Swiss Institute of Bioinformatics

Multiblock analysis for multiomics data integration

Florence Mehl, Josefina Lascano Maillard

Vital-IT

2/04/2024, Servier, Paris

elixir SWITZERLAND

Agenda

- **General introduction**
- **Methods based on components – 1 block**
  - PCA
  - PLS
  - OPLS
- **Multiblock analysis**
  - Introduction on Data Fusion
  - Unsupervised analysis – ComDim
  - Supervised analysis – consensus OPLS
- **Conclusion**
- **Hands-on**

The data explosion

Modern scientific technologies can generate **massive datasets** to describe specific phenotypes illustrating a biological phenomenon

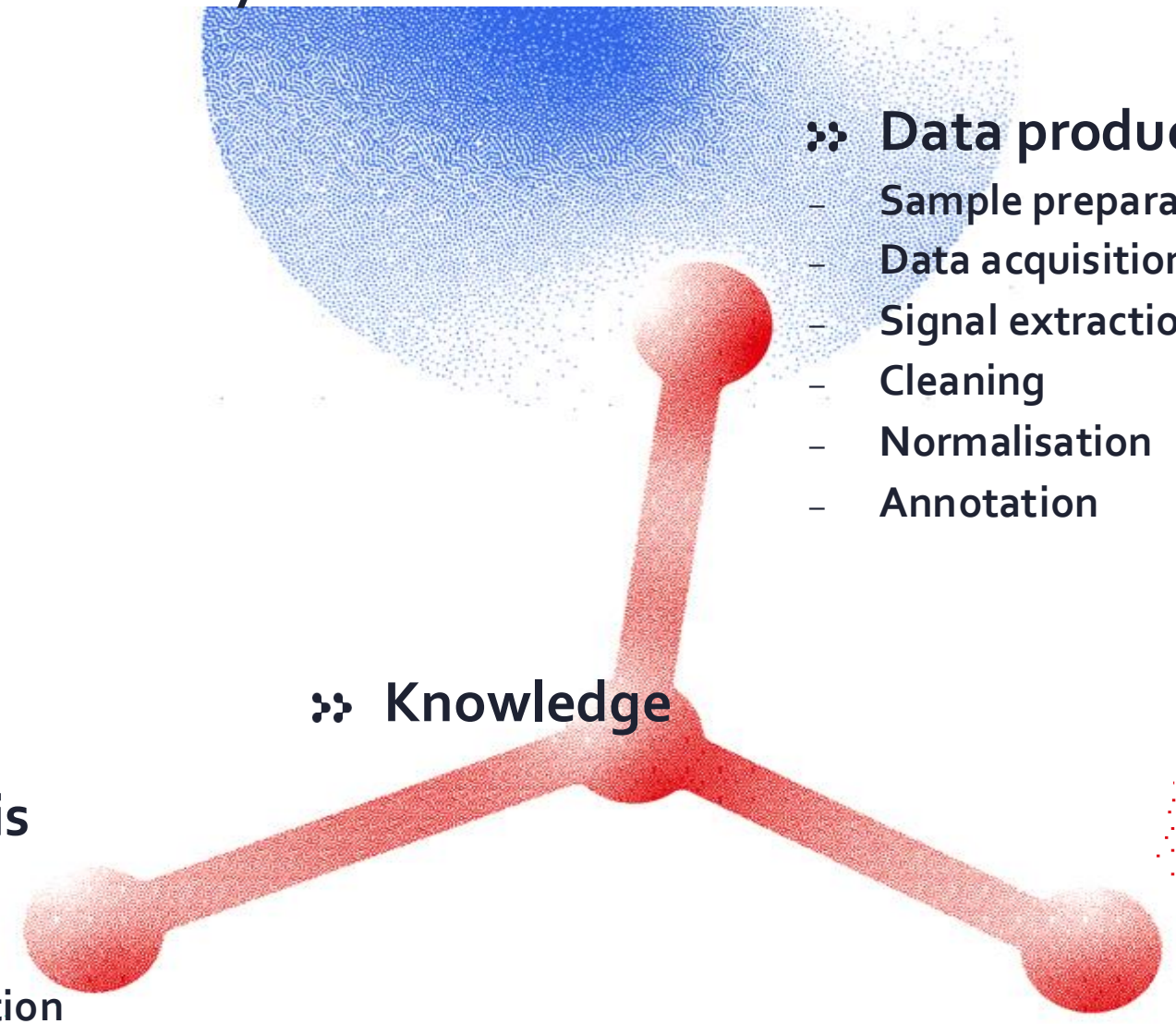
Hypothesis-driven approaches

- hypothesis
- experimental proof

Data-driven approaches

- data
- hypotheses

Knowledge discovery in Omics



Data production & processing

- Sample preparation
- Data acquisition
- Signal extraction
- Cleaning
- Normalisation
- Annotation

Knowledge

Data analysis

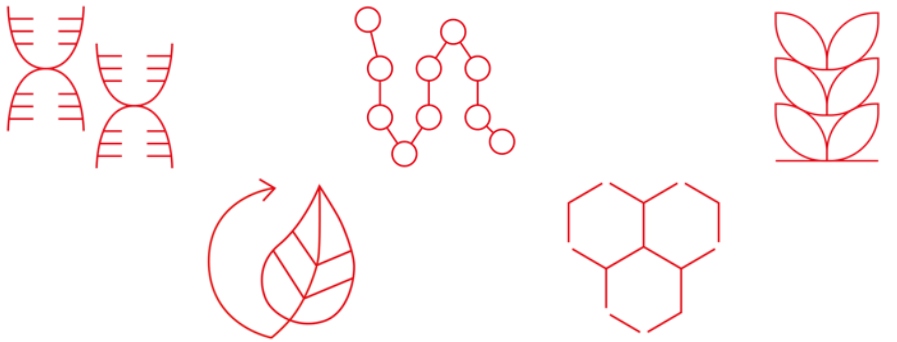
- Exploration
- Classification
- Pattern recognition
- Variables contribution

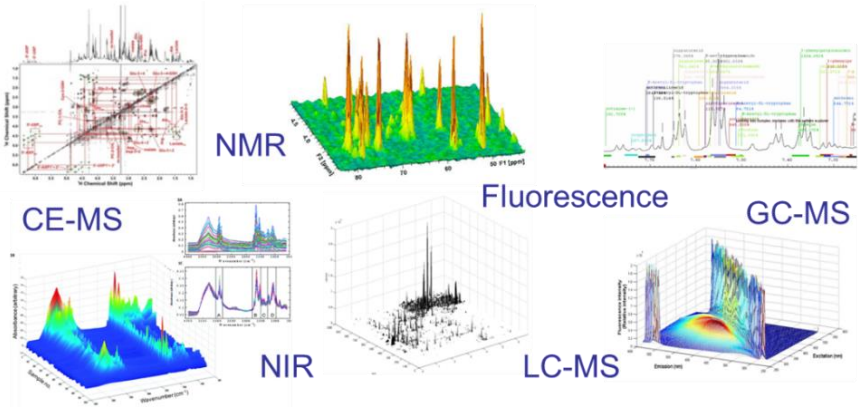
Interpretation

- Biological contextualisation
- Biological validation

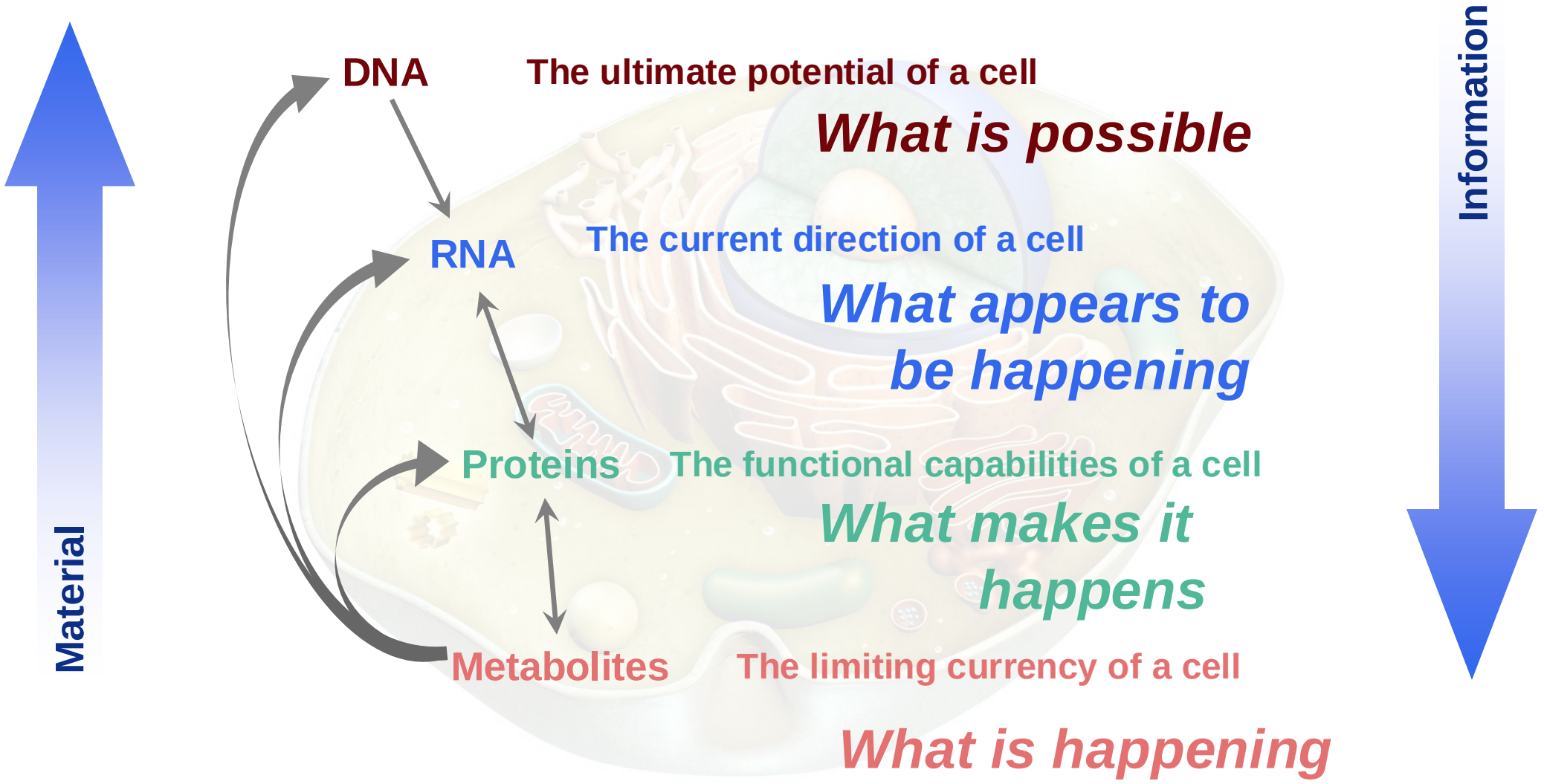
Multiple data sources

- **Different biological scales**
  - Cell/tissue/organism
  - Systems biology
- **Different stage of a process**
  - Dose
  - Toxicity
  - Disease progression
- **Different analytical techniques**
  - Heterogeneous data
  - Separation or spectral methods





Central Dogma & Systems Roles



Embracing Complexity – the Swiss Watch

How does a Swiss watch work?

✗ Examine separately the springs, gears, shafts, etc. how they fit together

or

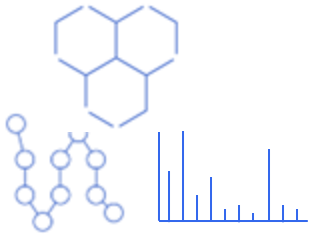
✓ Consider all the elements at once and how they fit and interact together

MULTIGROUP ANALYSIS
DATA FUSION
DATA INTEGRATION
MULTIVIEW ANALYSIS
MULTITABLE ANALYSIS
MULTISET ANALYSIS
MULTIBLOCK ANALYSIS

Nature of data

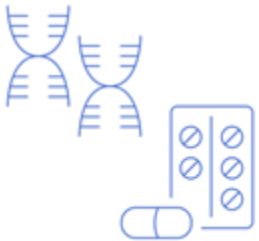
Homogeneous data: data blocks all measured on the same scale (e.g. quantitative data)

Heterogeneous data: data blocks measured on different scales (e.g. quantitative, ordinal, qualitative, binary)



- **Quantitative** (numeric variables that can take any value between of real numbers)
  - Transcripts / Proteins / Metabolites
  - Abundance / expression levels

- **Ordinal** (categorical variable that can be ranked)
  - Lifestyle



- **Qualitative** (categorical variable that cannot be ranked)
  - SNPs
  - Clinical

Objectives of data integration

Combine data sources

- Gain an extended understanding of complex systems
  - use complementarity of data to obtain more complete or new information
  - use redundancy of data to increase confidence
- Better prediction

Compare data sources

- Rank data sources (e.g. analytical methods)
- Block selection

Multiomics data integration strategies

Deep Learning

- deep neural networks, multiple kernels learning, autoencoders, recurrent neural networks, generative adversarial networks
- can automatically learn feature representations that capture the underlying relationships and patterns within these heterogeneous data sources

Similarity- (Kernel-) based methods

- Similarity Network Fusion, rMKL-LPP, NEMO
- calculate a similarity score between each pair of patients based on the combination of omics measurements

Matrix Factorization

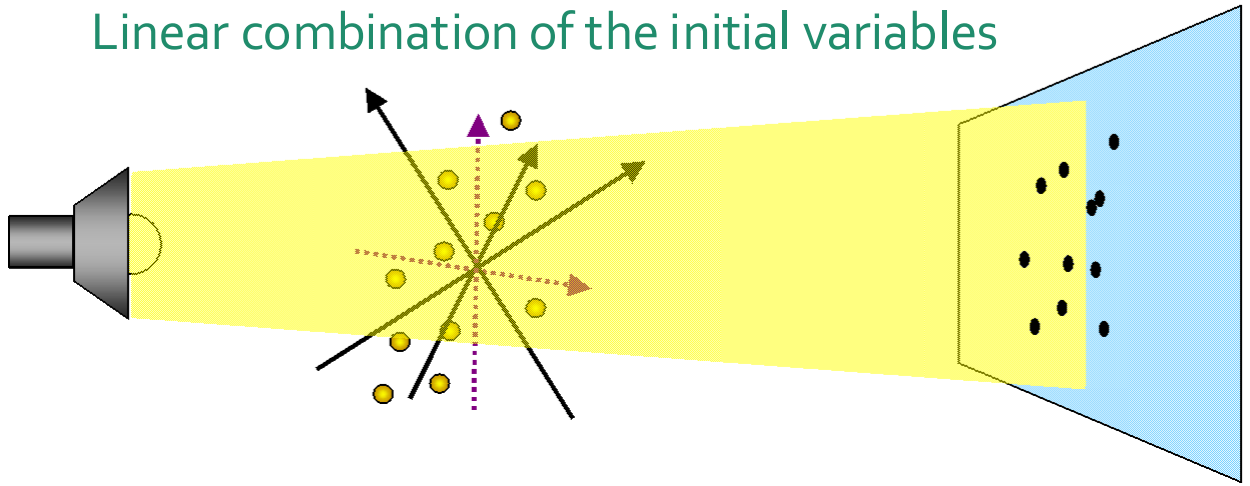
- MOFA, RGCCA, CCSWA/ComDim, consensus OPLS
- perform dimensionality reduction by decomposing the datasets into a smaller number of factors

Methods based on components

- Omics data
  - high dimensionality
  - correlated variables
  - biological variability
  - experimental noise

- Projection methods can
  - analyze datasets of high dimensionality
  - provide knowledge about systems
  - find unsuspected relationships
  - summarize the data with a small number of factors

Linear combination of the initial variables



Visualisation of

- samples' distribution
- correlations between variables

Variance, Covariance & Correlation

- **Variance**: an indicator of spread for one variable

The higher the variance, the most scattered the observations

- **Covariance**: an indicator of relationships for 2 sets of observations

The higher the covariance, the most related the 2 variables.

- **Correlation**: standardized version of the covariance that takes values between -1 and 1.

