U-BIOPRED Toolbox for Fingerprint and Handprint generation

Diane Lefaudeux
European Institute for Systems Biology & Medicine - CNRS-ENS-UCBL
Université de Lyon - France
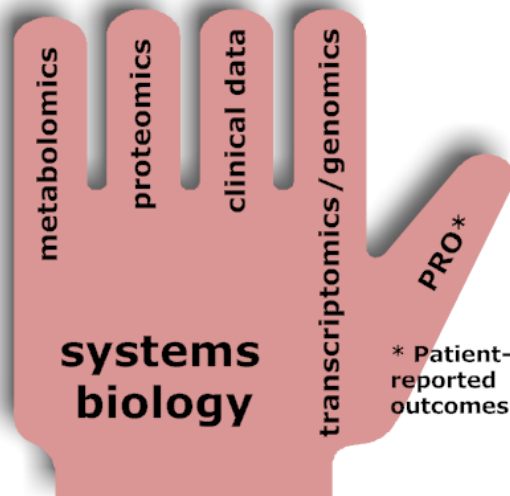**Symposium "Translating System medicine into practice"**
5th September 2014

The U-BIOPRED Project
Unbiased Biomarkers in the Prediction of Respiratory Disease Outcome
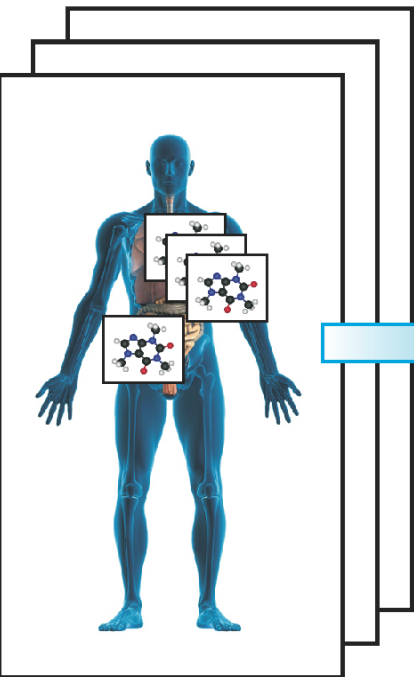
2009-2015

Coordinator: Peter Sterk
University of Amsterdam
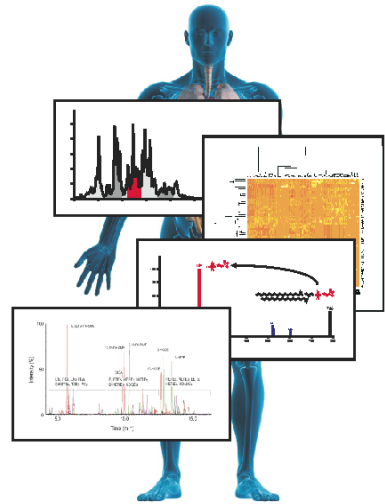
Phenotype Handprint of Severe Asthma

Patient recruitment

Severe asthmatics: Recruited from clinical centres across Europe

Sample collection
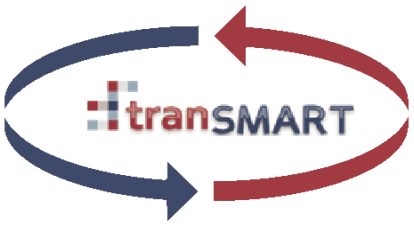
Multiple biomatrices:
- Plasma
- Sputum
- Urine

Biobank: Sample shipping & storage

'Omics data acquisition

Knowledge management platform

tranSMART

&

'Omics data integration: networks, pathway mapping, statistical analyses

U-BIOPRED

Handprint of severe asthma

U-BIOPRED

Severe asthma

- **Adult (617)**
  - Severe asthmatics (non smokers): 308
  - Severe asthmatics (smokers): 110
  - Moderate asthmatics: 98
  - Healthy controls: 101
- **Peadiatric (260)**
  - Pre-school severe asthmatics: 78
  - Pre-school mild/moderate asthmatics: 51
  - School-aged severe asthmatics: 85
  - School-aged mild/moderate asthmatics: 46

