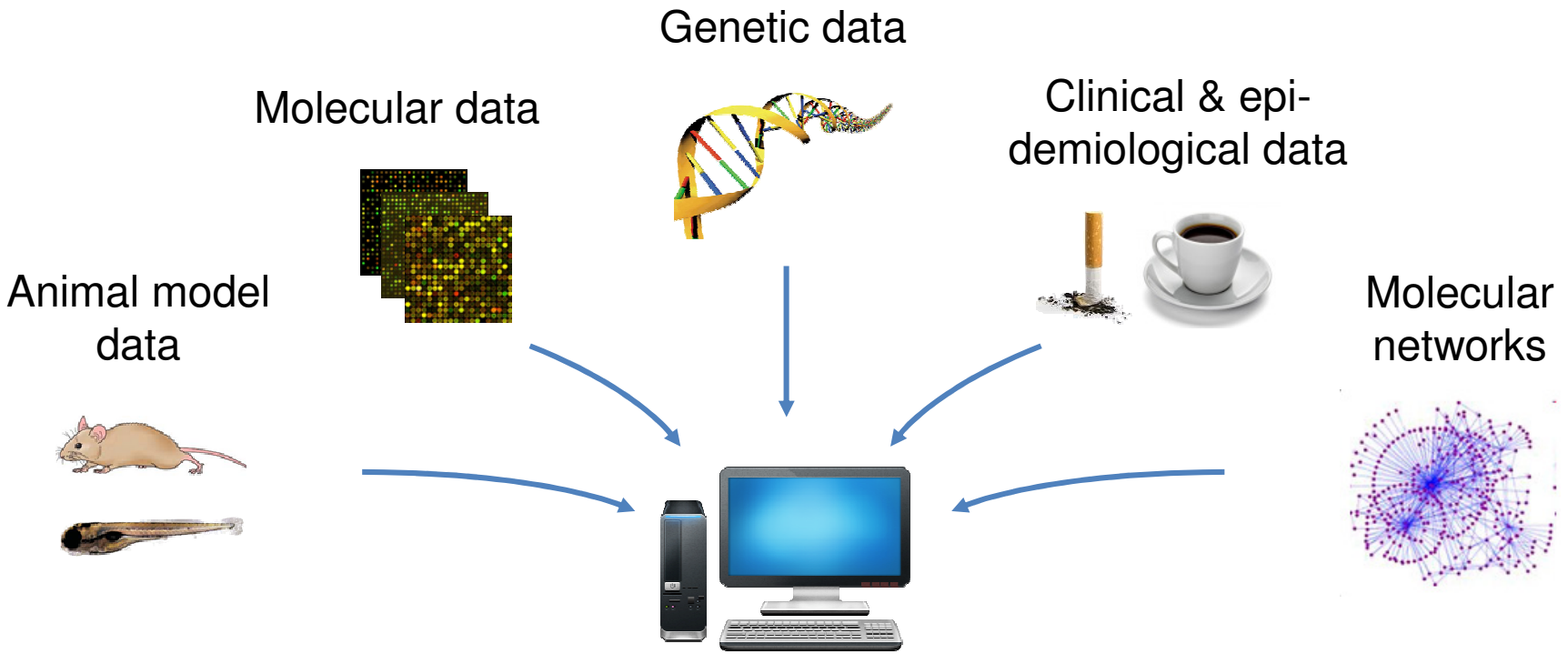


UNIVERSITÉ DU LUXEMBOURG

LCSB

LCSB Biomedical Data Science Group

Main research goal: Interpret molecular changes in complex diseases by exploiting prior biological knowledge and diverse sources of experimental data via integrative statistical analyses



Bioinformatics Services and Research Areas

- **Data integration and management**: Organize, store and categorize large amounts of data (PetaByte scale). Providing access and management for large compute farms

- **Automated pipelines for large-scale statistical data analysis:** Setup and administration of reproducible workflows to extract relevant information from heterogeneous data (R3 initiative: ensure **r**eproducible **r**esearch **r**esults)