  Correlation is widely used to describe the relationship between variables
  - linear relation between two variables (Pearson)
  - non-parametric relation based on the ranks of the variables (Spearman)

Unsupervised & supervised analysis

- **Unsupervised analysis**
  - explorative analysis looking for structures and patterns
  - no a priori knowledge about the observations

- **Supervised analysis**
  - predictive data analysis
  - emphasis is on a response block of data Y (a measured outcome)

Principal Component Analysis (PCA)

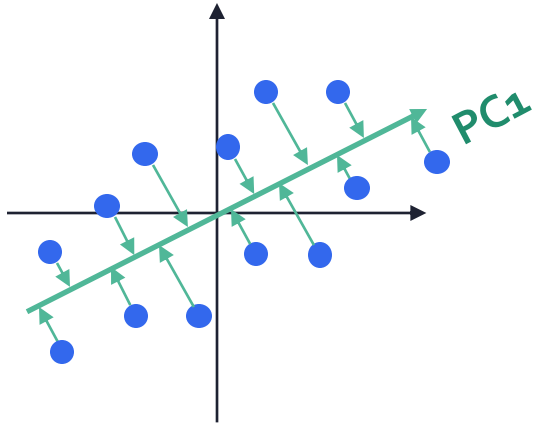
- **Unsupervised analysis**
  - explore pattern / groupings of observation
  - evaluate the variables contributions

- **Reduce the dimension of the feature space**
  - Feature selection: find a subset of input features
  - Feature extraction: project high-dimensional space into a space of fewer dimensions

Karl Pearson – mathematician & biostatistician (1901)

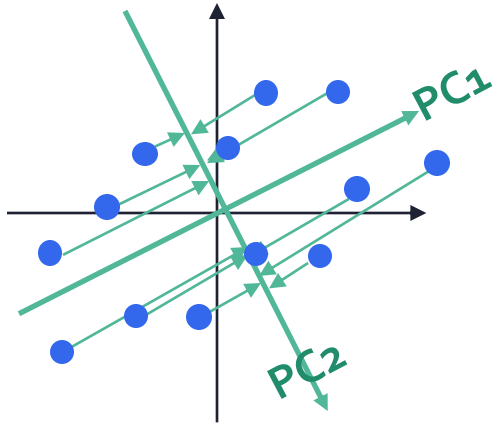
How does PCA work?

- The 1st component is the direction of maximum variance from origin



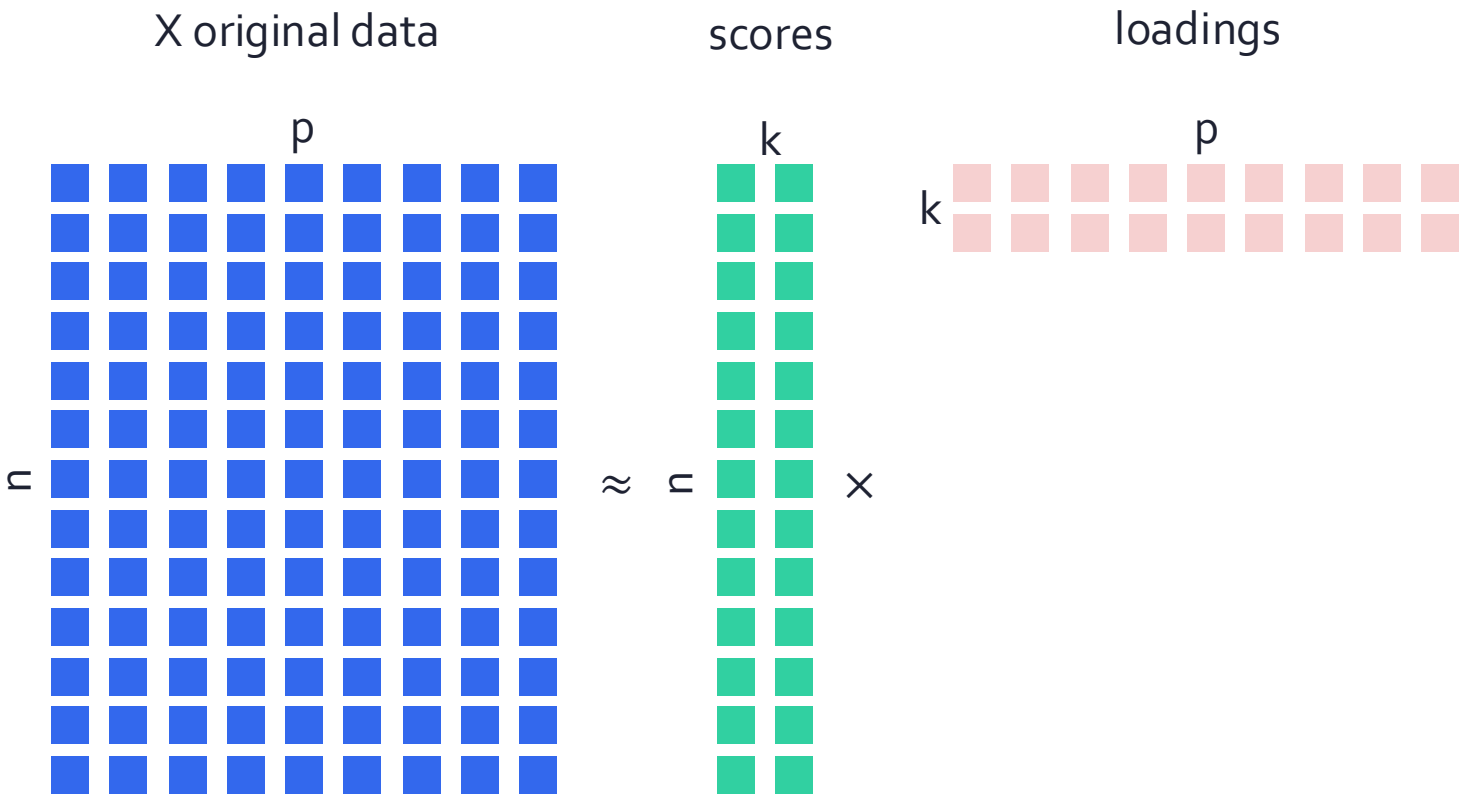
53% of the total variance

- The 2nd component is orthogonal to the 1st, describes the maximum residual variance and is therefore uncorrelated

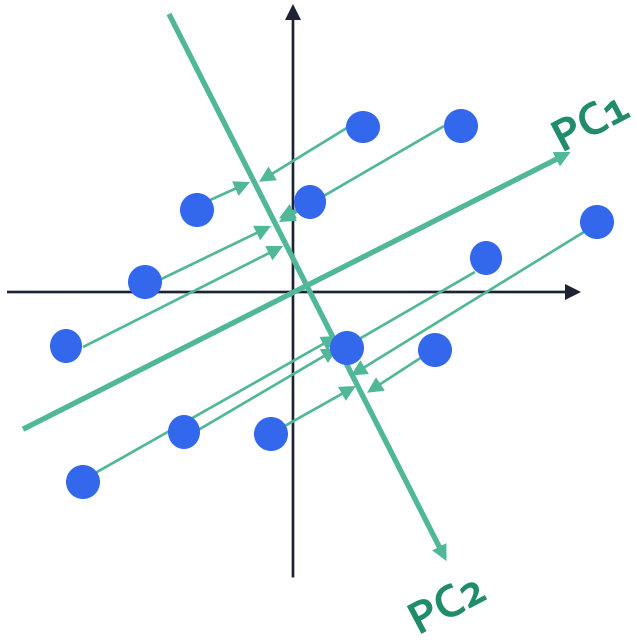


17% of the total variance

PCA matrix factorization



n observations × p variables
k components



PCA output

- **Scores**: projection measures of samples in each principal component

  Coordinate of samples on each axis PC

- **Loadings**: contribution of given features to each principal component

  Coordinate of given features on each axis PC

  Highly correlated features: similar weights in the loading vectors; close together in the loading plots of all dimensions.

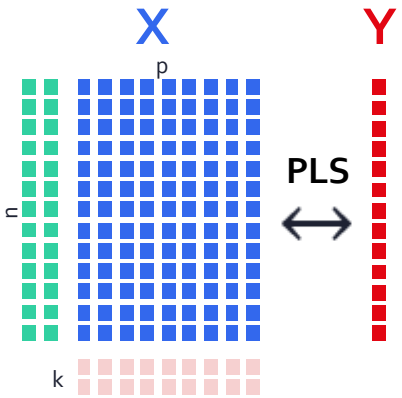
Partial Least Square (PLS)

- **Supervised analysis**
  - Many (predictors or independent/explanatory variables) vs one (response or dependent variable)
  - Many vs Many

- **Reduce the dimension of the feature space**
  - Feature extraction: project high-dimensional space into a space of fewer dimensions

- **Find the relation between the two sets of variables**
  - covariance

Herman Wold – econometrician (1966)

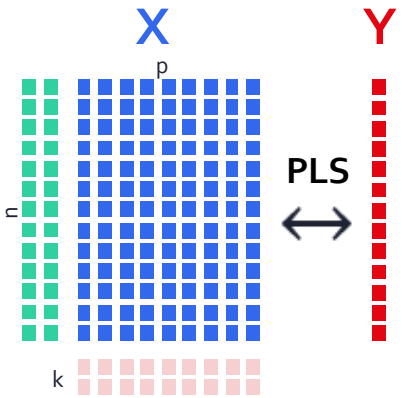


PLS Discriminant Analysis (PLS-DA)

X : n observations × p variables

Y : n observations × 1 categorical variable

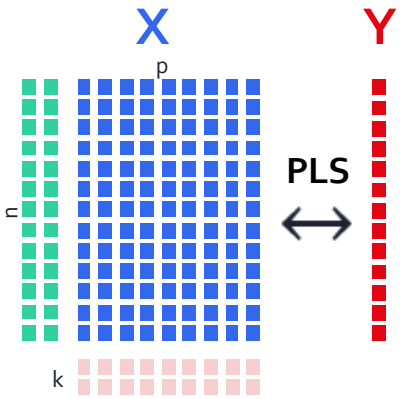
- **Supervised version of PCA**
  - Dimensionality reduction, feature selection, classification
  - Maximize covariance between each latent variable and the labelling



How does PLS work?

- **PCA objective**
  - calculate latent variables (components) that best explaining variance in X
  - Maximize variance in X components

- **PLS objective**
  - calculate components that best explain variance in X
  - calculate components that best explain variance in Y
  - calculate components that have greatest relationship between X and Y
  - Maximize covariance between X and Y



Orthogonal PLS (OPLS)

- **Orthogonal Projection on Latent Structures**
  - removes variation from predictors X that is **not correlated** to response(s) Y
  - maximize explained variance on the first component(s) of predictors
  - remaining components capturing variance that is orthogonal to responses
  - **Model separately variations of X correlated and uncorrelated to Y**

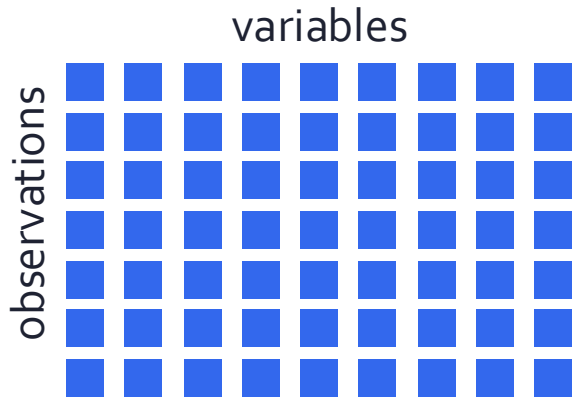
- Reduce model complexity: lower the number of latent variables
- Allow identification and investigation of the source of orthogonal variation
- Interpretation is easier
- Produce more efficient predictive model, particularly when structured noise dominates

Johan Trygg and Svante Wold (2002)

Data structures

- One-way data is a **vector**, with a single data value for each element of the single dimension (i)

- Two-way data is a **matrix**, with a single data value for each element of two separate dimensions (i,j)

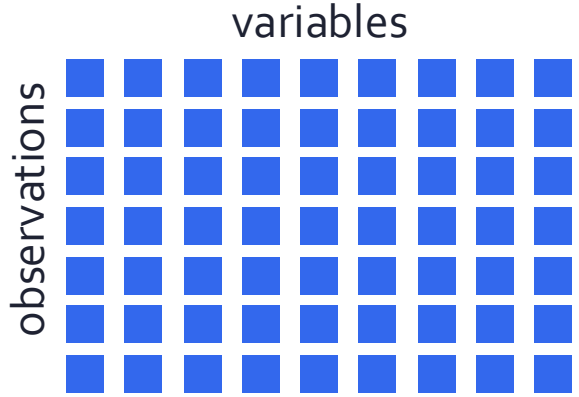


PCA, PLS and OPLS are performed on matrices

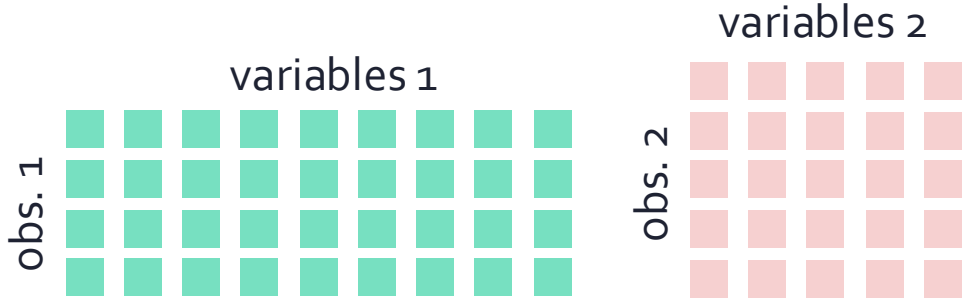
Data structures

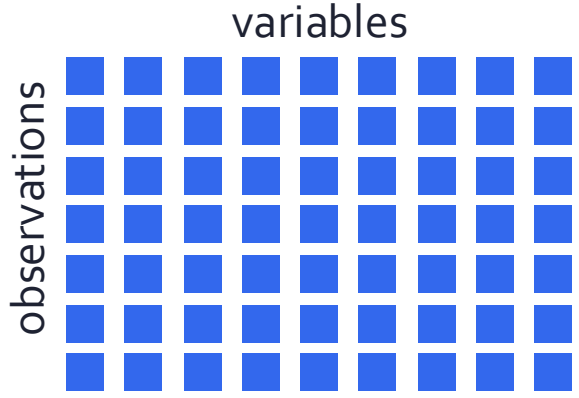
- One-way data is a **vector**, with a single data value for each element of the single dimension (i)

- Two-way data is a **matrix**, with a single data value for each element of two separate dimensions (i,j)



- Multiblock data can be seen as a **collection of matrices**





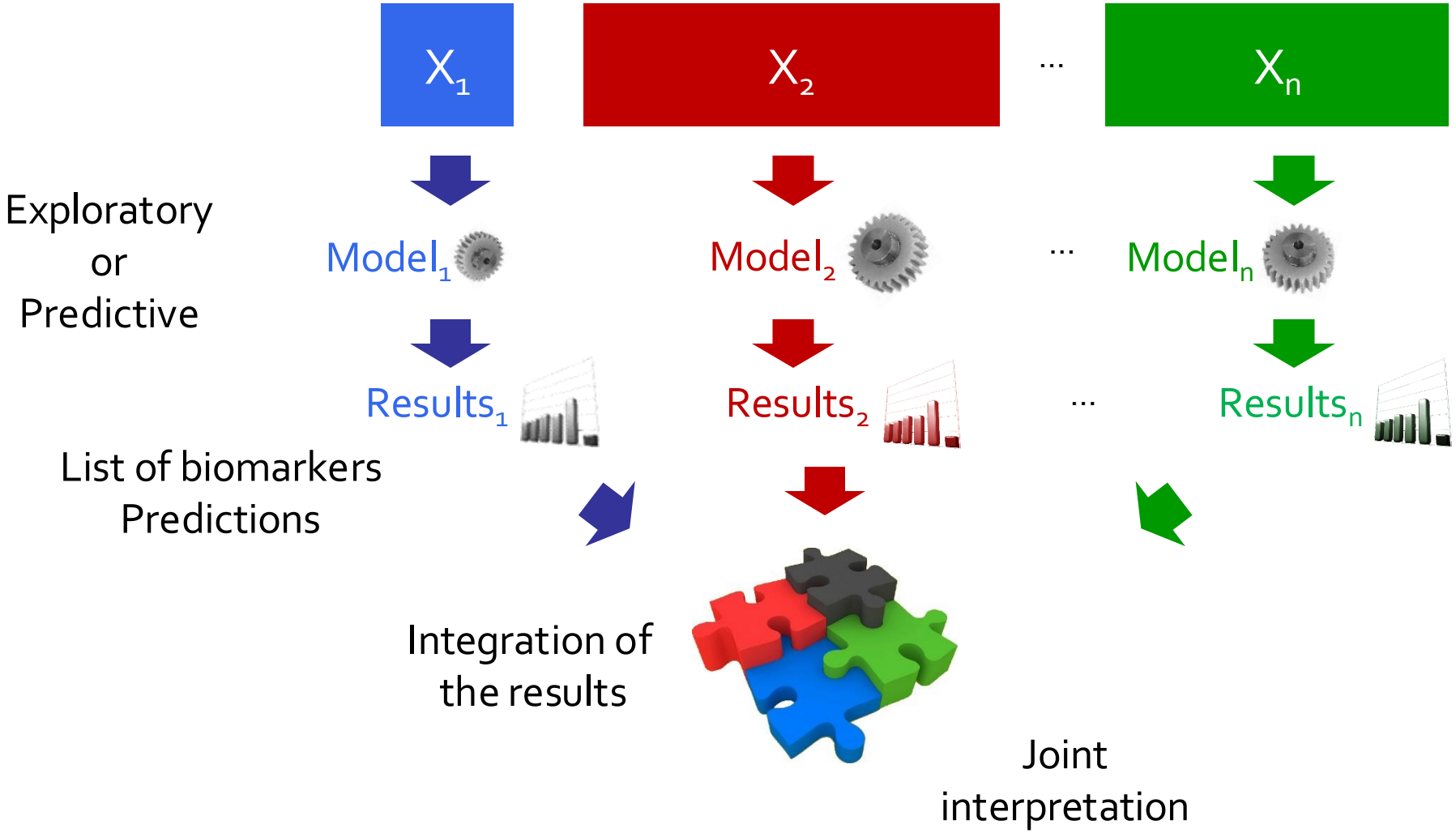
Data fusion / integration

Strategies for data integration based on abstraction levels

- **High-level (symbolic representations or decisions) / late data integration**
  - information/decision fusion
- **Mid-level (patterns or subsets extracted from the sources) / intermediate**
  - characteristics employed for other tasks
- **Low-level (signals) / early data integration**
  - data fusion/aggregation/association