U-BIOPRED: Technology & Samples
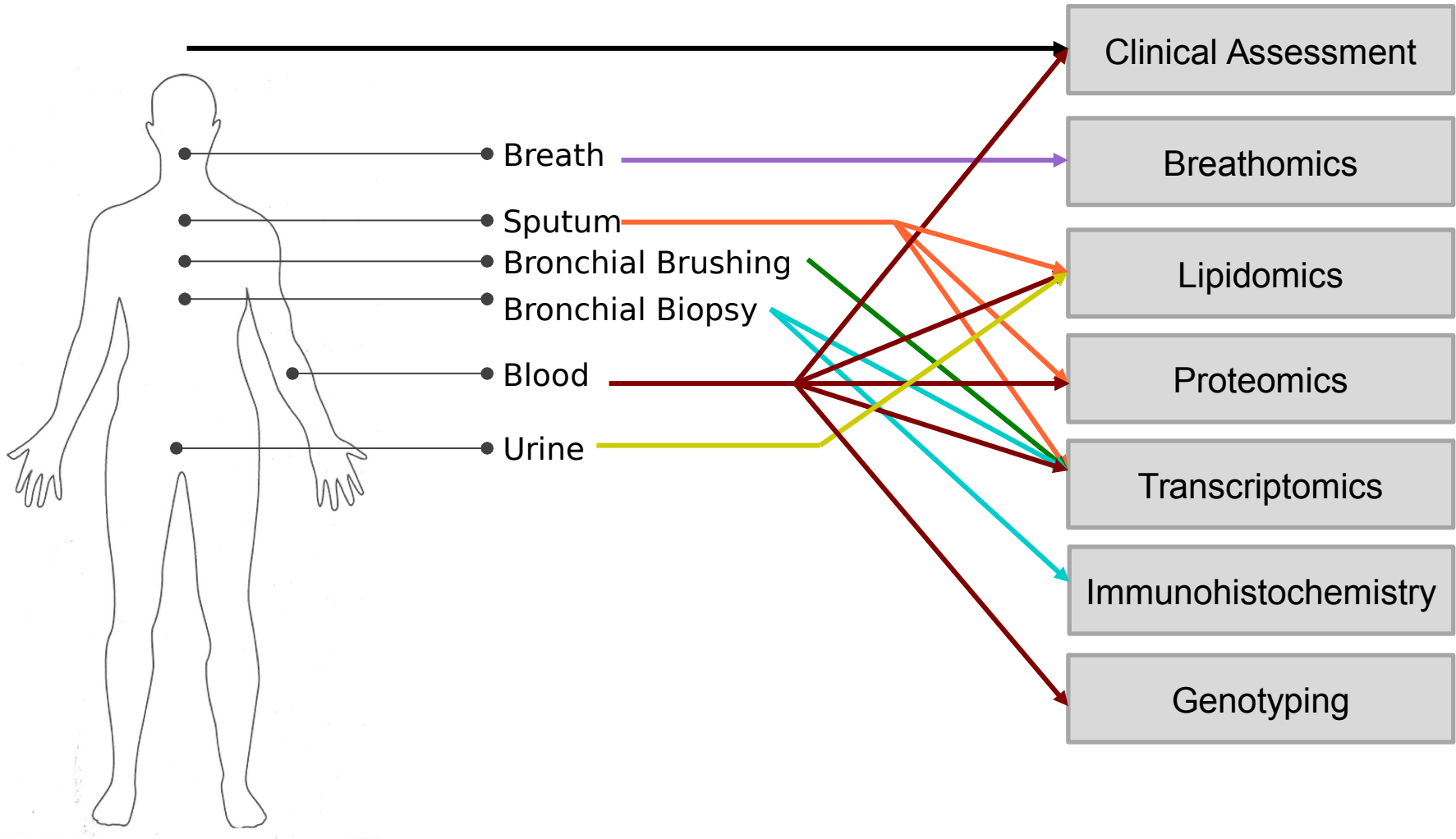
What is a "fingerprint"?

Combination of:

- Group definition (a priori or new one)

- Predictors – Biomarkers for 1 type of data
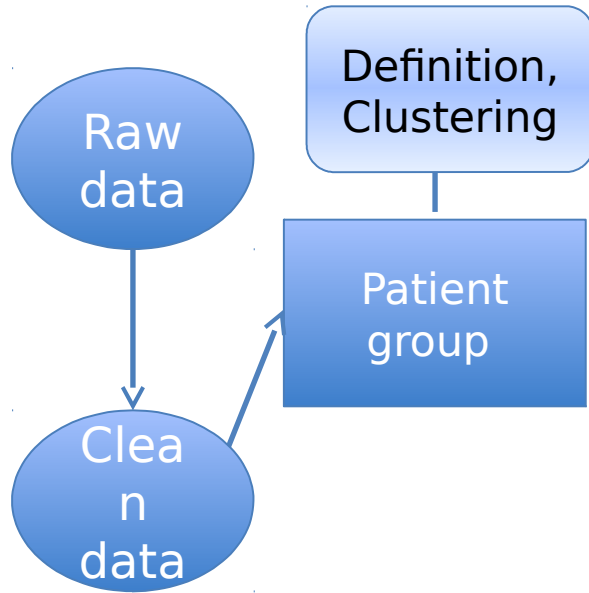
What is a "handprint"?

Combination of:

- Multiple fingerprints

OR

- Group definition (a priori or new one)
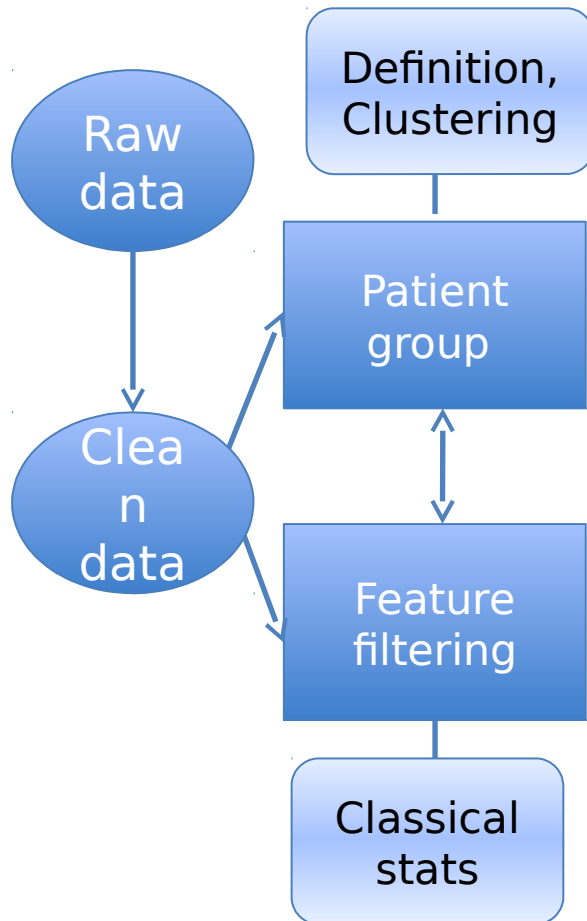- Integrated multi-omics predictors-biomarkers

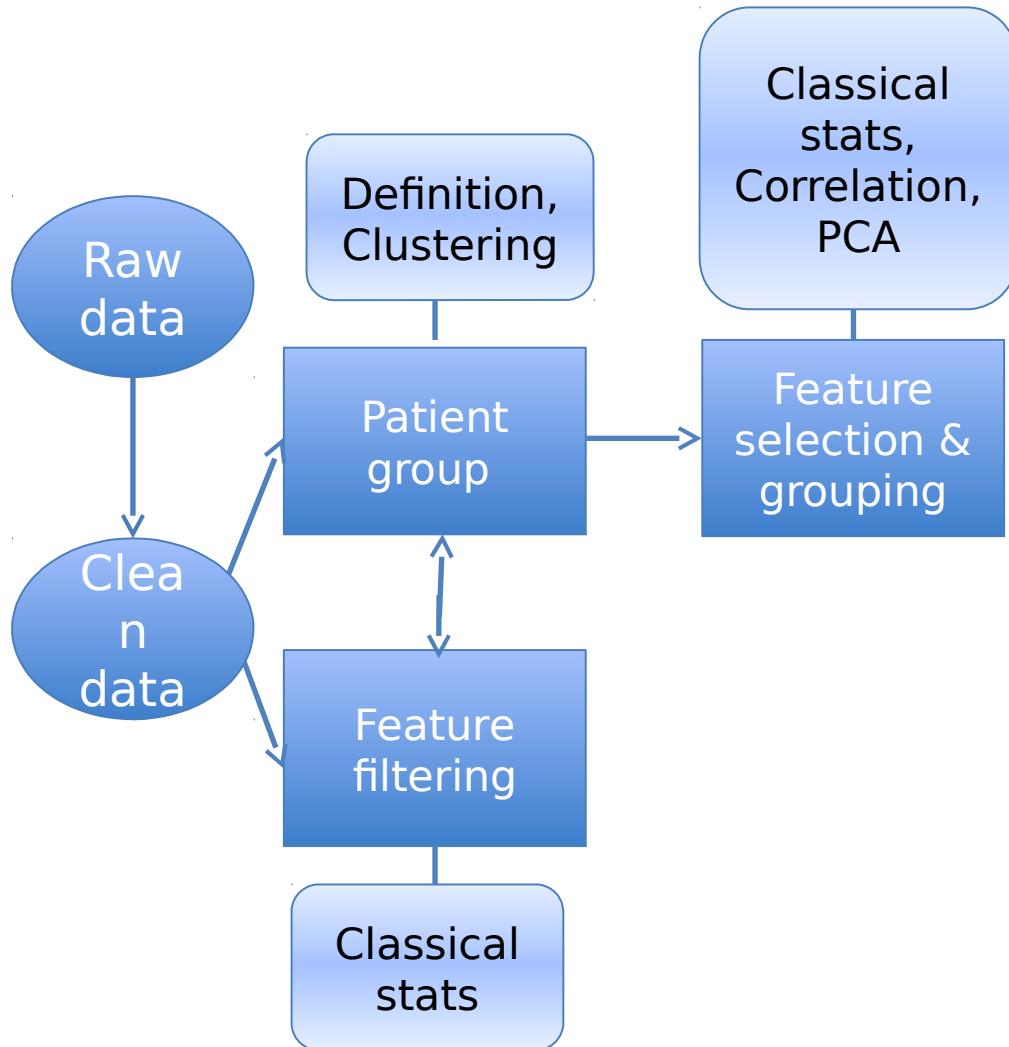# "Fingerprint" and "handprint" generation workflow

