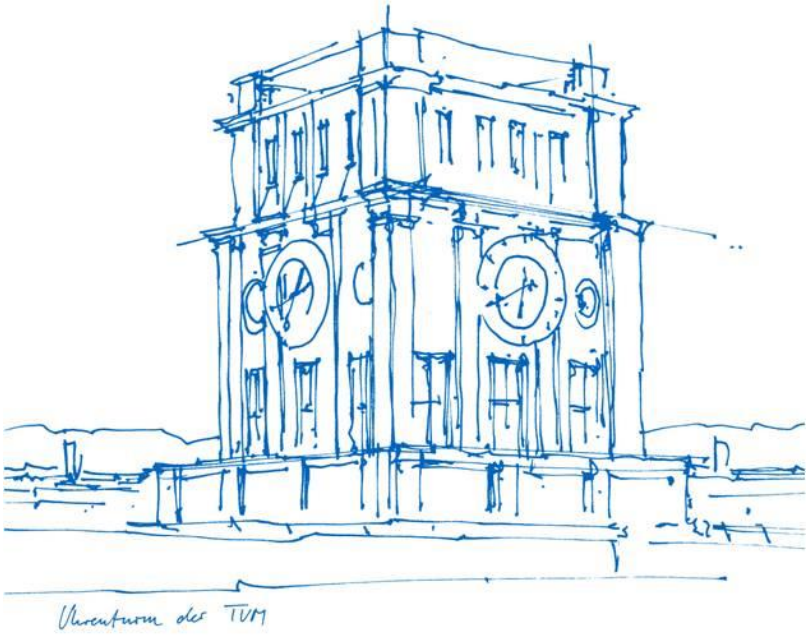
Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Multi-block PCA in omics data integration

- A brief overview



Dr. Chen Meng

Bavarian Center for Biomolecular Mass Spectrometry

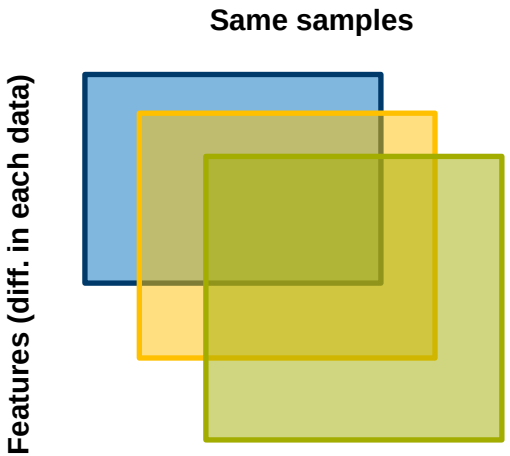
TU Munich, Freising, Germany

BIRS BioIntegrationWorkshop (Online)