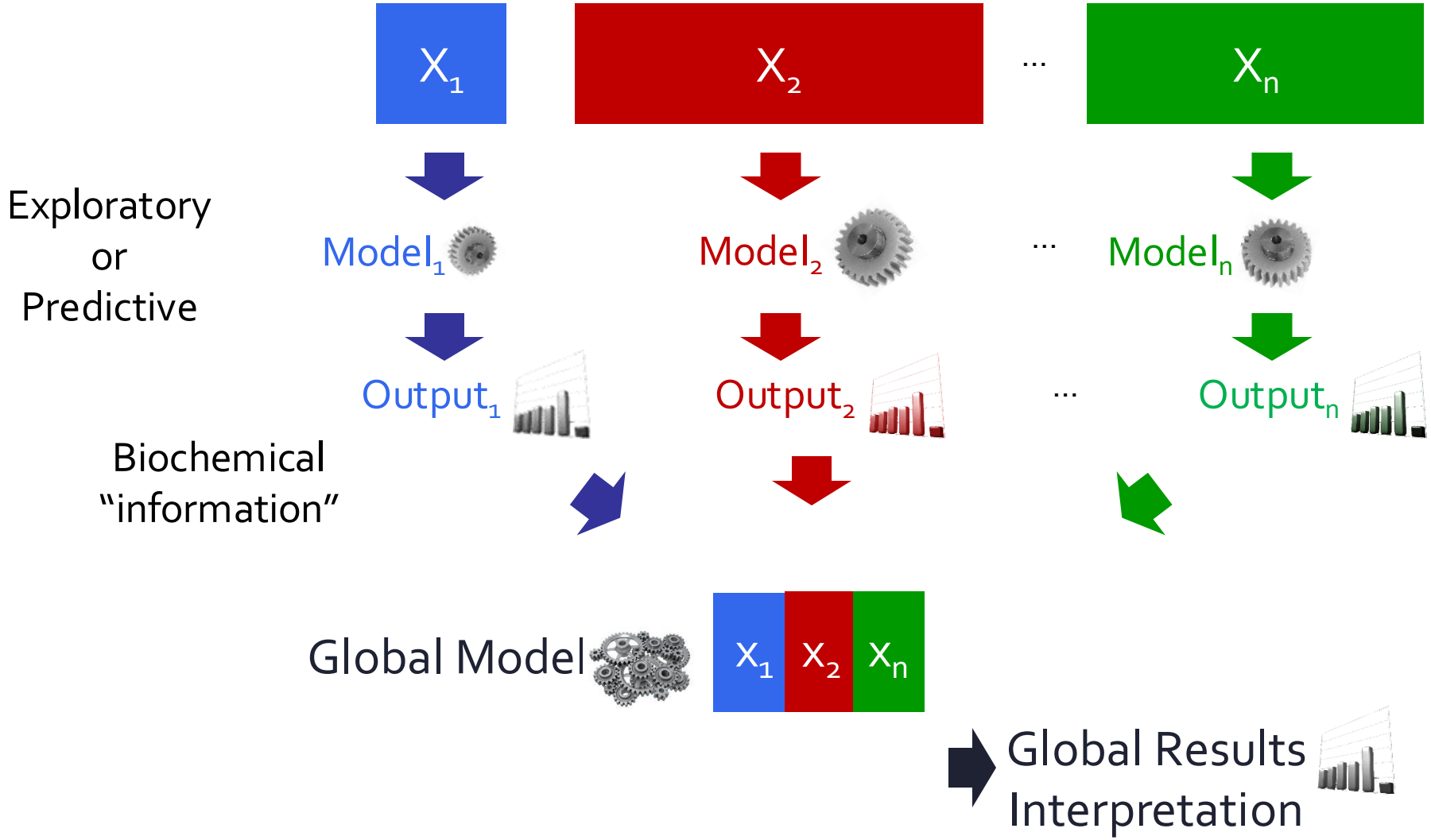
High level data fusion

Integration of results from single blocks models



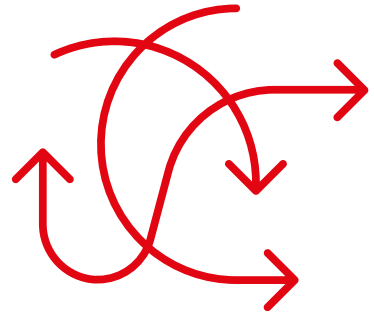
Mid-level data fusion

Integration of results from single blocks models



Limitations of High/Mid-level Data Integration

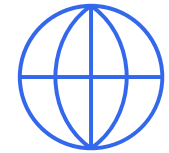
- Results may be contradictory/heterogeneous
  - Unsupervised: biological interpretation can be tedious
  - Supervised: the result can be inconclusive in the case of ties

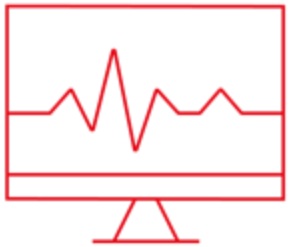


- No insight into the links between measurements
- Individual models may lead to a substantial loss of biological information
- The combination of information may not be relevant

Low-level data fusion

- **Think global** by building **a compromise** accounting for all data with **adequate weights**
- **Act local** by maximizing the link between data blocks **under a specific criterion**, e.g. canonical correlation, co-inertia, partial least squares



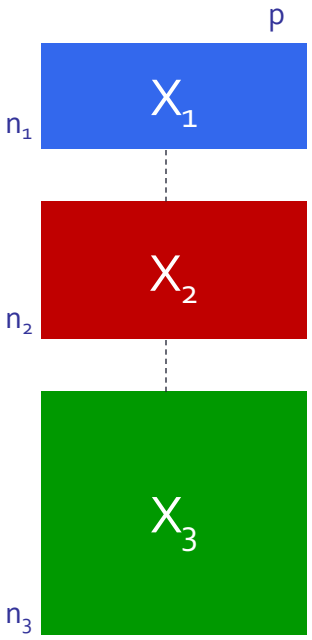


Find the relevant information

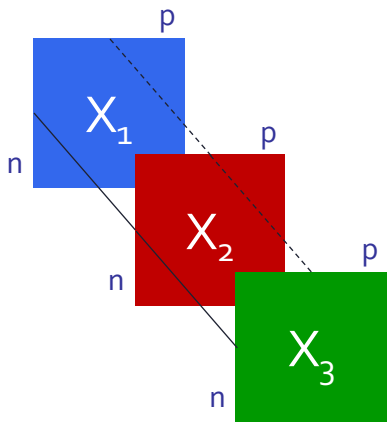
- Role / importance of each data table
- Common / specific trends
- Links between variables of different nature

Multiblock data structure

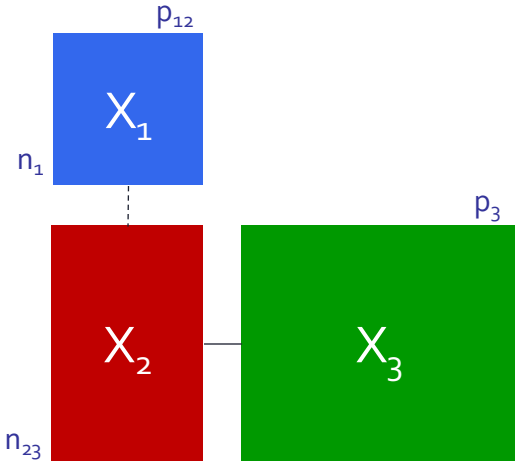
- Shared Variable mode ?
- Shared Observation or Variable mode ?
- Shared Observation and Variable mode ?
- Shared Observation mode ?



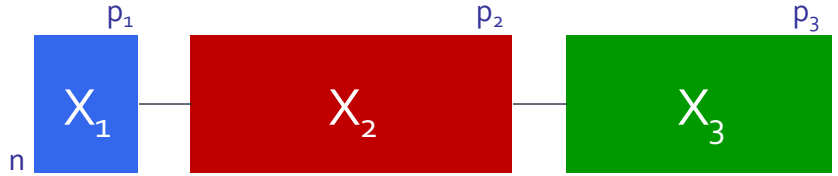
Vertical



Tensorial (multiway)



L-shape



Horizontal

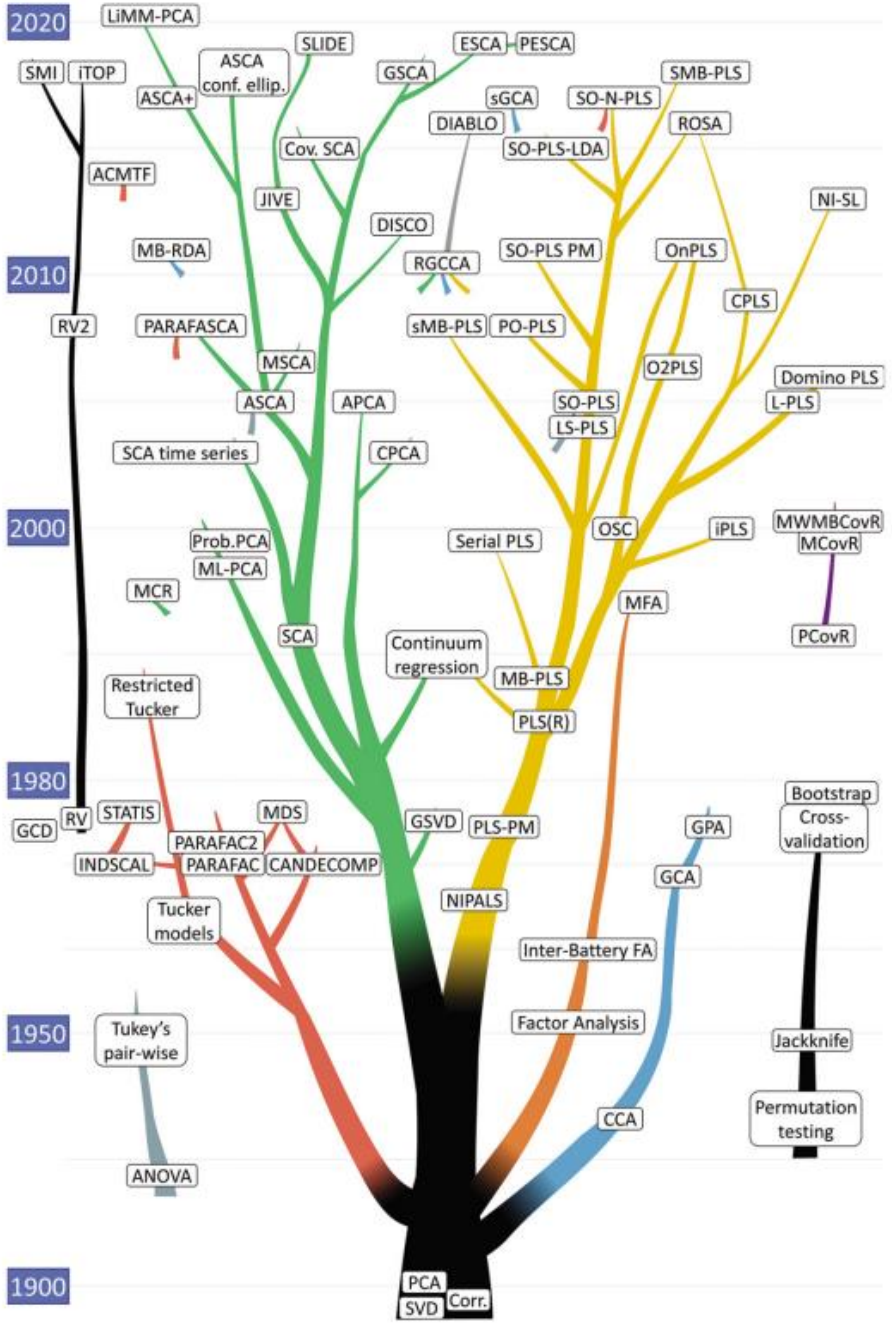
A whole forest of methods

Phylogeny of some multiblock methods and relations to basic data analysis methods

Green branch: Unsupervised multiblock
Yellow branch: Supervised multiblock

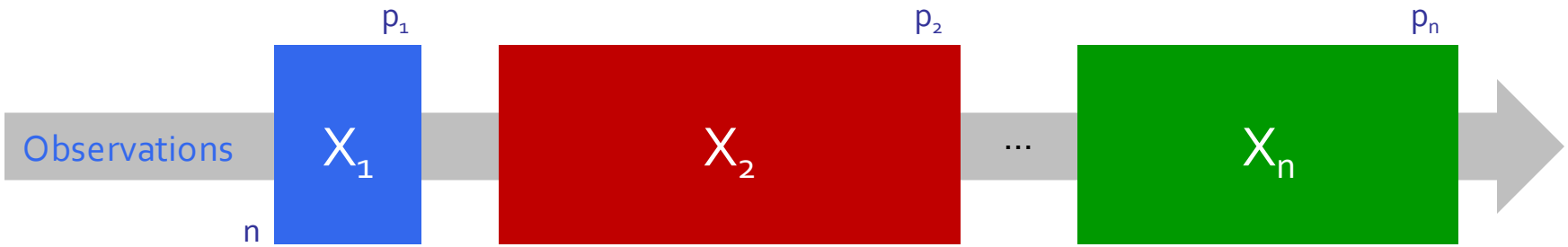
Red branch: Multiway
Blue branch: Correlation
Orange branch: Factor analysis

Black branch: Model validation



Multiblock Data Fusion in Statistics and Machine Learning: Applications in the Natural and Life Sciences
Age K. Smilde, Tormod Næs, Kristian Hovde Liland. Wiley 2022

Horizontal multiblock structure



- The **Observation mode (rows) is shared** (n observations)
- The **Variables mode (columns) is specific** (p1, p2, p3 variables)
- Many more variables than observations

Goals are the same as single-block data analysis

- **Describe – Classify – Discriminate – Predict**

Horizontal multiblock analysis

- The **Observation mode (rows) is shared** (n observations)
  - assess the relationship between the variables and the data tables
- **Unsupervised analysis**
  - explorative analysis looking for structures and patterns
  - links between variables in a single data block
  - links across data blocks
- **Supervised analysis**
  - predictive data analysis, emphasis is on a response block of data Y
  - connections to one or more blocks of data
  - some blocks are dependent, and others are independent