- **Network and pathway analysis:** Develop and apply software tools that exploit existing data on biomolecular interactions to investigate perturbations of cellular signalling pathways in complex diseases

- **Text-mining and visualization:** Build 2D/3D visualizations tools and large-scale literature mining methods for data exploration and hypothesis generation

Hardware resources: High-performance computing servers

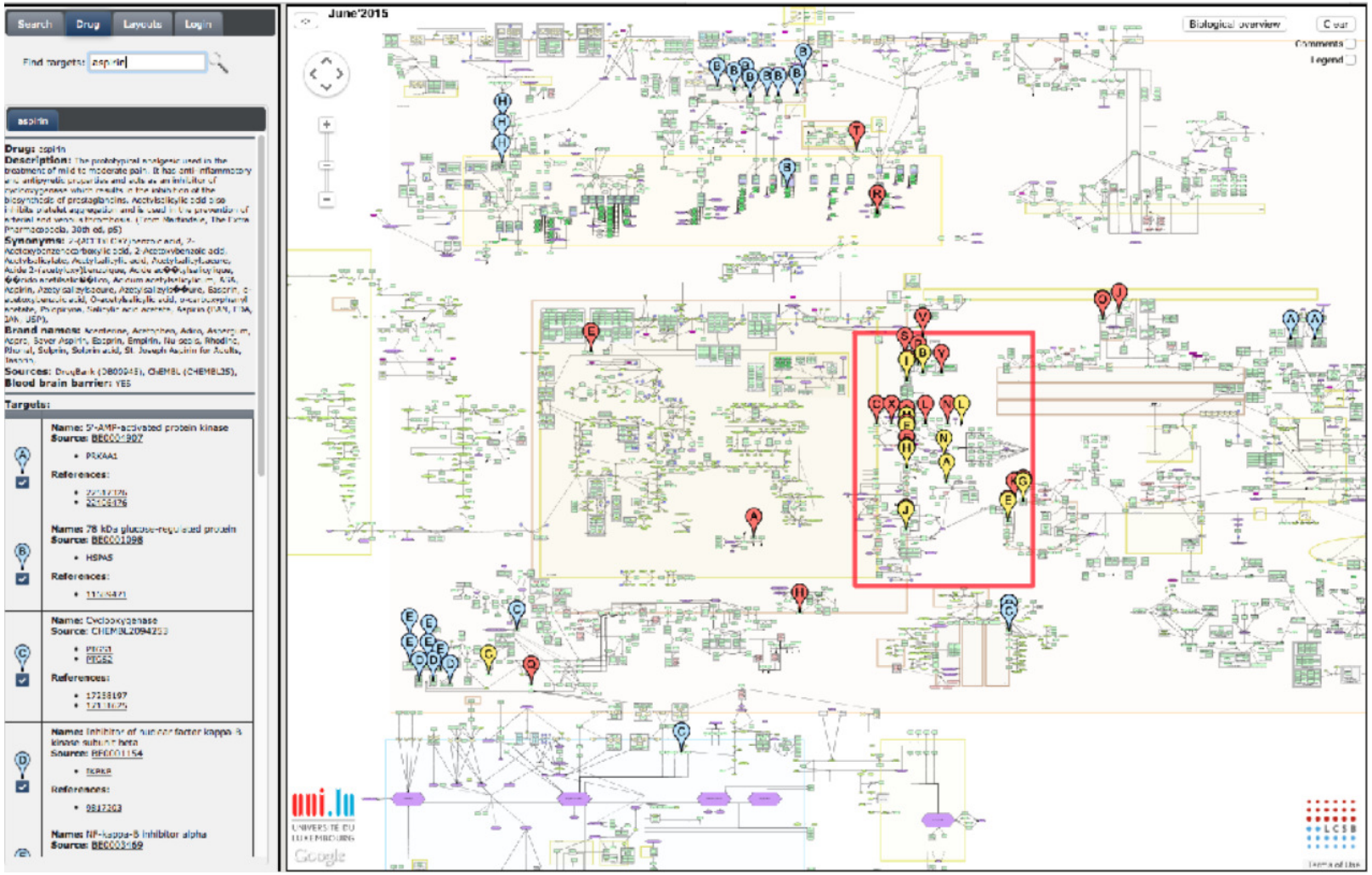
- **Kirchberg Campus (2009)**: 14 racks, 1120 cores, 180 TB storage + 180 TB backup