Raw data

Clean data

Definition, Clustering

Patient group

We can use clean data from each platform to find clusters

"Fingerprint" and "handprint" generation workflow

Raw data

Definition, Clustering

Clean data

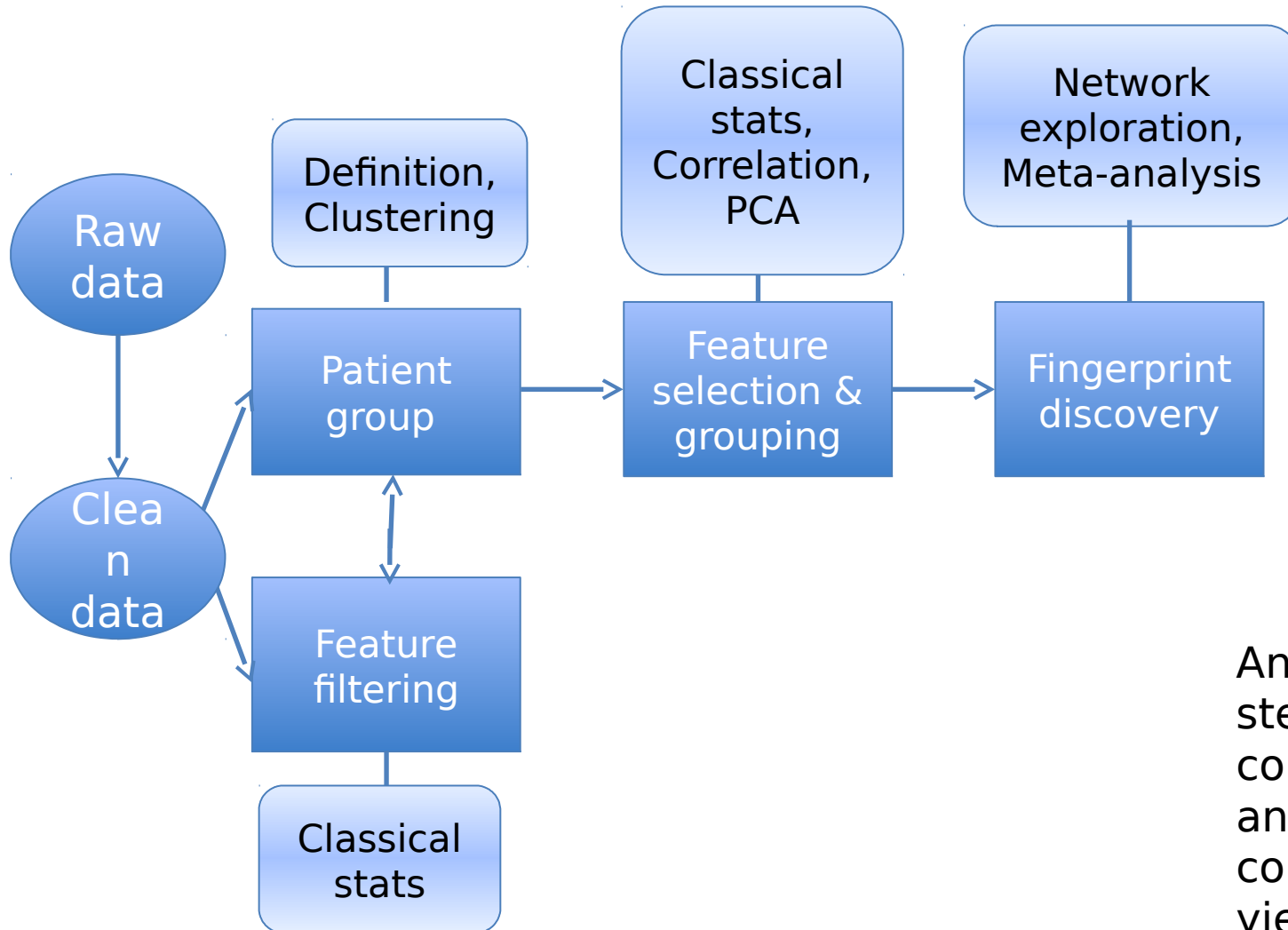**Patient group**

Feature filtering

Classical stats

Or we can add a step of feature filtering before searching for clusters (dependent on the plaform)
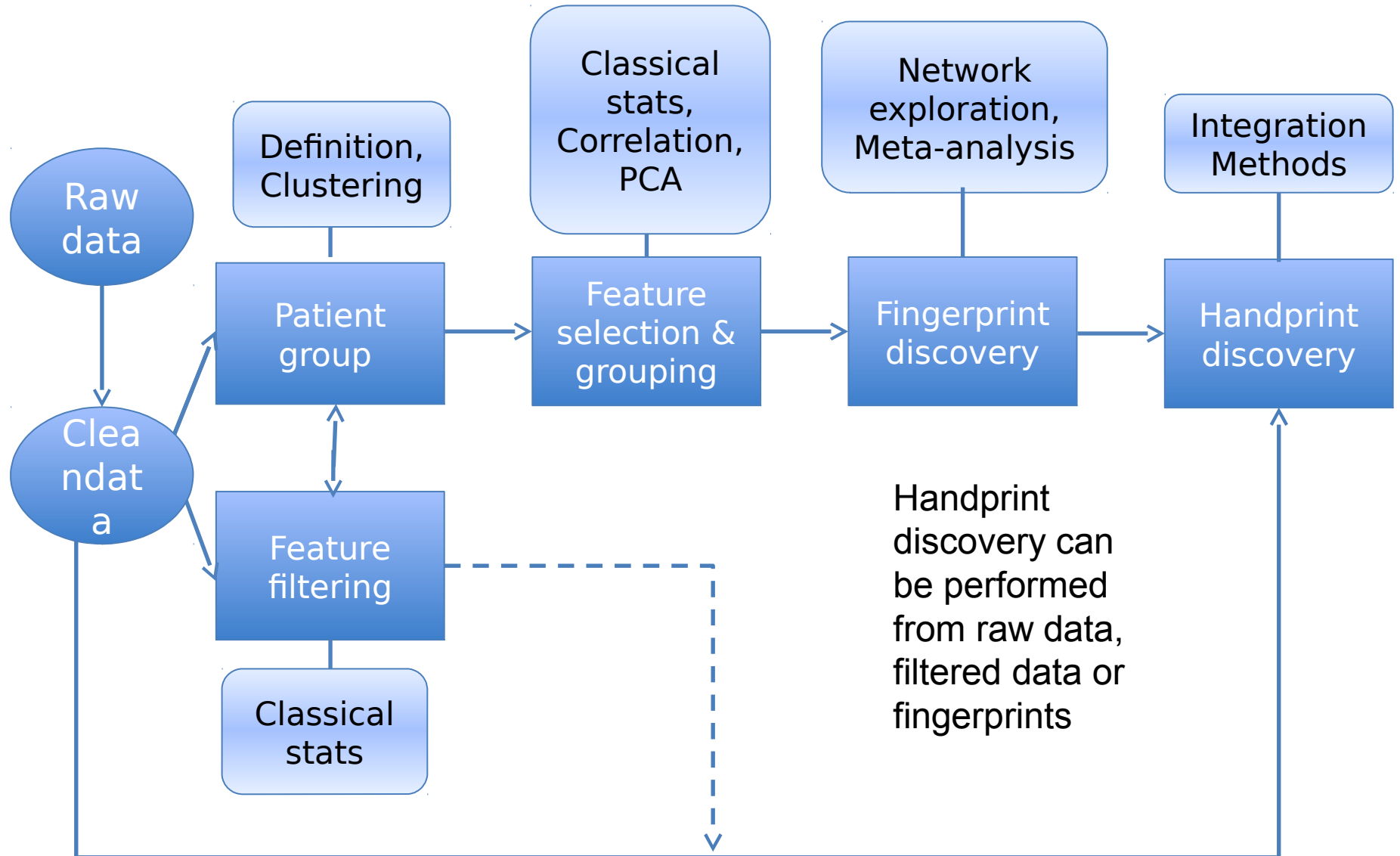
"Fingerprint" and "handprint" generation workflow

Classical stats, Correlation, PCA

Definition, Clustering

Raw data

Clean data

Patient group

Feature selection & grouping

Feature filtering

Classical stats

Once we have patient groups, we can look for features differentiating them

"Fingerprint" and "handprint" generation workflow

Raw data

Clean data

Definition, Clustering

Patient group

Feature filtering

Classical stats

Classical stats, Correlation, PCA

Feature selection & grouping

Network exploration, Meta-analysis

Fingerprint discovery

An additional step of completion and/or comparisons yields fingerprints

"Fingerprint" and "handprint" generation workflow

Raw data

Clean data

Definition, Clustering

Patient group

Classical stats, Correlation, PCA

Feature selection & grouping

Network exploration, Meta-analysis

Fingerprint discovery

Integration Methods

Handprint discovery

Feature filtering

Classical stats

Handprint discovery can be performed from raw data, filtered data or fingerprints

"Fingerprint" generation: blood transcriptomics

Biased:

Group definiton: Cases vs Control
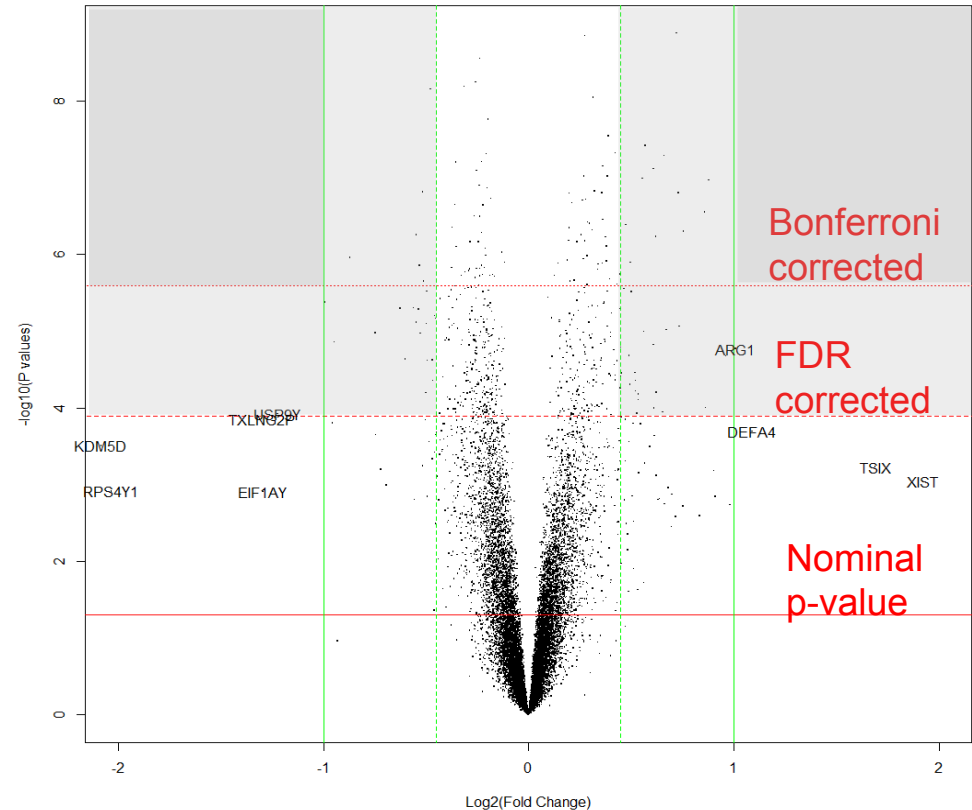**Biomarkers:** Linear Model for MicroArray data (LIMMA) Analysis
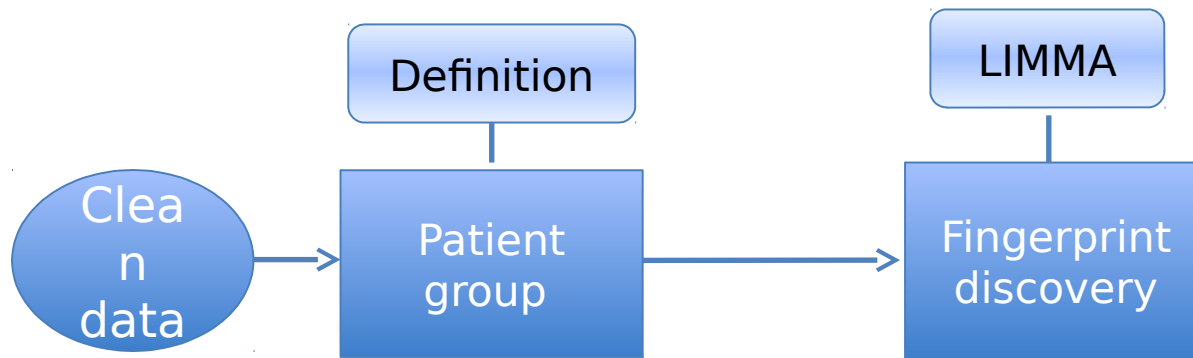
"Fingerprint" generation

Group Definition:
Severe Asthma vs Healthy

LIMMA analysis to find differentially expressed genes (DEGs):
- $|\log(\text{Fold Change})| > 1$
- Bonferroni corrected p-values < 0.05
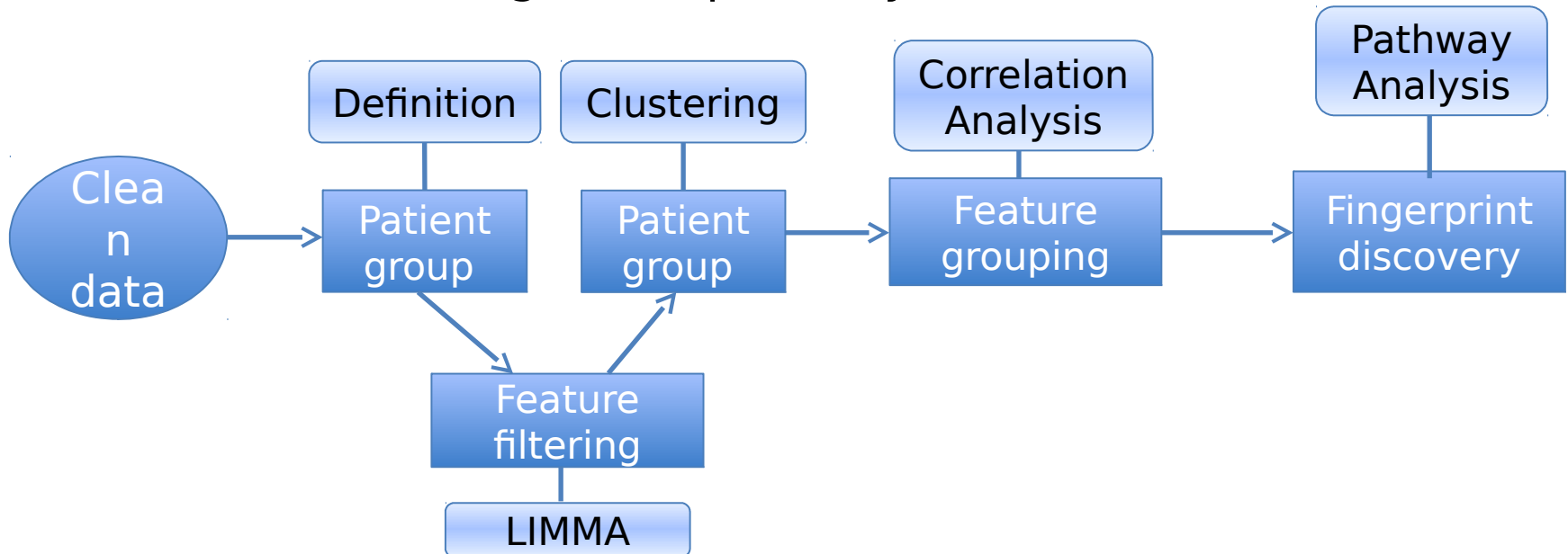
"Fingerprint" generation: blood trancriptomics