Matrix factorisation methods

Factor analysis can be applied to blocks instead of initial variables

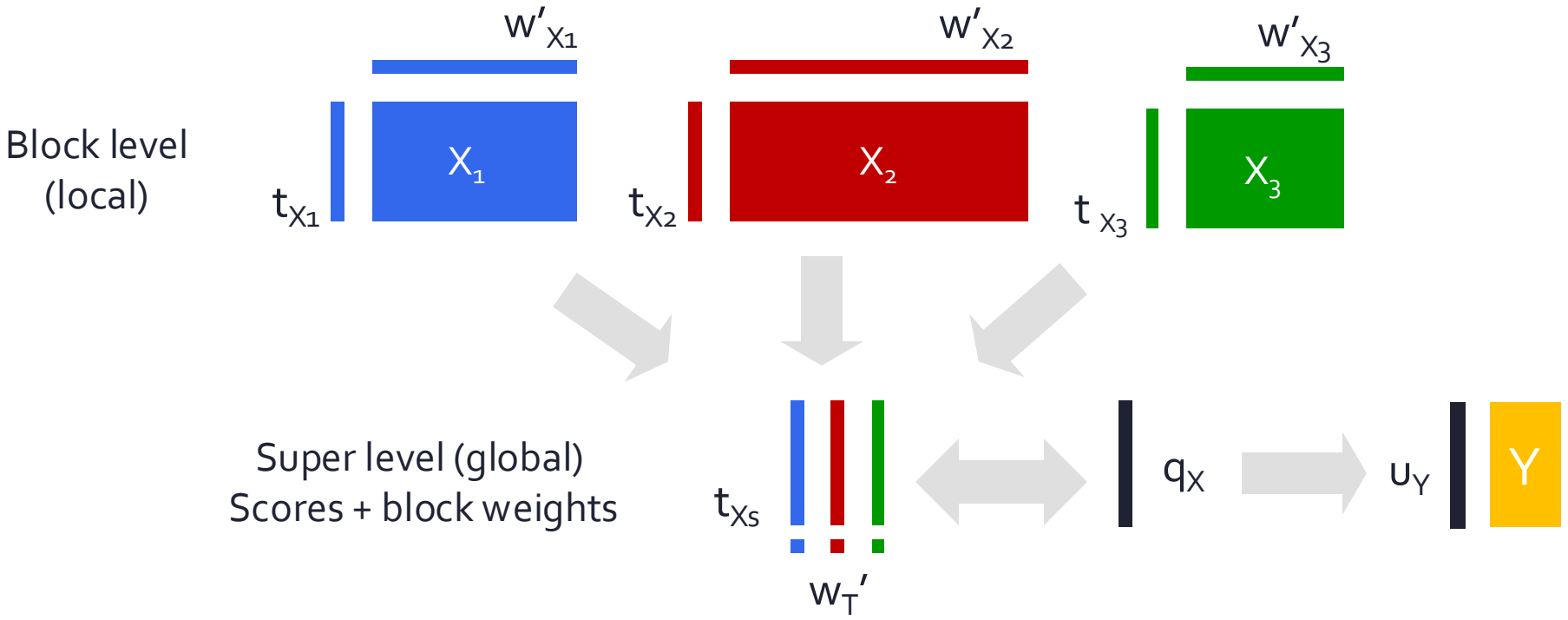
- Each component related to one block is connected to all the components related to the other blocks and/or to a global component

- Block components should verify two properties simultaneously
  - Block components **explain well their own block**
  - Block components are as **correlated** as possible **with related blocks**

- The multiblock model build components as **a compromise** for explaining **between-block** and **within-block variation**

- Different methods favor explaining more within- or between-block variation

Exploratory Multiblock Analysis



How to build the super level ?

It depends on the model!

Within/Between block scaling

- **Within block**

  If the variables have wide numerical scales, their variances will greatly depend on the range of their values

- **Between blocks**

  High-dimensional blocks will have more influence

  → Scaling according to the number of variables (1/VarNb)

  Block with large range will have more influence

  → Scaling according to block inertia/norm

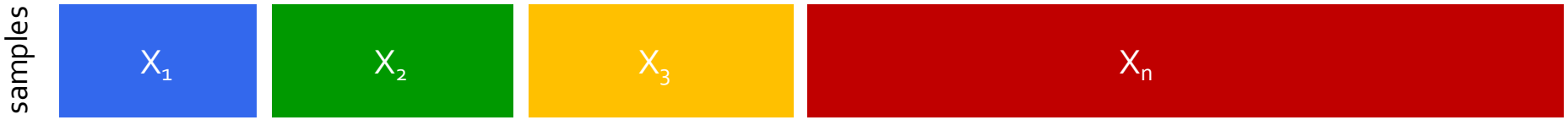
Different Weighting Strategies

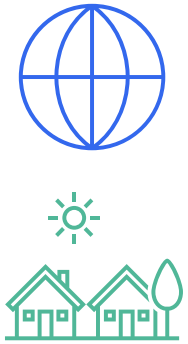
How to **balance the influence** of the different blocks in a global analysis?

The block combination is based on **specific weighting schemes**:

- Data concatenation
  → each block as a **weight of one** (SUM-PCA, MBPCA)
- Unsupervised methods
  → weights depend on **block dispersion or agreement with a compromise** (Multiple Factor Analysis, STATIS, CCSWA)
- Supervised methods
  → block weights are **driven by the Y response** (MBPLS, block-PLS, consensus OPLS)

Multiblock Data Modeling





✓ **Think global** by building **a compromise** accounting for all data with **adequate weights**

✓ **Act local** by maximizing the link between data blocks **under a specific criterion**, *e.g.* canonical correlation, co-inertia, partial least squares

✓Find the relevant information
✓Role/importance of each data table
✓Common/specific trends
✓Links between variables of different nature

Multiblock model outputs

- **New common subspace**

Common/distinct component(s)

- **Global/local observations scores**

Pattern recognition

Blocks weights (contributions)

Loadings of the initial variables

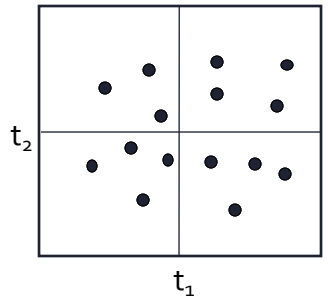
- Common/specific variation(s)

Balance between block weights

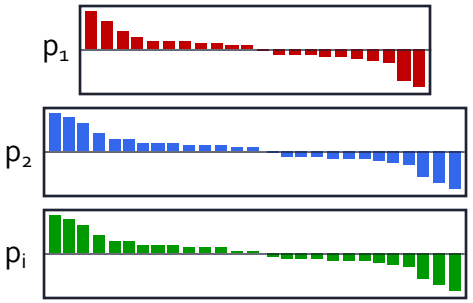
- More complete interpretation:

Links between variables

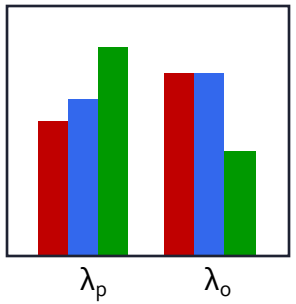
Links between blocks



Global/Local scores



Variables loadings



Blocks weights

Unsupervised multiblock analysis

- Generate hypotheses from the data blocks
- Undirected links
- All blocks are treated in the same way
  - the blocks are exchangeable
  - no block sequence

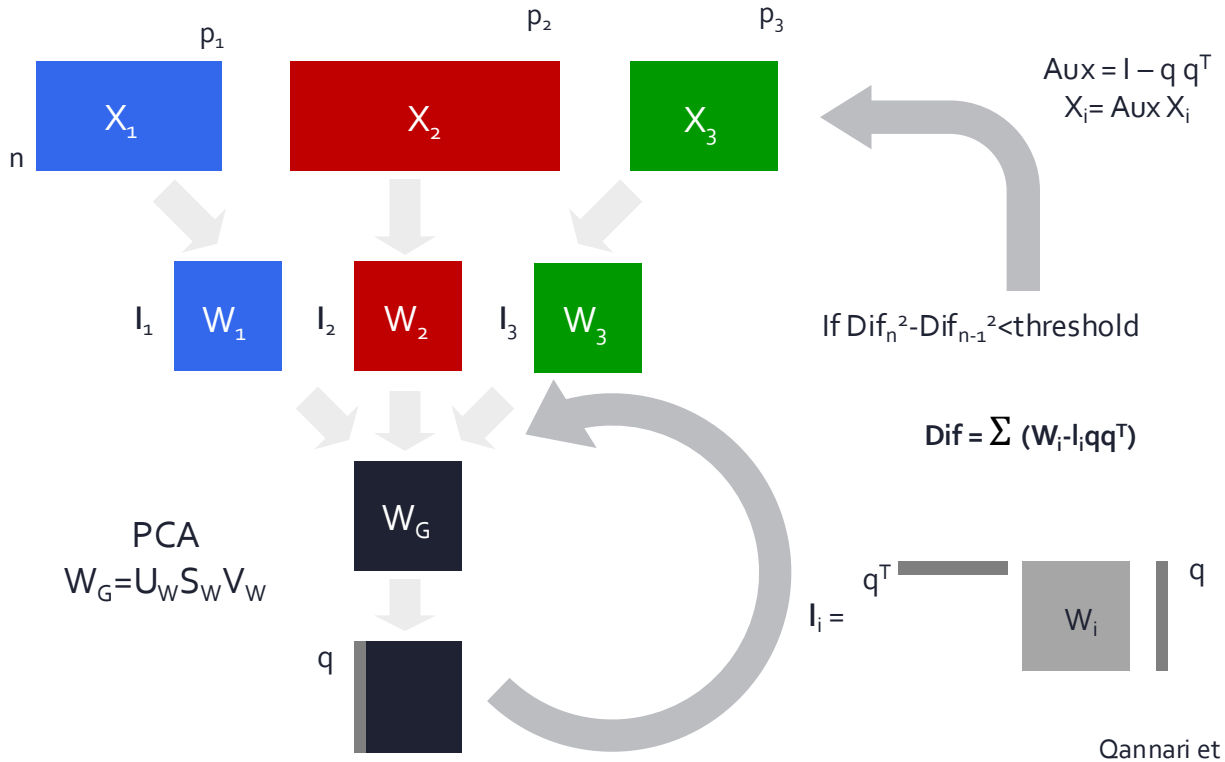
- Needs
  - focus on common & specific variation
  - fairness between block

ComDim / CCSWA Method

Common Dimensions
Common Components and Specific Weights Analysis (CCSWA)

Scalar product (defines an association matrix for each data block, similarity between observations within a block (covariances)) **+ iterative block weighting**

→ Block weighting is different from a component to another



Link between blocks: covariance[4]

$$max \sum_{k=1}^{n} cov^4(X_k u_k, q)$$

Qannari et al., *Food Quality Preference* 11, 151-154, 2000.

ComDim / CCSWA Method

- CCSWA components are extracted according to their explained variance
  → Similar to PCA

- But **more flexibility is included** !

- Data blocks can **contribute or not** to a component
  - First components will tend to aggregate several data blocks
    → large variance

- Higher components may grasp more specific trends
  → lower variance

Use case

Nutrimouse dataset

The data sets come from a nutrigenomic study in the mouse (Martin et al., 2007) in which the effects of five regimens with contrasted fatty acid compositions on liver lipids and hepatic gene expression in mice were considered.

- observations: 40 mice

- **genotype**: 2-levels factor, wild-type (WT) and PPARα -/- (PPAR)

- **diet**: 5-levels factor. Oils used for experimental diets preparation were corn and colza oils (50/50) for a reference diet (REF), hydrogenated coconut oil for a saturated fatty acid diet (COC), sunflower oil for an Omega6 fatty acid-rich diet (SUN), linseed oil for an Omega3-rich diet (LIN) and corn/colza/enriched fish oils for the FISH diet (43/43/14)

- datasets:

- **genes**: expressions of 120 genes measured in liver cells, selected (among about 30,000) as potentially relevant in the context of the nutrition study. These expressions come from a nylon macroarray with radioactive labelling

- **lipids**: concentrations (in percentages) of 21 hepatic fatty acids measured by gas chromatography

R package MBAnalysis

MBAnalysis: Multiblock Exploratory and Predictive Data Analysis

Exploratory and predictive methods for the analysis of several blocks of variables measured on the same individuals.

Version:	2.0.2
Depends:	R (≥ 2.10)
Imports:	ggplot2, ggrepel, grDevices, stats, utils
Published:	2023-10-24
Author:	Benjamin Mahieu [aut, cre], Essomanda Tchandao Mangamana [aut], Evelyne Vigneau [aut], Veronique Cariou [aut]
Maintainer:	Benjamin Mahieu <benjamin.mahieu at oniris-nantes.fr>
License:	GPL (≥ 3)
NeedsCompilation:	no
CRAN checks:	MBAnalysis results

Documentation:

Reference manual: MBAnalysis.pdf

Downloads:

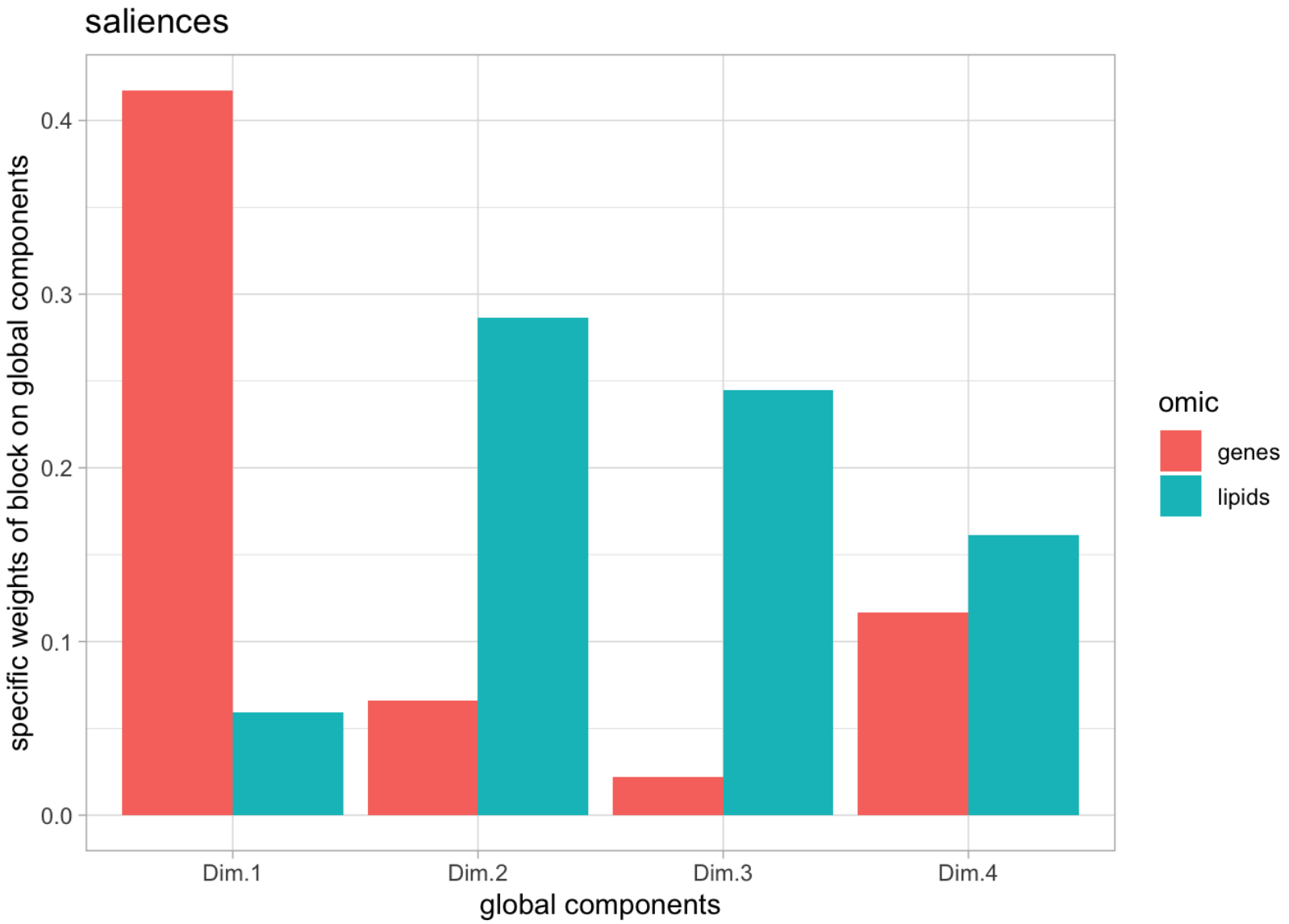
Package source:	MBAnalysis_2.0.2.tar.gz
Windows binaries:	r-devel: MBAnalysis_2.0.2.zip, r-release: MBAnalysis_2.0.2.zip, r-oldrel: MBAnalysis_2.0.2.zip
macOS binaries:	r-release (arm64): MBAnalysis_2.0.2.tgz, r-oldrel (arm64): MBAnalysis_2.0.2.tgz, r-release (x86_64): MBAnalysis_2.0.2.tgz
Old sources:	MBAnalysis archive