2020.06.16

Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

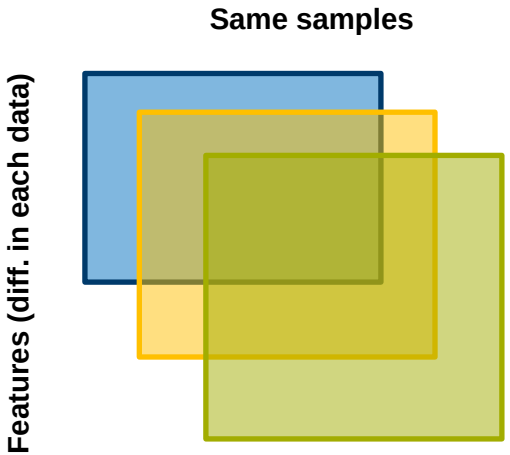
Omics integration - Overlapping samples

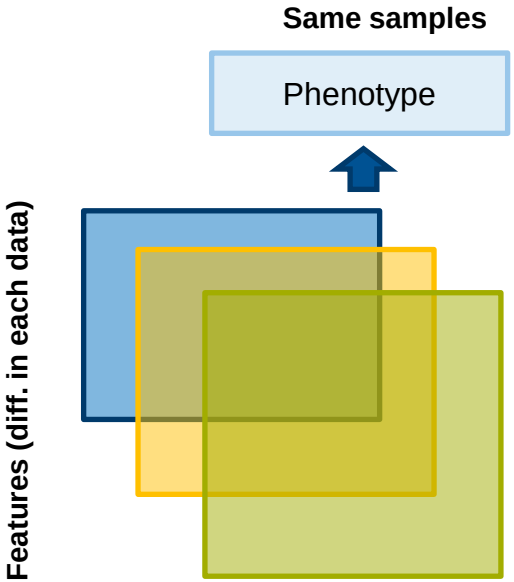


Exploratory analysis:
- PCA/MFA
- Correspondence Analysis or multi-block PCA (CCA/MCIA)

Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration - Overlapping samples



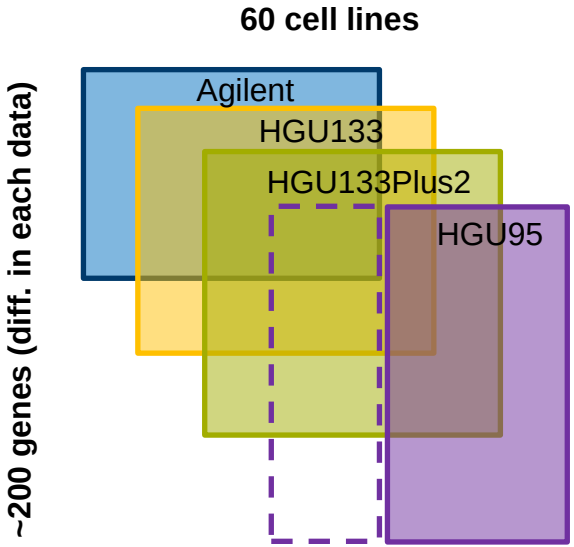


Exploratory analysis:
- PCA/MFA
- Correspondence Analysis or multi-block PCA (CCA/MCIA)

Prediction problem - Generalization of PLS
- Concordance analysis/(sparse) multi-block PLS
- Implemented in MixOmics?

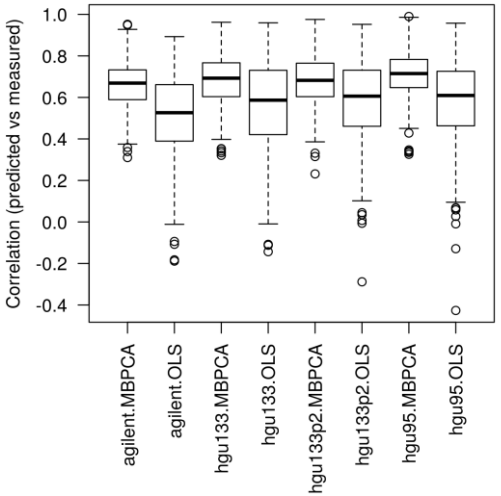
Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration – Partially overlapping samples



How should we approach integrating partially-overlapping proteomic data collected on different patients with similar phenotypes?

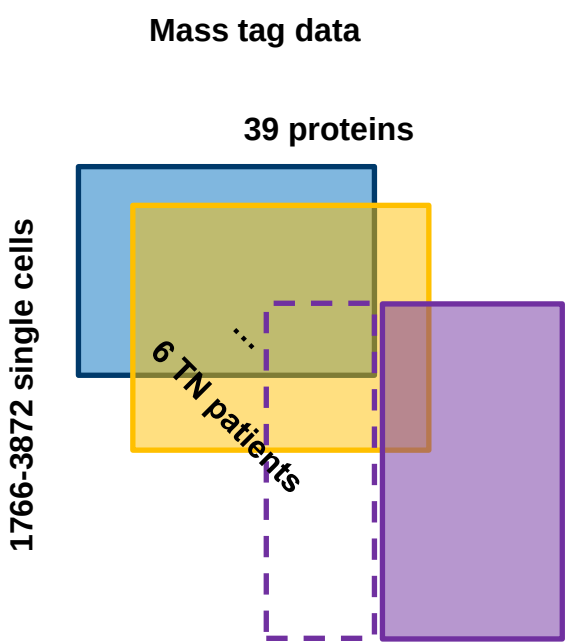
MBPCA – projection-re-construction the missing samples in HUG95
Least square regression: training linear model using Agilent, HGU133 and HGU133Plus2