- **Belval Campus (2011)**: 14 racks, 3440 cores, 3564 TB storage + 1336 TB backup


Software resources: The Parkinson‘s disease map

→ An interactive, literature-curated map for data integration & exploration



UNIVERSITÉ DU LUXEMBOURG

LCSB

Developed network analysis tools

- **EnrichNet**, a web-service to score the associations of gene/protein lists with cellular pathways using the graph structure of molecular networks (Bioinformatics 2012)

- **TopoGSA**, a web-application for network topological analysis of gene/protein lists to find candidate disease genes with outstanding network properties (Bioinformatics, 2010)

- **PathExpand**, a web-application to extend disease-related pathways via graph-theoretical analyses on molecular networks to identify new candidate disease genes (BMC Bioinformatics, 2010)

Developed machine learning tools for molecular data

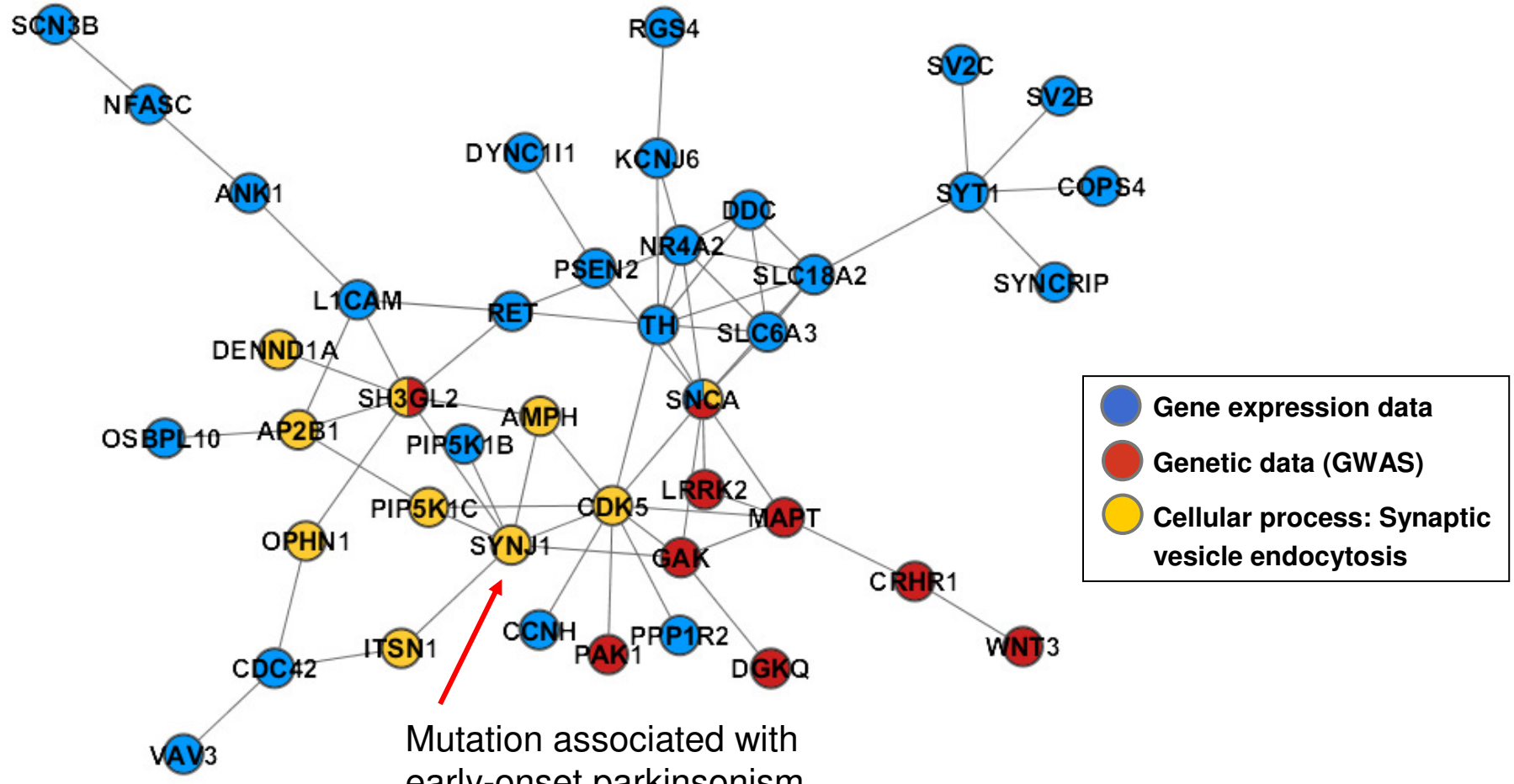
- **PathVar**, a web-application for sample clustering and diagnostic classification using pathway information (Bioinformatics, 2012)