Linking:

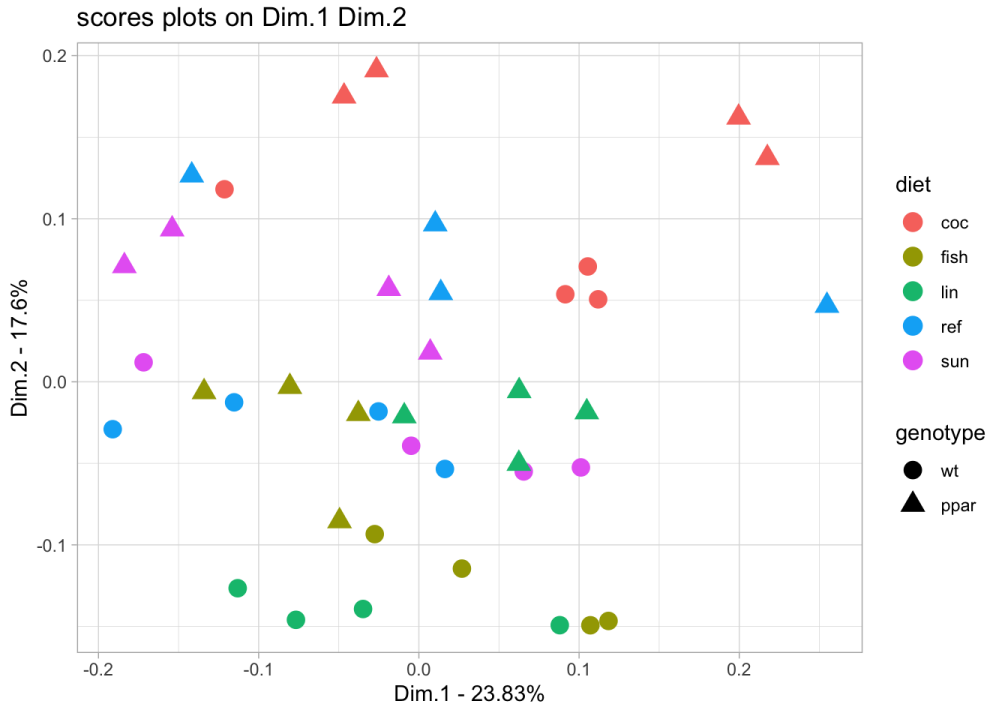
Please use the canonical form https://CRAN.R-project.org/package=MBAnalysis to link to this page.

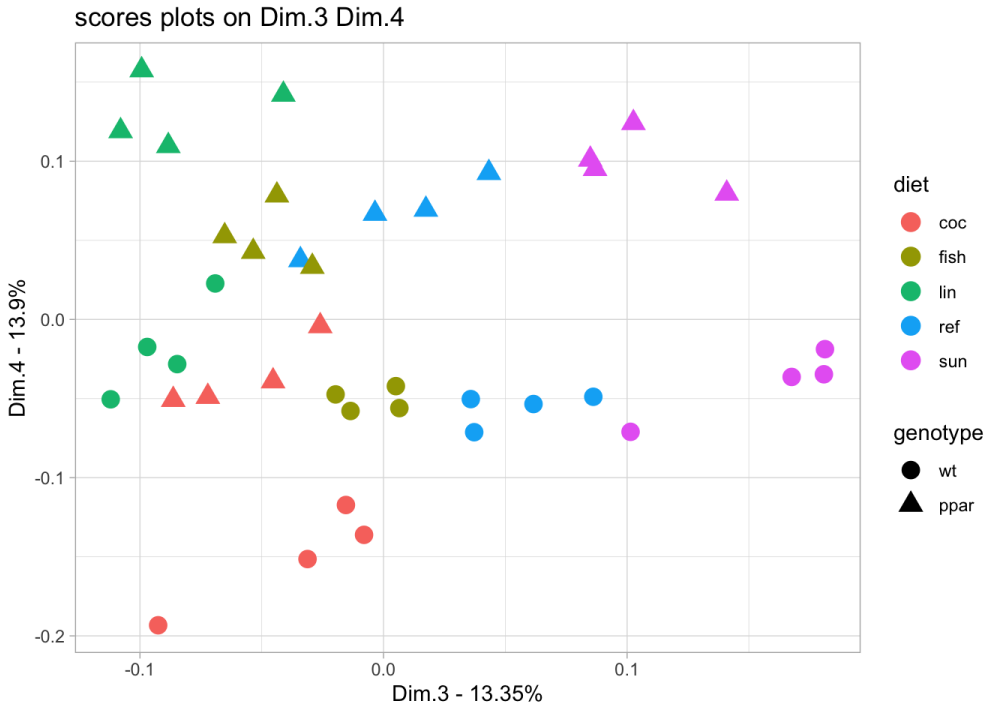
```r
# prepare dataset
ComDim_data <- cbind.data.frame(genes, lipids)
n_group <- c(dim(genes)[[2]], dim(lipids)[[2]])

# run analysis
ComDim_res <- ComDim(X = ComDim_data,
                     block = n_group,
                     name.block = c("genes", "lipids"),
                     scale = T,
                     scale.block = T)
```

How do both blocks contribute to each dimension?

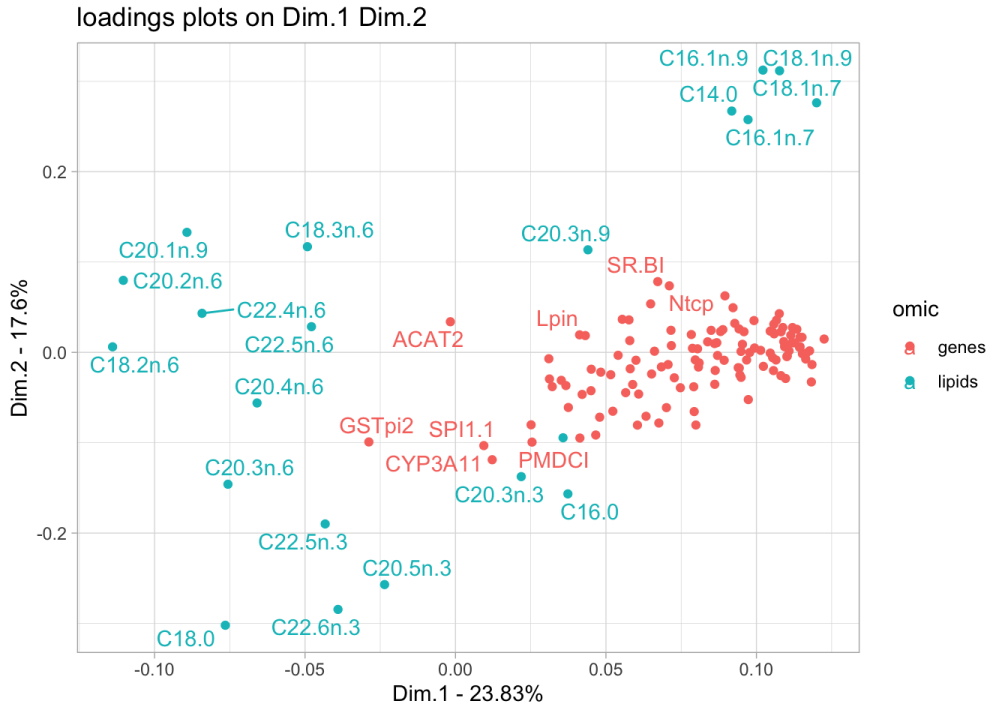


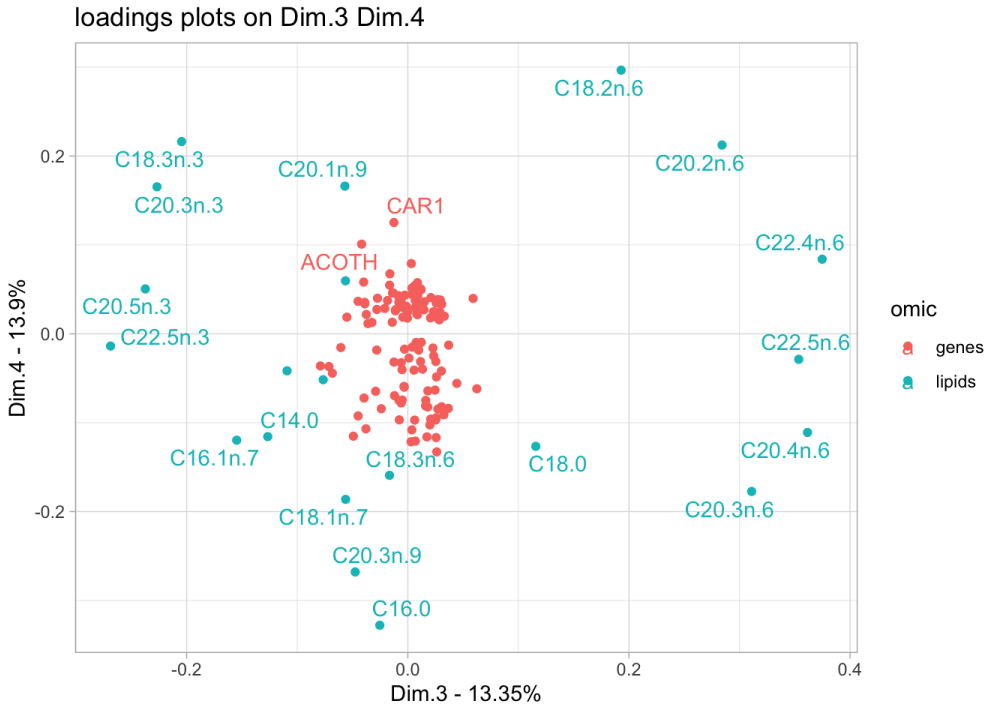
Do we observe a grouping of observations?





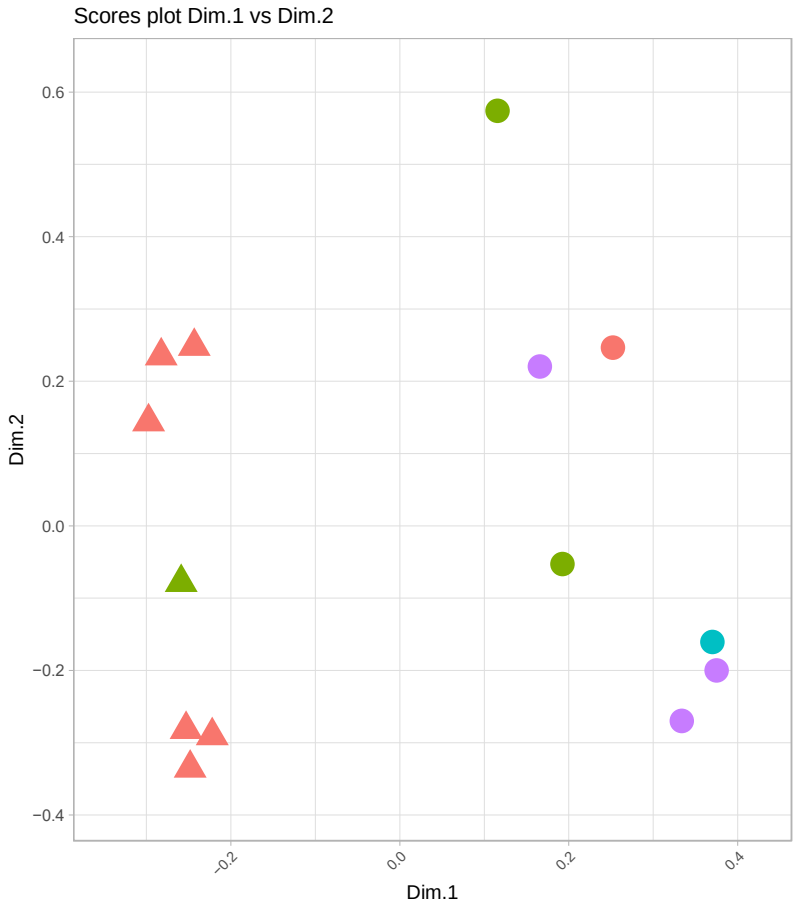
Which variables contribute to the observed grouping?