https://github.com/mengchen18/BIRSBioIntegrationWorkshop

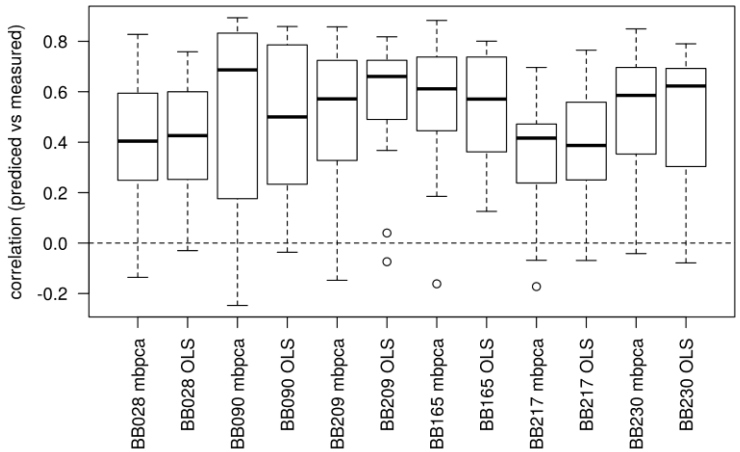
Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration – Partially overlapping proteins



How should we approach integrating partially-overlapping proteomic data collected on different patients with similar phenotypes?

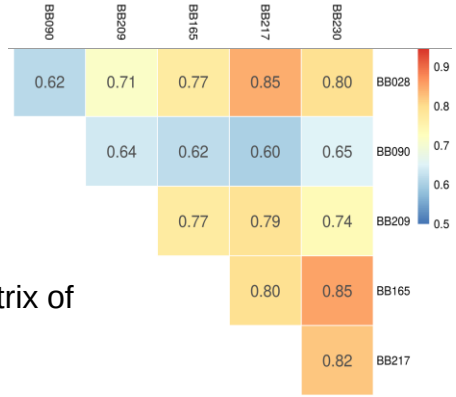
MBPCA – projection-reconstruction the missing proteins in the purple patient
Least square regression: training linear model using other patients than the purple and make predictions using purple



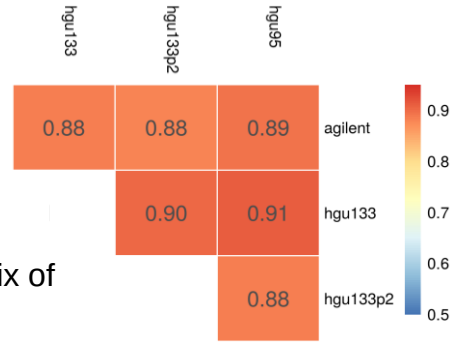
https://github.com/mengchen18/BIRSBioIntegrationWorkshop

Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Can MBPCA be better? – RV coefficient



RV coefficient matrix of the massTag data



RV coefficient matrix of the NCI60 data

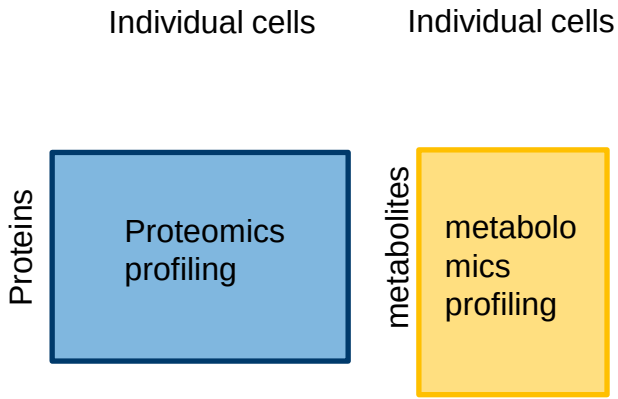
Normalization

- Row wise (centering, scaling)
- Column wise (centering, VSN)
- Data wise (MFA, STATIS, weighting matrices according to their similarity to the matrix to be predicted)

Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration – phenotype overlapping

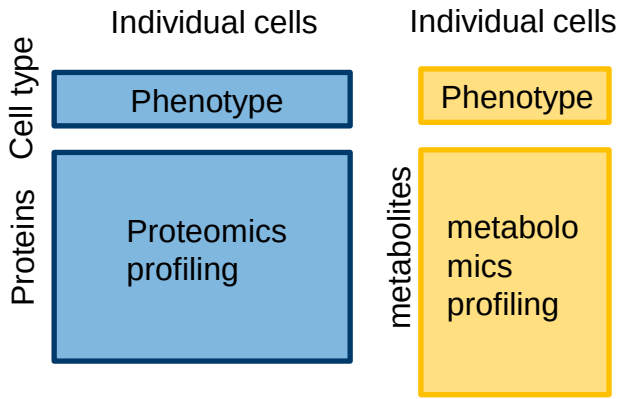
No overlapping samples/cells
Different molecule measured
All we have is the samples/cells shares similar phenotypes e.g. cell types



Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration – phenotype overlapping

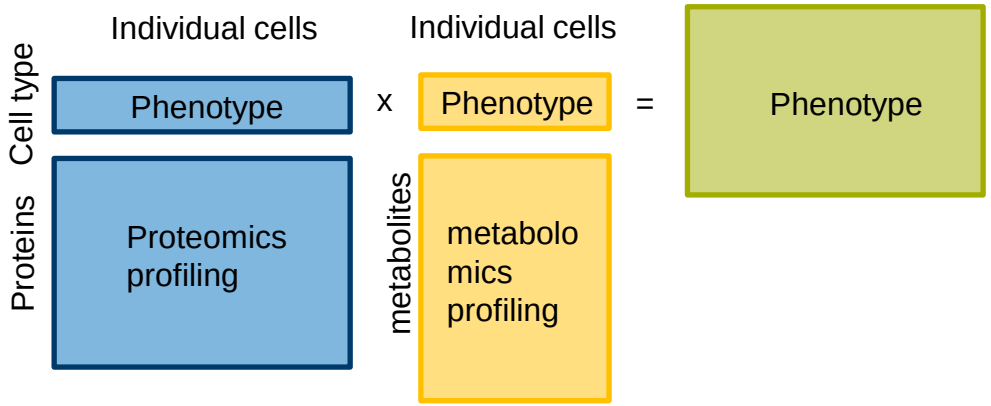
No overlapping samples/cells
Different molecule measured
All we have is the samples/cells shares similar phenotypes e.g. cell types



Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration – phenotype overlapping

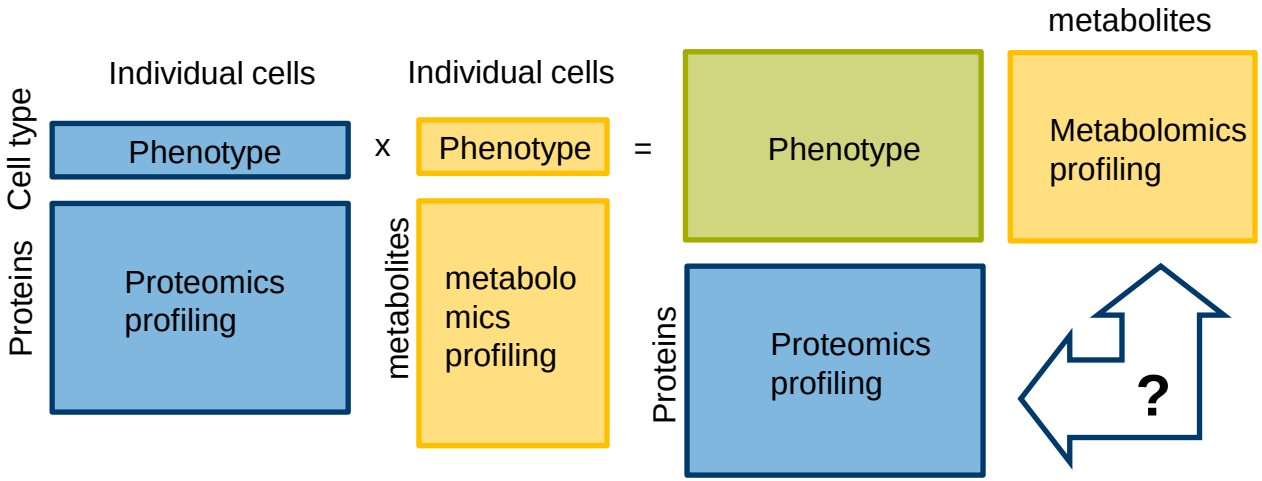
No overlapping samples/cells
Different molecule measured
All we have is the samples/cells shares similar phenotypes e.g. cell types



Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration – phenotype overlapping

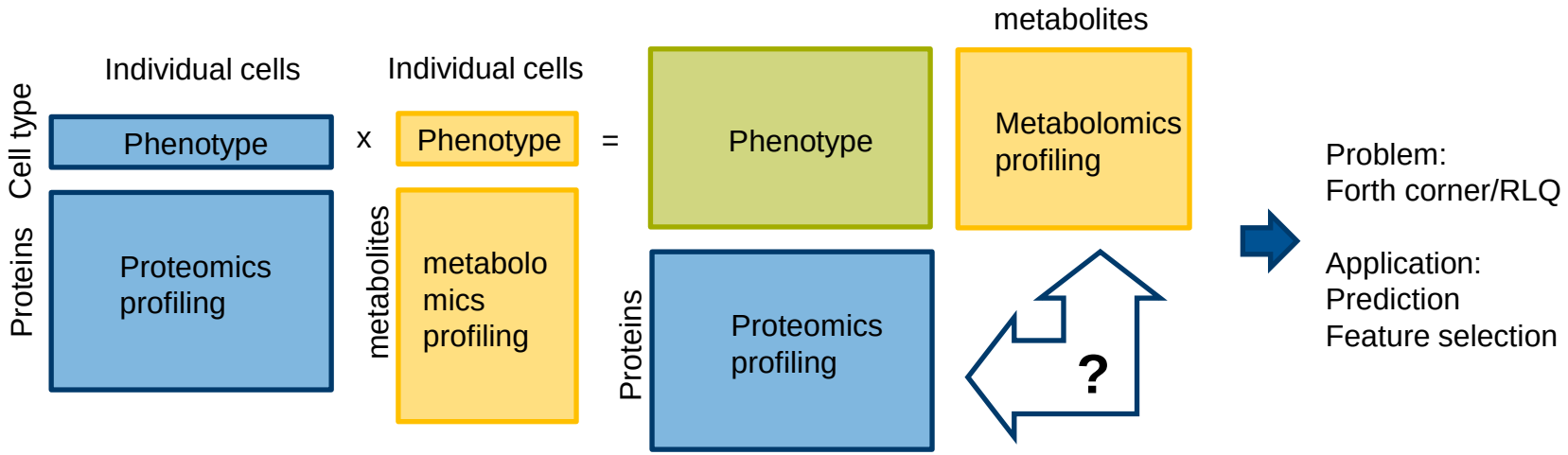
No overlapping samples/cells
Different molecule measured
All we have is the samples/cells shares similar phenotypes e.g. cell types



Bavarian Center for Biomolecular Mass Spectrometry
TUM School of Life Sciences Weihenstephan
Technical University of Munich

Omics integration – phenotype overlapping

No overlapping samples/cells
Different molecule measured
All we have is the samples/cells shares similar phenotypes e.g. cell types



Final remarks

- Matrix decomposition has a great potential for integration analysis of multi-omics data
- A proper normalization of dataset (on all levels) is essential
- Adaptation/extension of current method is important to fit the needs of specific biological questions
- (Interactive) visualization and integration with prior knowledge (GO, pathway) will be a critical factor whether people will use it