Semi-biased:

    Feature Selection: Non stringent (according to a priori group definition)

    Group Definition: Clustering of patients according to their gene profile

    Feature grouping: Correlation Analysis

    Biomarker: Disregulated pathway
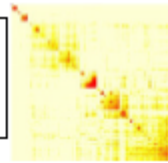
"Fingerprint" generation: blood trancriptomics

Correlation Analysis:
WGCNA: Weighted Gene Correlation Network Analysis

Construct a gene co-expression network
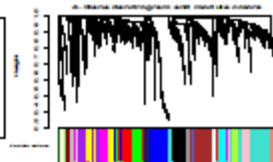Rationale: make use of interaction patterns among genes
Tools: correlation as a measure of co-expression

Identify modules
Rationale: module (pathway) based analysis
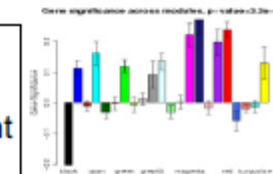Tools: hierarchical clustering, Dynamic Tree Cut

Relate modules to external information
Array Information: clinical data, SNPs, proteomics
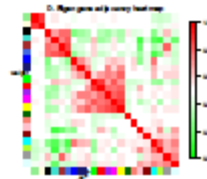Gene Information: ontology, functional enrichement
Rationale: find biologically interesting modules

Study module relationships
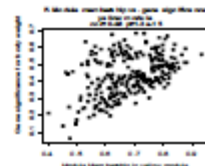Rationale: biological data reduction, systems-level view
Tools: Eigengene Networks

Find the key drivers in *interesting* modules
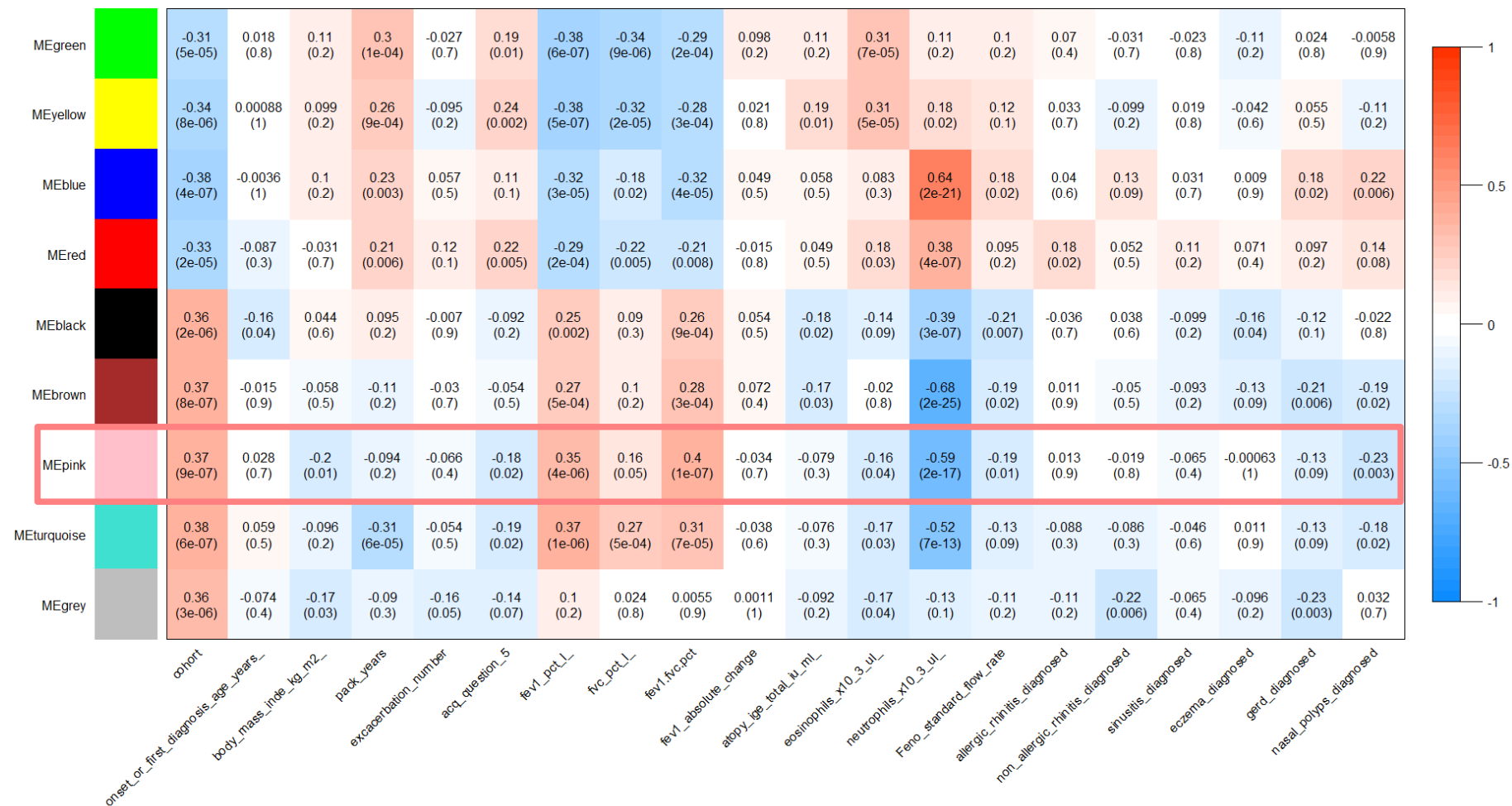Rationale: experimental validation, biomarkers
Tools: intramodular connectivity, causality testing