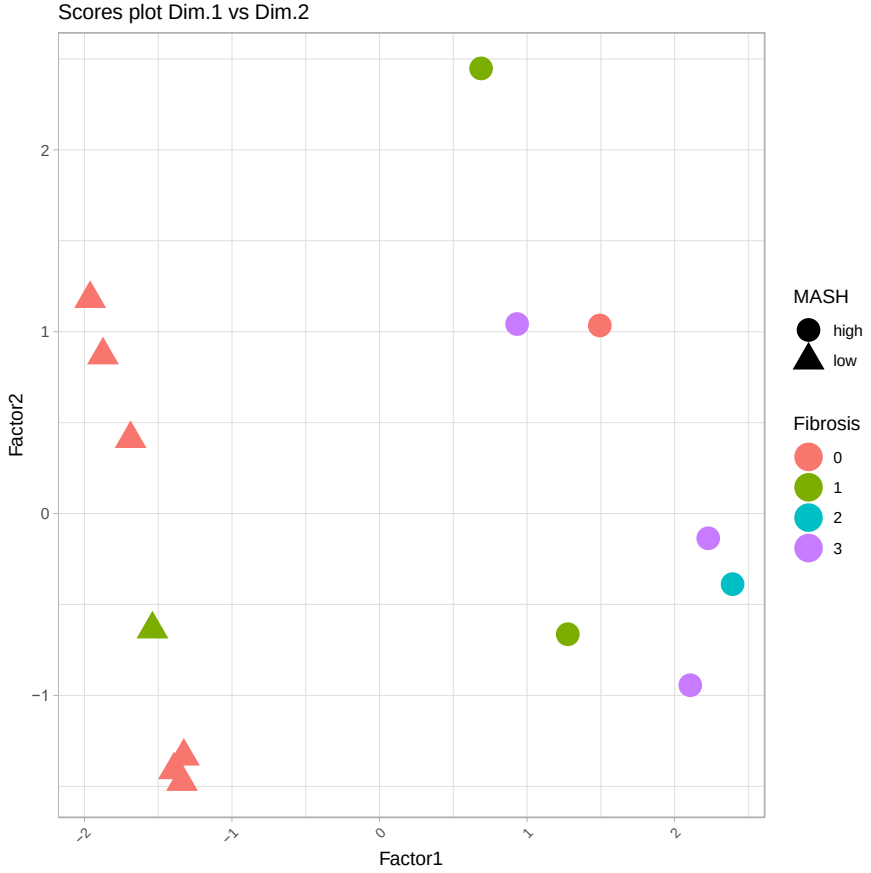


ComDim and MOFA show similar distribution of samples

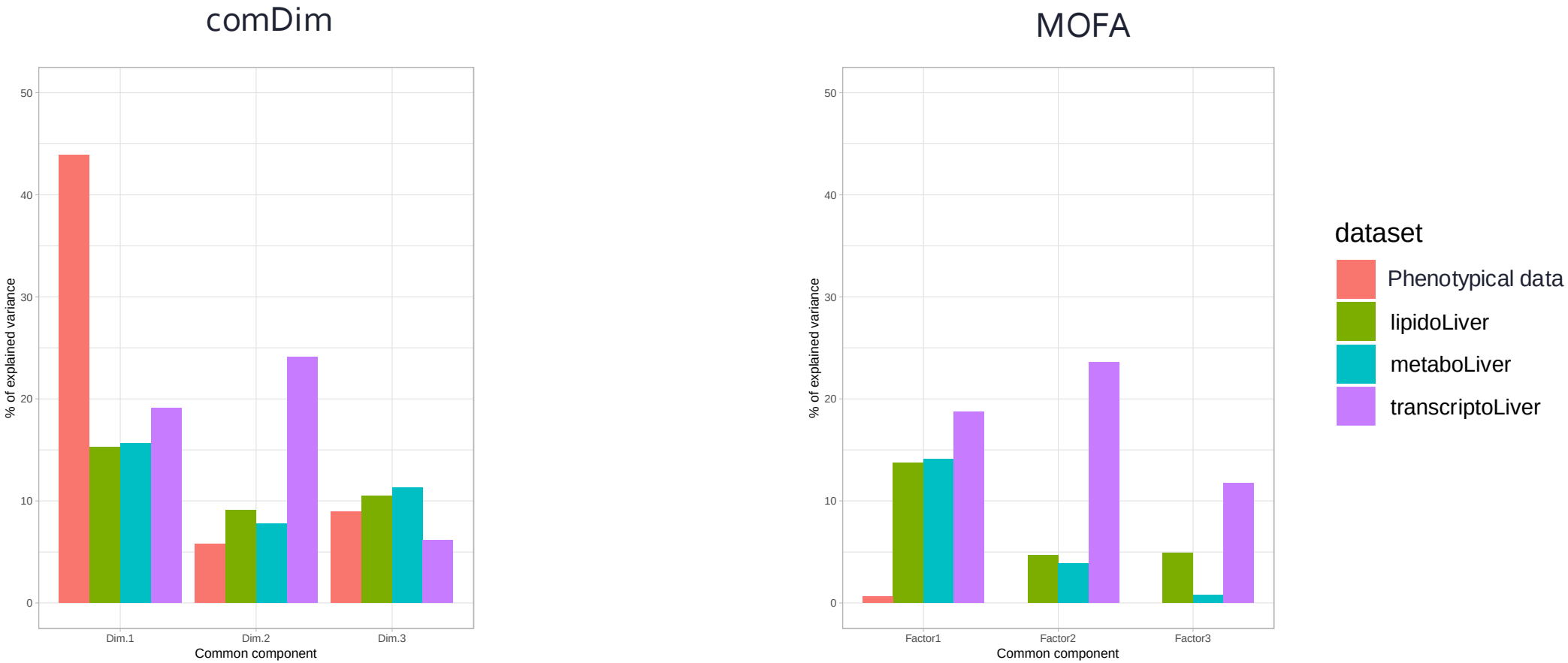
comDim



MOFA

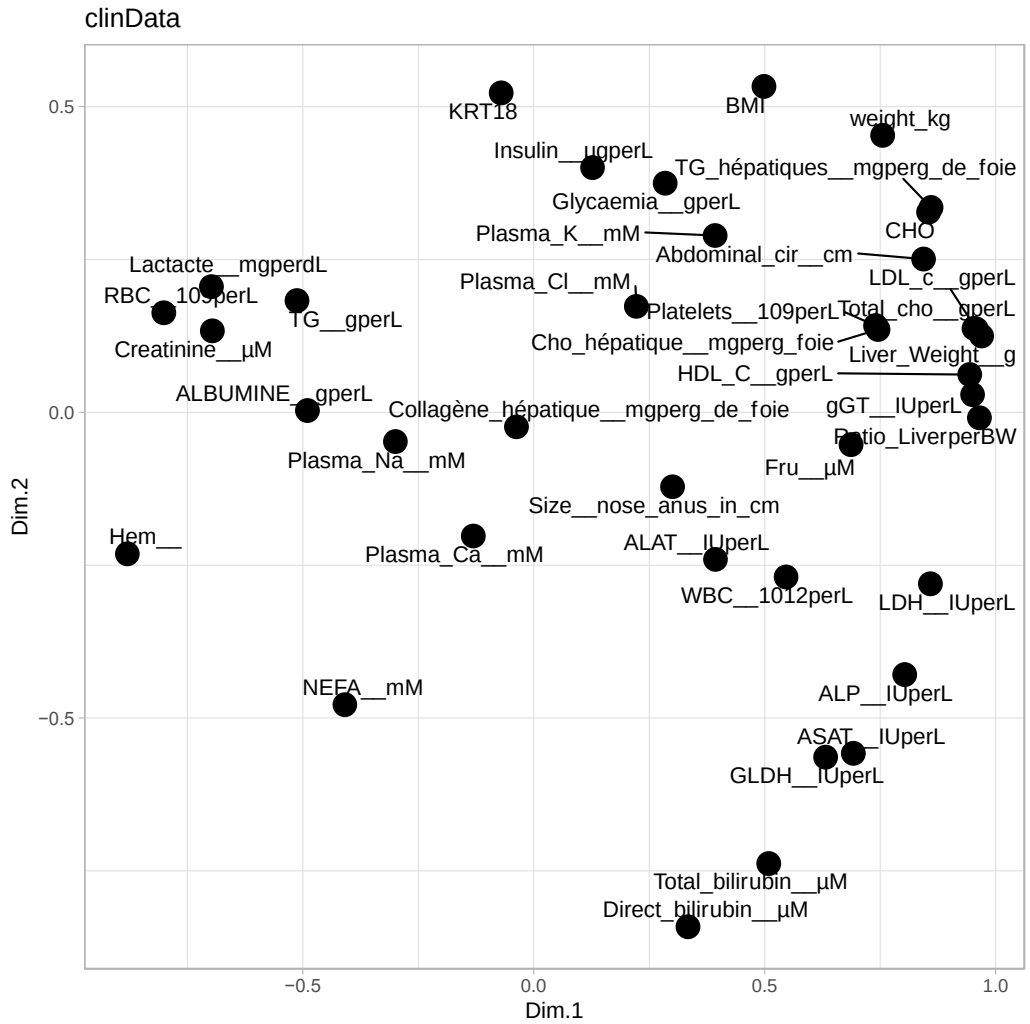


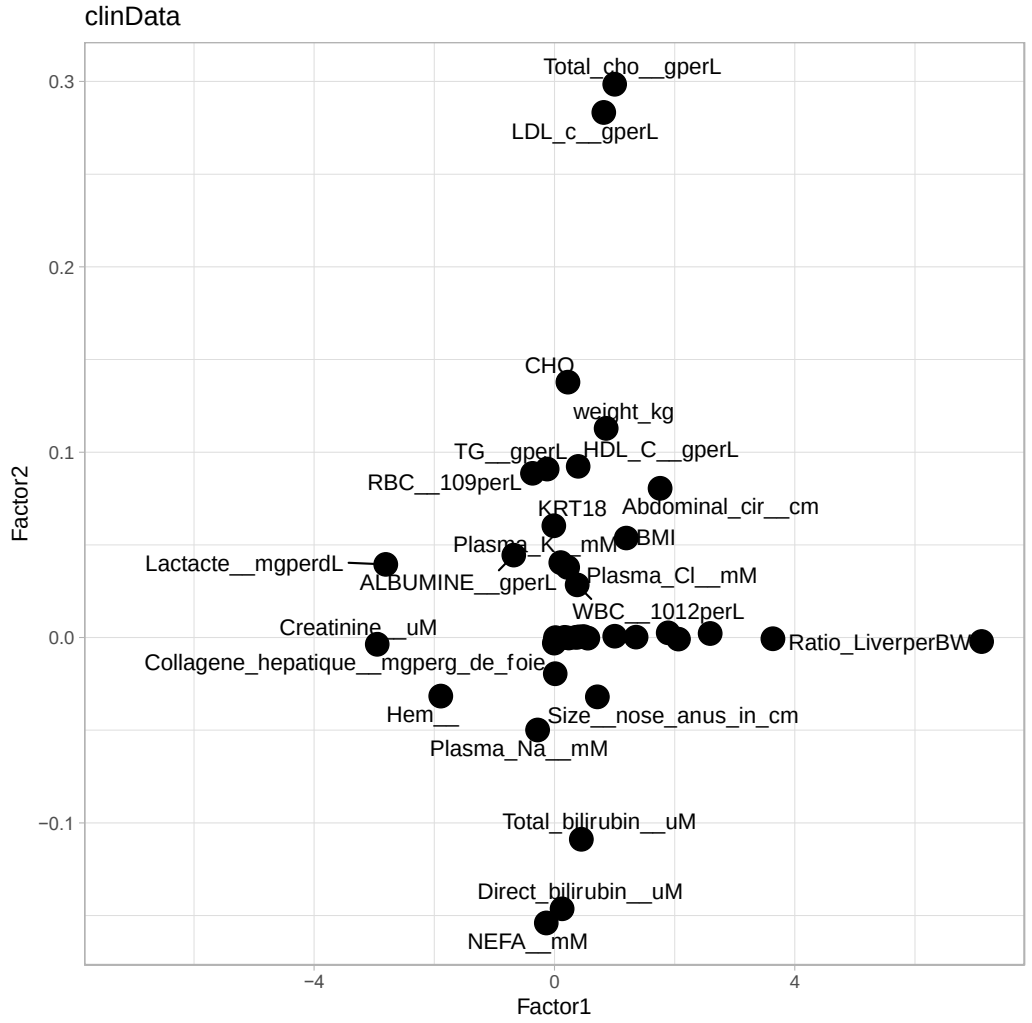
MOFA and ComDim capture differently the variance of blocks



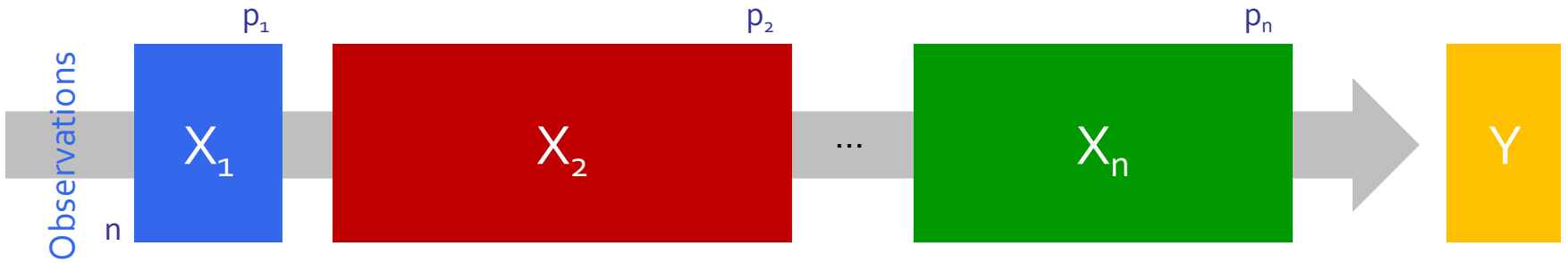
Samples phenotypical data: height, weight, liver markers, etc, but no fibrosis or nash scores (stay unsupervised)

Loadings of minipigs phenotypical data on the first two components – this could come from chosen MOFA parameters





Supervised multiblock analysis



Add another block, but with a different role in the system

- the blocks are **no longer exchangeable**

Predictive relationship

- regression approach

Some methods account for the **sequence of blocks (hierarchy)**
Some methods separate **common and distinct components** (not discussed here)

- Choosing the linking structure is an a priori decision (domain-specific)

Consensus OPLS workflow

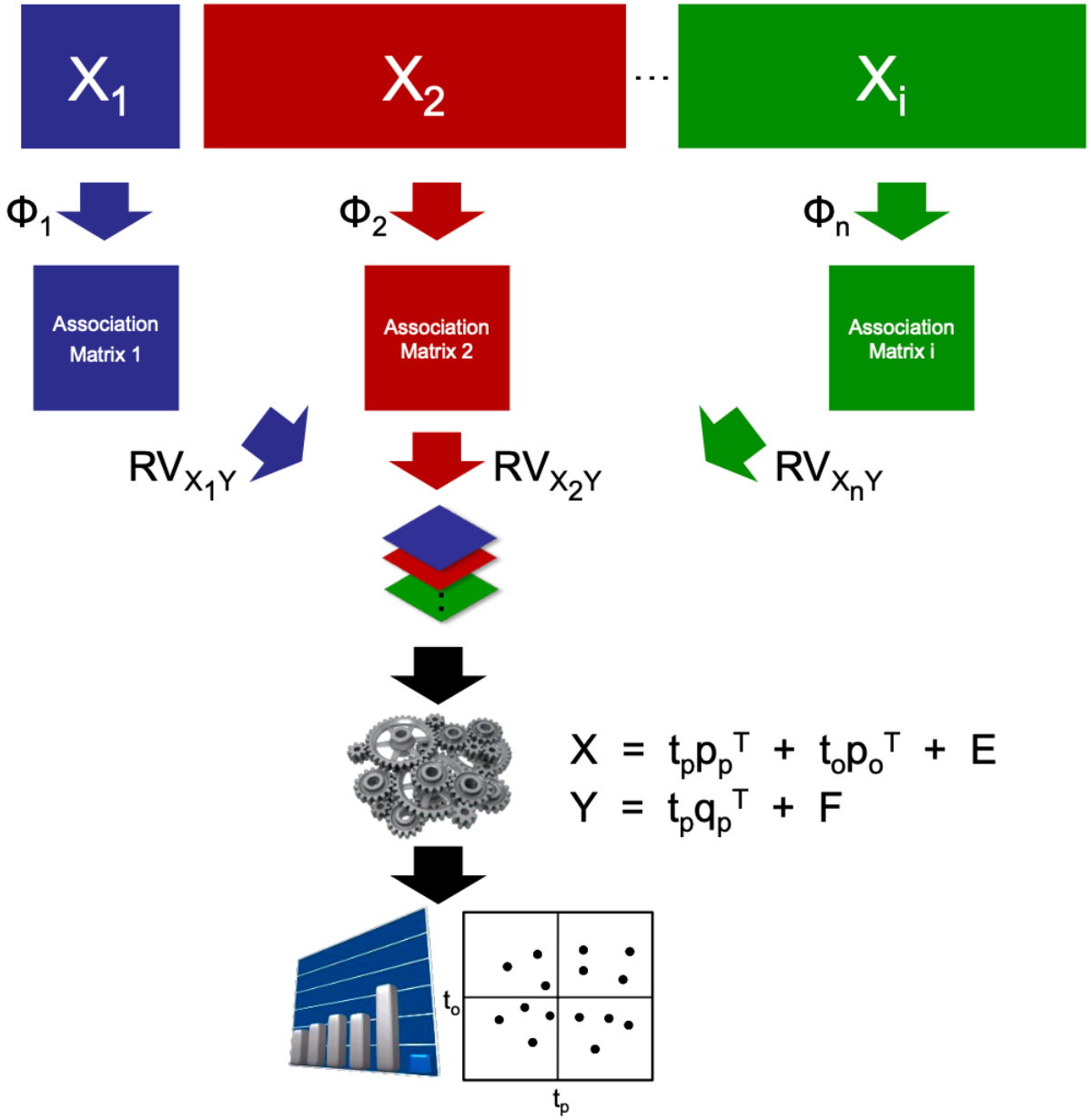
Heterogeneous data
($n \times k_i$)

Association matrices XX^T
(linear kernel, n x n)

Weighted sum matrix
(linear combination, RV coefficients)

Kernel OPLS algorithm
(dual form)

Global results & joint interpretation

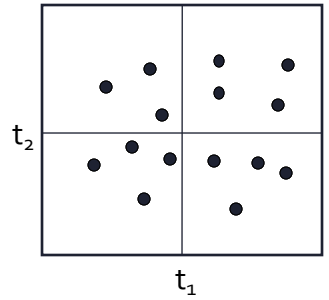


Consensus OPLS model

- **Multiblock predictive model**

New common subspace

Common/distinct component(s)

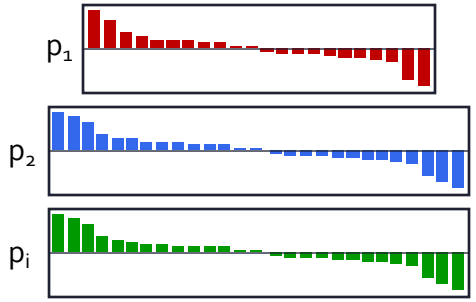


Global/Local scores

- **Consensus observations scores**

Blocks weights (contributions)

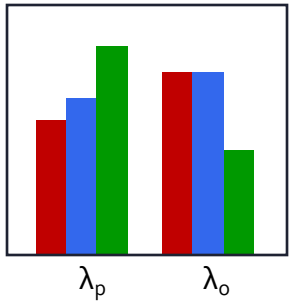
Loadings of the initial variables



Variables loadings

- **Common/specific variation(s)**

Balance between block weights



Blocks weights

- Easy interpretation

Compromise between **prediction ability** and **data description**

R package consensusOPLS

- R package under development (INRAe, SIB, UniGE)

https://cran.r-project.org/web/packages/ConsensusOPLS/index.html

- based on the original MATLAB code

https://gitlab.unige.ch/Julien.Boccard/consensusopls

Use case – Discriminate wt vs ppar samples

Nutrimouse dataset

The data sets come from a nutrigenomic study in the mouse (Martin et al., 2007) in which the effects of five regimens with contrasted fatty acid compositions on liver lipids and hepatic gene expression in mice were considered.