- **GenePEN**, a machine learning approach to identify coordinated alterations of genes or proteins in biological networks using omics data (Stat. Appl. Genet. Mol. Biol., 2015)

- **ArrayMining.net**, a server for automated analysis of gene expression data, including Feature selection, Clustering, Prediction, Co-Expression & Pathway Analysis (BMC Bioinformatics, 2009)

Diseases as network perturbations (EnrichNet software)

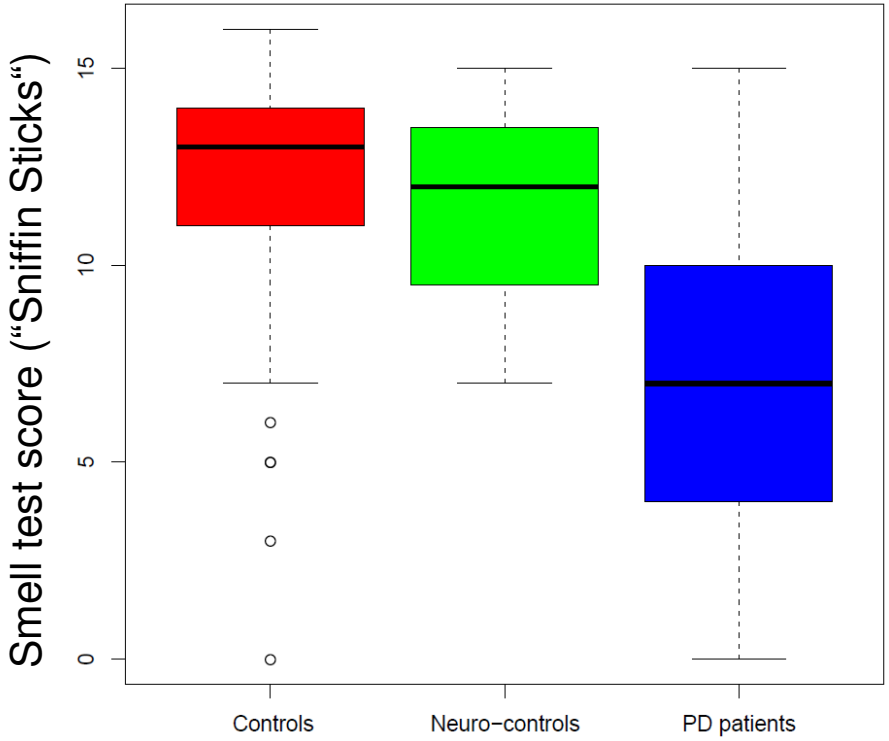
→ Identify disease-associated network alterations in Parkinson's molecular data:



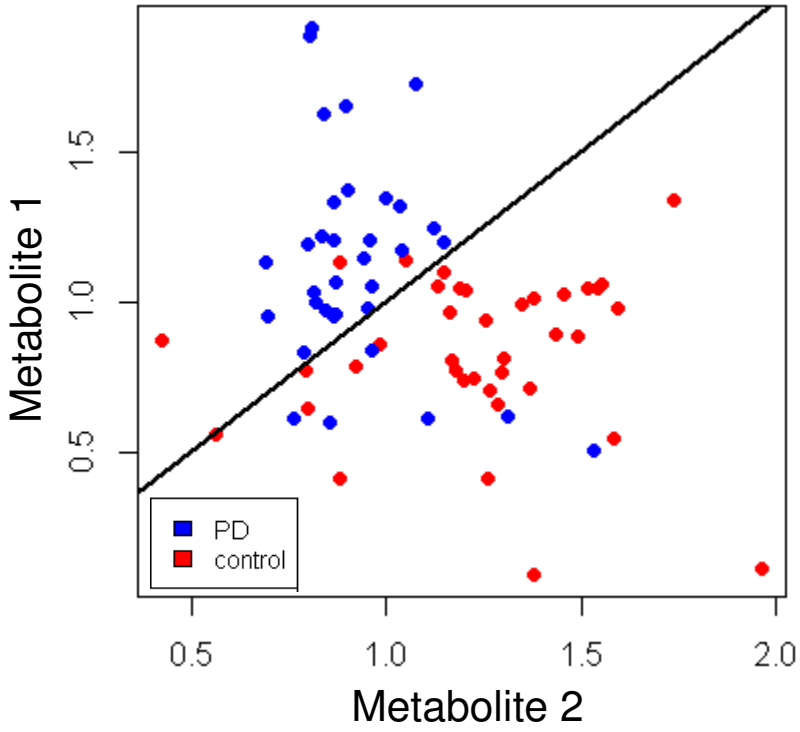
Mutation associated with
early-onset parkinsonism
(Quadri et al., 2013; Krebs et al., 2013)

Integrative machine learning applied to Parkinson‘s disease

Parkinson‘s clinical feature (Smell test)



Molecular features (Metabolites)



Combining standardized molecular and clinical features for diagnostic classification increases Parkinson vs. control cross-validated classification accuracy from 86% to 90%

(DeNoPa Parkinson cohort, 72 metabolite samples from cerebrospinal fluid)

UNIVERSITÉ DU LUXEMBOURG

LCSB

Summary

- Both hardware and software resources for efficient biomedical data processing and interpretation have been set up, enabling integrated and reproducible analyses

- Prior biological knowledge is exploited via new graph-based analysis tools to identify disease-linked network perturbations across multiple datasets and their associations with cellular pathways

- Machine learning methods integrating complementary data sources (e.g. clinical and molecular data) provide prediction models with increased robustness for diagnostic biospecimen classification

References

1. E. Glaab, *Using prior knowledge from cellular pathways and molecular networks for diagnostic specimen classification*, Briefings in Bioinformatics (2015), 17(3), pp. 440
2. 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
3. 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
4. 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)
5. 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
6. 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
7. E. Glaab, R. Schneider, *RepExplore: Addressing technical replicate variance in proteomics and metabolomics data analysis*, Bioinformatics (2015), 31(13), pp. 2235
8. E. Glaab, *Building a virtual ligand screening pipeline using free software: a survey*, Briefings in Bioinformatics (2015), 17(2), pp. 352
9. E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, A. Valencia. *EnrichNet: network-based gene set enrichment analysis*, Bioinformatics, 28(18):i451-i457, 2012
10. E. Glaab, R. Schneider, *PathVar: analysis of gene and protein expression variance in cellular pathways using microarray data*, Bioinformatics, 28(3):446-447, 2012
11. 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
12. E. Glaab, A. Baudot, N. Krasnogor, A. Valencia. *TopoGSA: network topological gene set analysis*, Bioinformatics, 26(9):1271-1272, 2010
13. 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
14. 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
15. 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
16. 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
17. 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