- observations: 40 mice

- **genotype**: 2-levels factor, wild-type (WT) and PPARα -/- (PPAR)

- **diet**: 5-levels factor. Oils used for experimental diets preparation were corn and colza oils (50/50) for a reference diet (REF), hydrogenated coconut oil for a saturated fatty acid diet (COC), sunflower oil for an Omega6 fatty acid-rich diet (SUN), linseed oil for an Omega3-rich diet (LIN) and corn/colza/enriched fish oils for the FISH diet (43/43/14)

- datasets:

- **genes**: expressions of 120 genes measured in liver cells, selected (among about 30,000) as potentially relevant in the context of the nutrition study. These expressions come from a nylon macroarray with radioactive labelling

- **lipids**: concentrations (in percentages) of 21 hepatic fatty acids measured by gas chromatography

ConsensusOPLS()

Arguments

`data`	A list of data blocks. Each element of the list must be of matrix type. Rows and columns can be identified (names), in which case this will be retained during analysis. Any pre-processing of the data (e.g. scaling) must be carried out before building the model.
`Y`	A vector, factor, dummy matrix or numeric matrix for the response. The type of answer given will condition the model to be used: a numeric vector for linear regression, a factor or dummy matrix for logistic regression or a discriminant model.
`maxPcomp`	Maximum number of Y-predictive components used to build the optimal model. Default, 1.
`maxOcomp`	Maximum number of Y-orthogonal components used to build the optimal model. Default, 5.
`modelType`	String for type of OPLS regression model, either `reg` for regression or `da` for discriminant analysis. Default, `da`.
`nperm`	Number of random permutations desired in response Y. Default, 100.
`cvType`	String for type of cross-validation used. Either `nfold` for n-fold cross-validation, where `nfold` is look up, or `mccv` for Monte Carlo cross-validation, or `mccvb` for Monte Carlo class-balanced cross-validation, where `nMC` and `cvFrac` are used. Default, `nfold`, i.e. `nMC` and `cvFrac` are ignored.
`nfold`	Number of folds performed in n-fold cross-validation. This can be set to the number of samples to perform Leave-One-Out cross validation. Default, 5.
`nMC`	An integer indicating the number of rounds performed when `cvType` is `mccv` or `mccvb`. Default, 100.
`cvFrac`	A numeric value indicating the fraction of observations from `data` used in the training set for `mccv` or `mccvb` cross-validation. Default, 4/5 = 0.8.
`kernelParams`	List of parameters for the kernel. Either `p` for polynomial kernel, which implies specifying the order of the polynomial by the `order` parameter, or `g` for Gaussian kernel. Default, `list(type='p', params = c(order=1.0))`.
`mc.cores`	Number of cores for parallel computing. Default, 1.
`verbose`	A logical indicating if detailed information (cross validation) will be shown. Default, FALSE.

Plotting fonctions

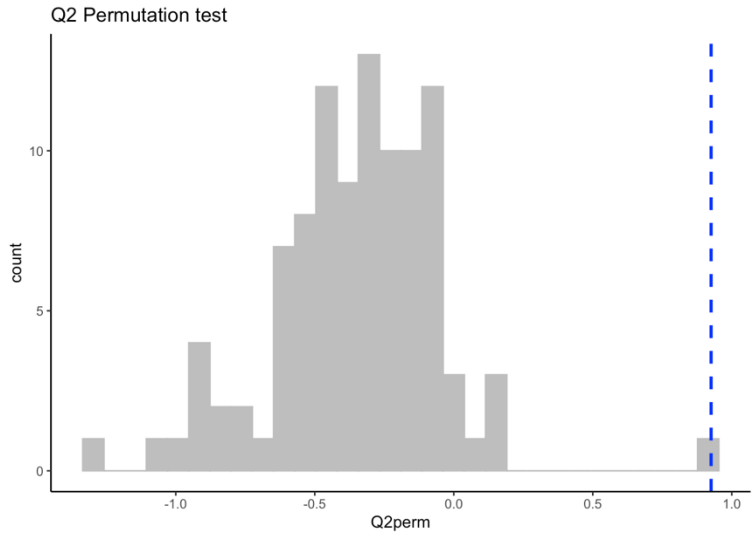
```
plotScores()
plotLoadings()
plotVIP()
plotR2()
plotQ2()
plotDQ2()
```

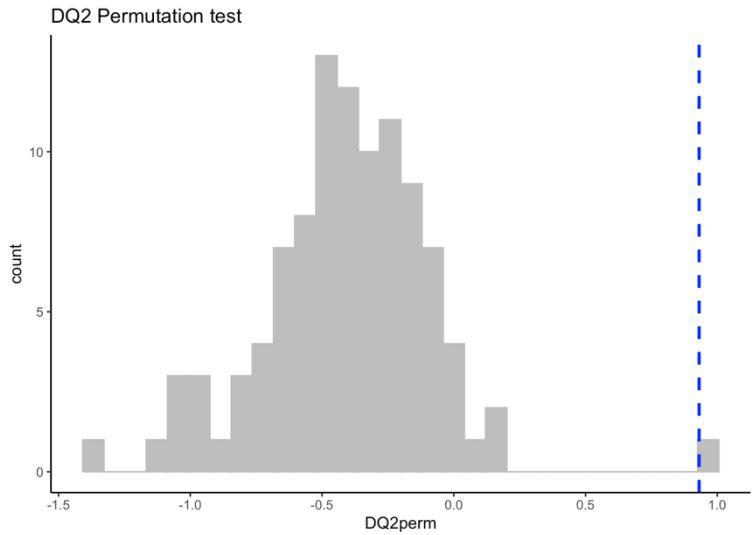
Application to the nutrimouse dataset

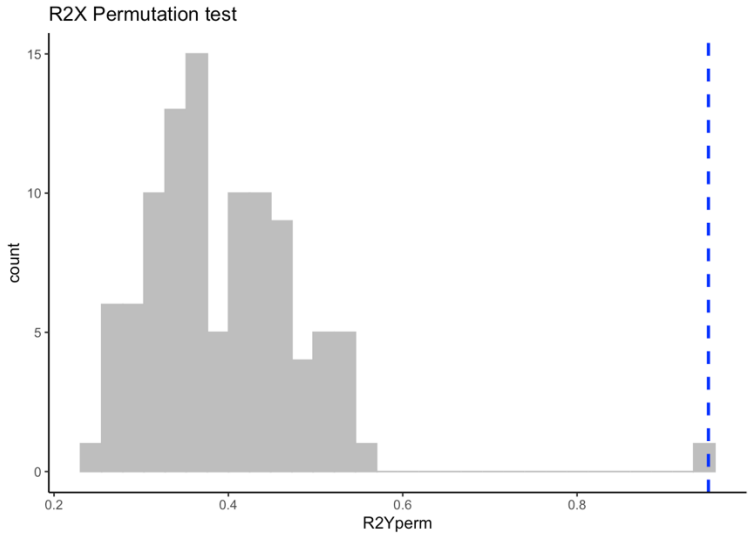
Aim: discriminate wt vs ppar samples

```
COPLS_data <- list(genes=as.matrix(genes), lipids=as.matrix(lipids))
COPLS_data <- lapply(COPLS_data, scale)
genotype <- metadata$genotype
dummy_genotype <- as.matrix(data.frame(wt = ifelse(genotype == "wt", 1, 0),ppar = ifelse(genotype == "ppar", 1, 0)))

COPLS_res <- ConsensusOPLS(
  data = COPLS_data,
  Y = dummy_genotype,
  maxPcomp = 1,
  maxOcomp = 1,
  modelType = "da",
  cvType = "nfold",
  nfold = 40,
  nperm = 100,
  verbose = T)
```

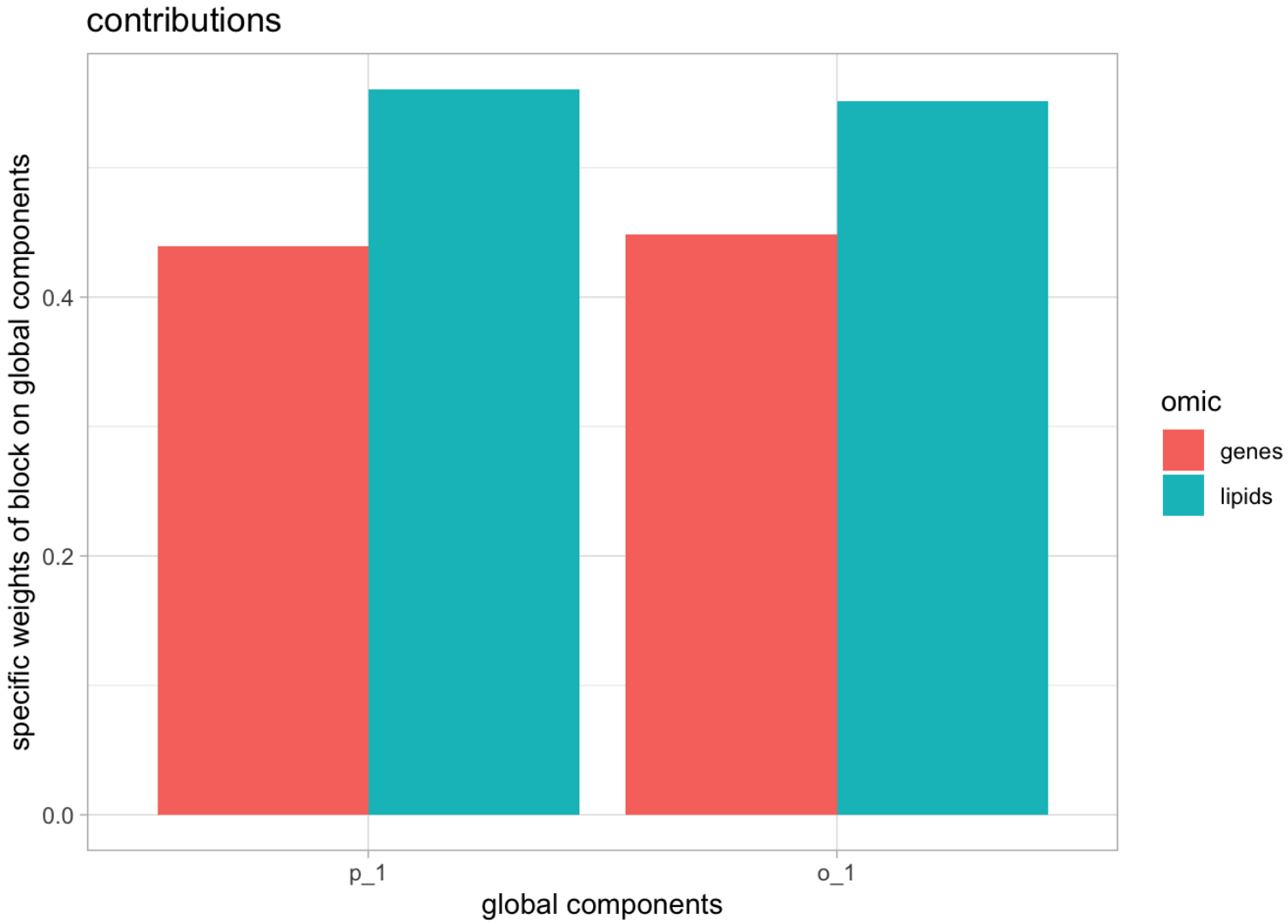
Validation of the model



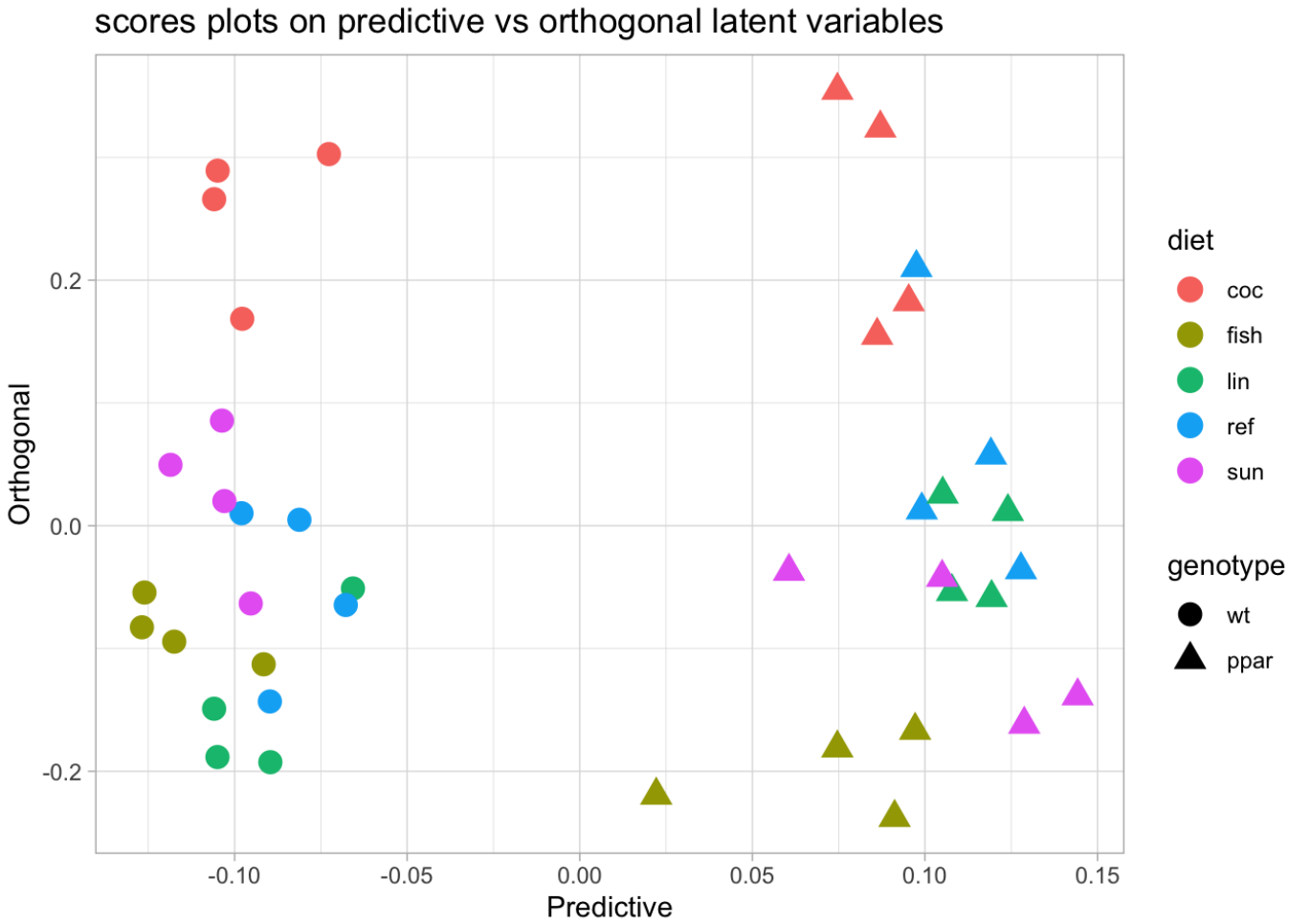




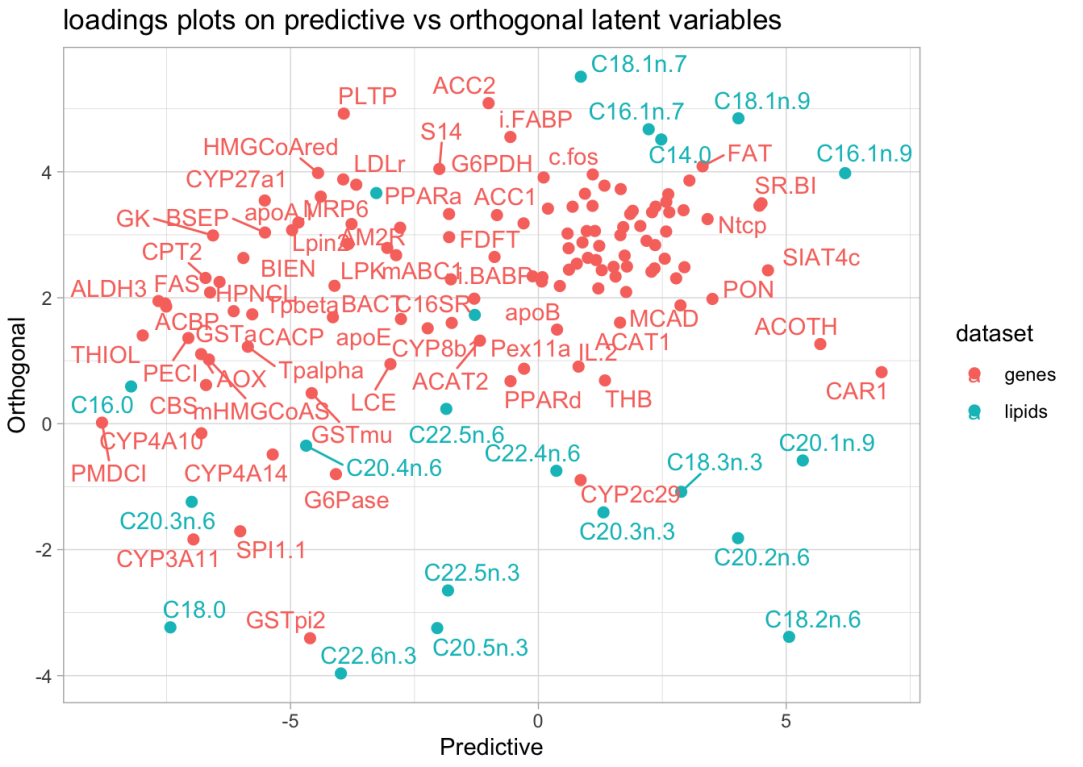
Contribution of the data blocks

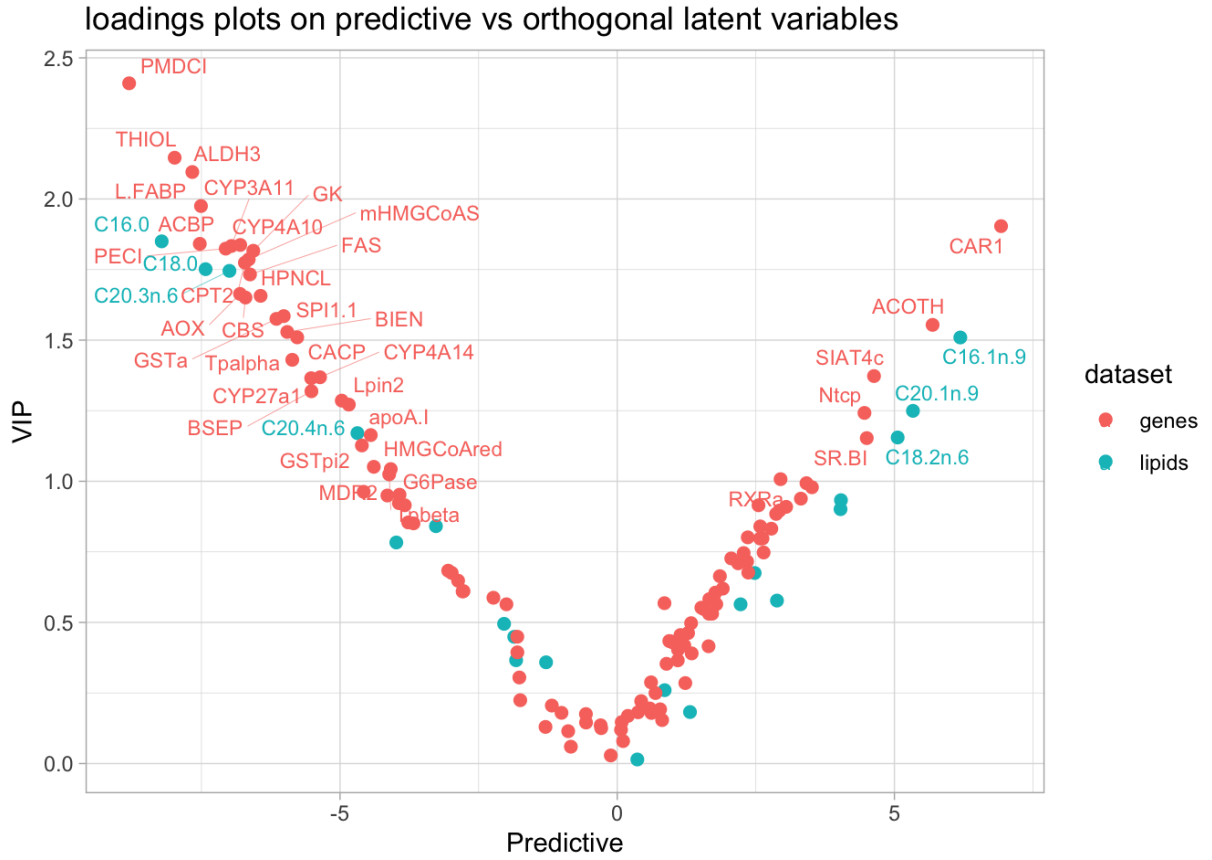


Scores plot


scores plots on predictive vs orthogonal latent variables
Orthogonal
0.2
0.0
-0.2
-0.10
-0.05
0.00
0.05
0.10
0.15
Predictive
diet
coc
fish
lin
ref
sun
genotype
wt
ppar

Contribution of the variables





Other multiblock methods

Models based on factor analysis with penalties

Experience is required for selecting adequate parameters

- **Multi-omics factor analysis (MOFA) (unsupervised)**

Model parameters estimated within a Bayesian approach with priors

Different penalties can be imposed on the weights

Probabilistic estimation (different distributions can be used for heterogeneous data)

Properties of the estimated scores and loadings are not always clear

- **Data Integration Analysis for Biomarker discovery using a Latent component method for Omics studies (DIABLO) (supervised)**

Sparse implementation of RGCCA

Lasso-type penalties on the weights

Some Take Home Messages

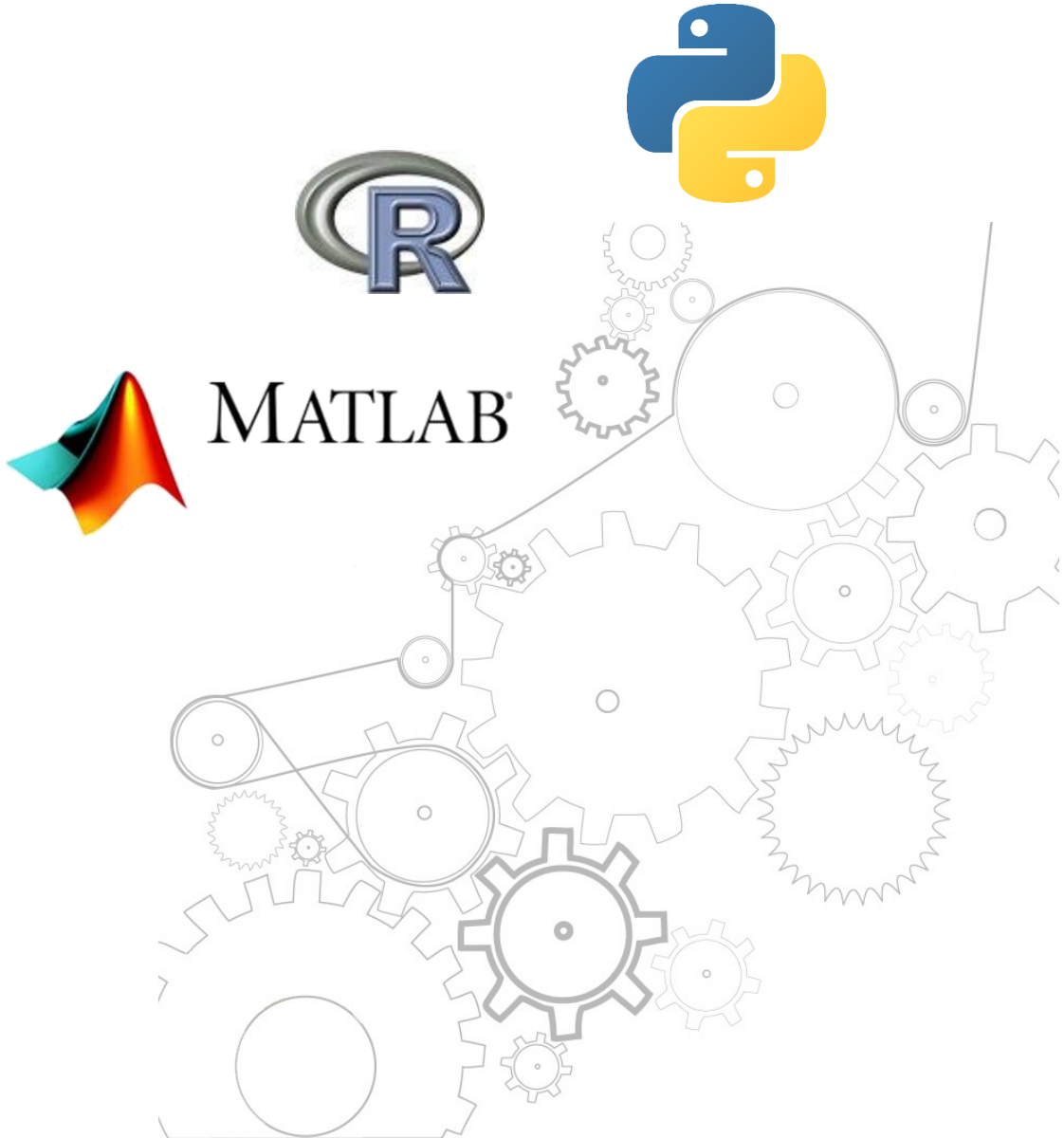
- Different data sources can be combined for a **more complete description** of complex data using two-level of interpretation (global and local)
- Dedicated chemometric methods allow **common and/or specific directions of variations** to be extracted from the data blocks
- Unsupervised multiblock analysis takes the **relationship between X_i blocks** into account
- Supervised multiblock analysis takes **the relationship with the Y response(s)** into account

Toolboxes & packages

MBAnalysis
CCA
mixOMics
ade4
omicade4
multiblock

SAISIR
MBToolbox

...many others



What to do next?

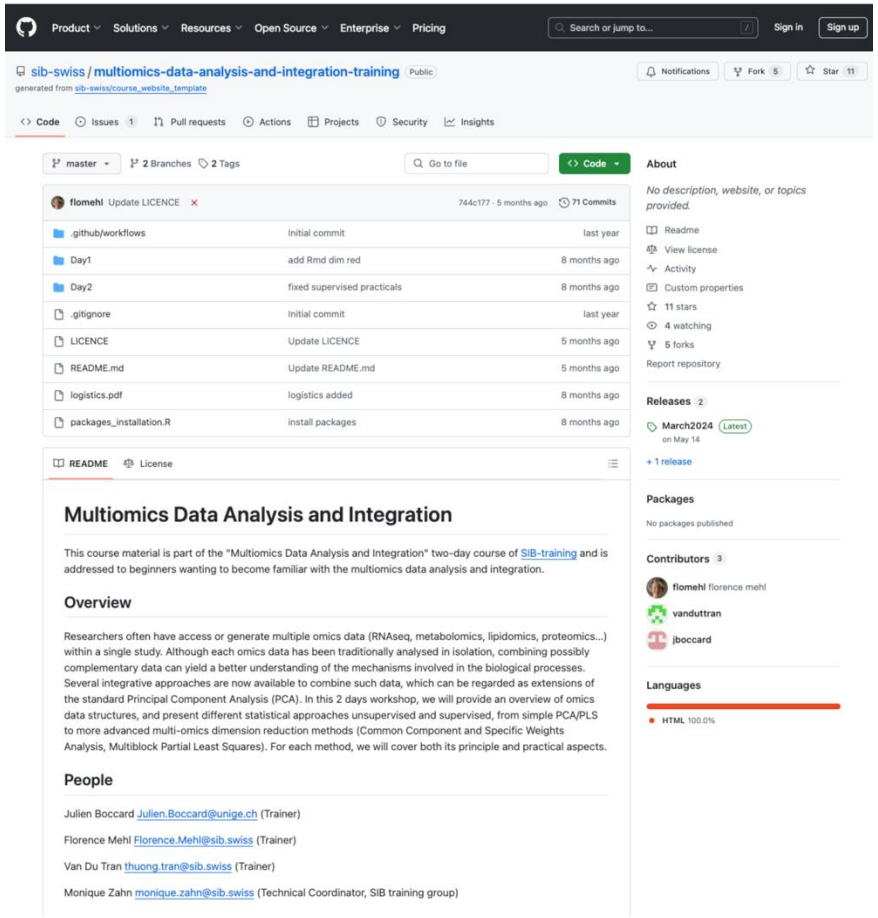
- **Put data in the biological context**
  - Need stable identifiers (https://www.metanetx.org/cgi-bin/mnxweb/id-mapper may help)
  - Perform enrichment analyses (over-representation analysis, GSEA)

https://github.com/sib-swiss/summer-school-multiomics-data-analysis-and-integration

- **Validate the obtained hypotheses**
  - Data driven knowledge discovery workflow

Additional course material

https://github.com/sib-swiss/multiomics-data-analysis-and-integration-training



Other bioinformatics courses from SIB training

SIB repository https://github.com/sib-swiss

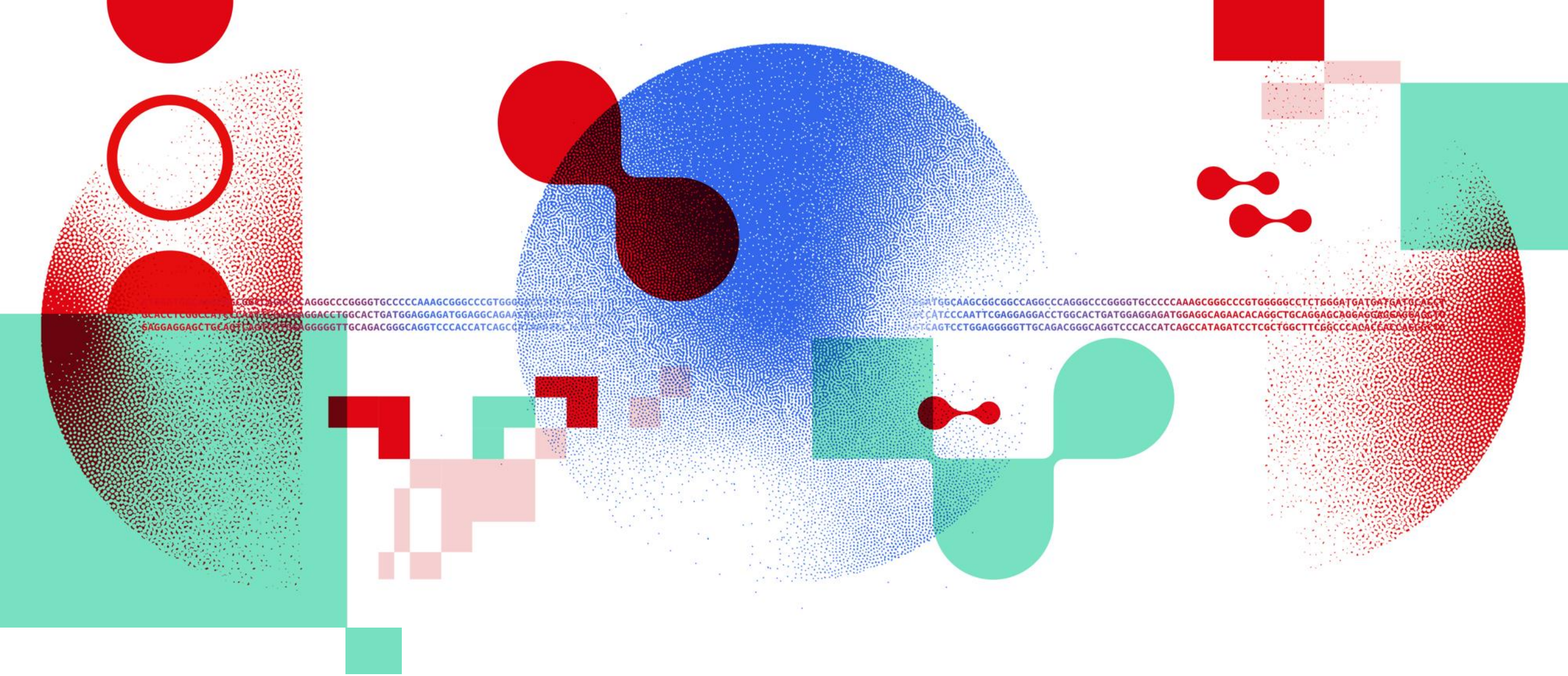
SIB course catalogue https://www.sib.swiss/training/upcoming-training-courses

Credits & Acknowledgements

Julien Boccard
Lecturer
University of Geneva – Analytical Sciences

Van Du Tran
Senior computational biologist
SIB - Vital-IT

This presentation was developed by
Julien Boccard, Van Du Tran and Florence Mehl



Thank you

DATA SCIENTISTS FOR LIFE

sib.swiss