"Fingerprint" generation: blood trancriptomics

WGCNA

Module-Trait relationships

"Fingerprint" generation: blood trancriptomics

Pathway Analysis: Genetrail

Module	#Genes	Category	Subcategory	Expected	Observed	FDR
Pink	53	KEGG	B-Cell receptor signaling pathway	0.1942	7	4.02E-9
		GO	Immune response	1.5081	9	0.00261
		KEGG	Asthma	0.0803	2	0.01640
		KEGG	Hematopoïetic cell lineage	0.2279	2	0.02976
		GO	Signal transducer activity	2.8597	9	0.03464
		GO	Plasma membrane	6.1922	14	0.03464

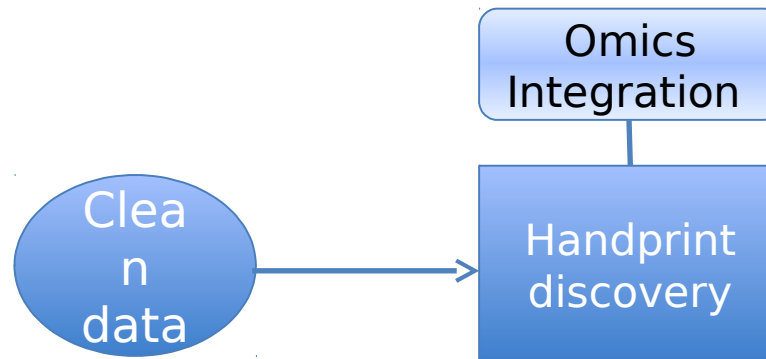# "Handprint" generation

Completely unbiased:

 Integrate different omics
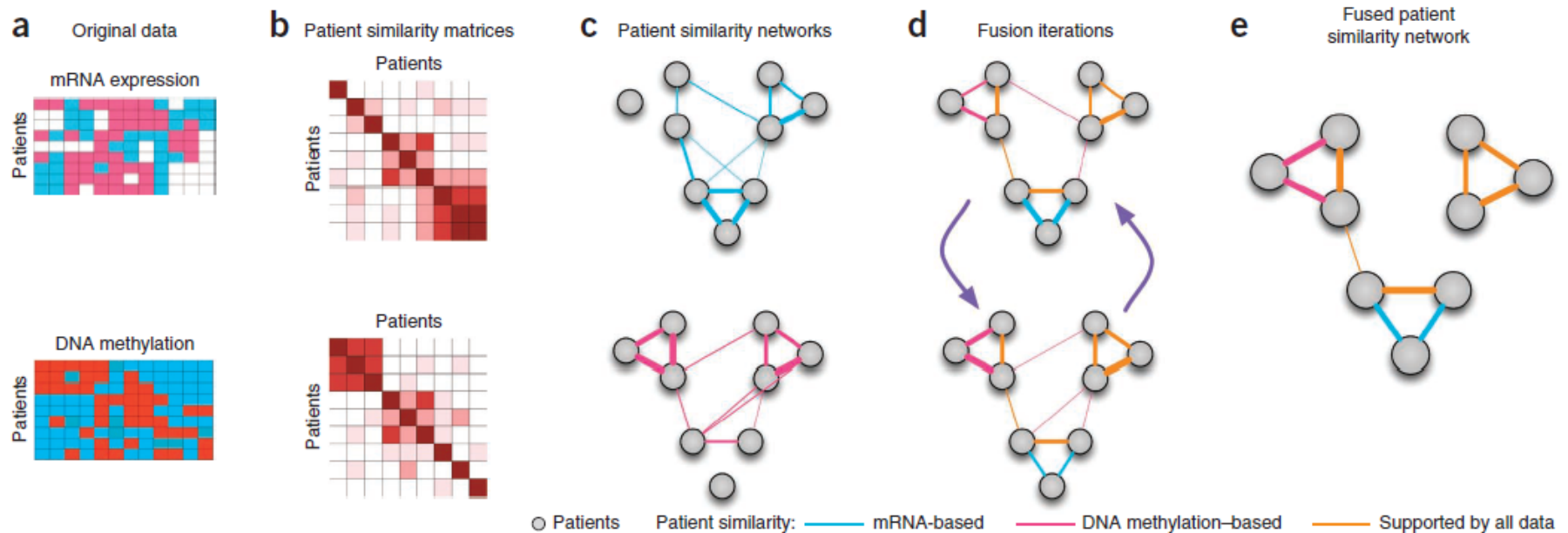 Subset of patients having all omics platform data
 Cluster according to unfiltered cleaned omics data
 Find Predictors

Omics Integration

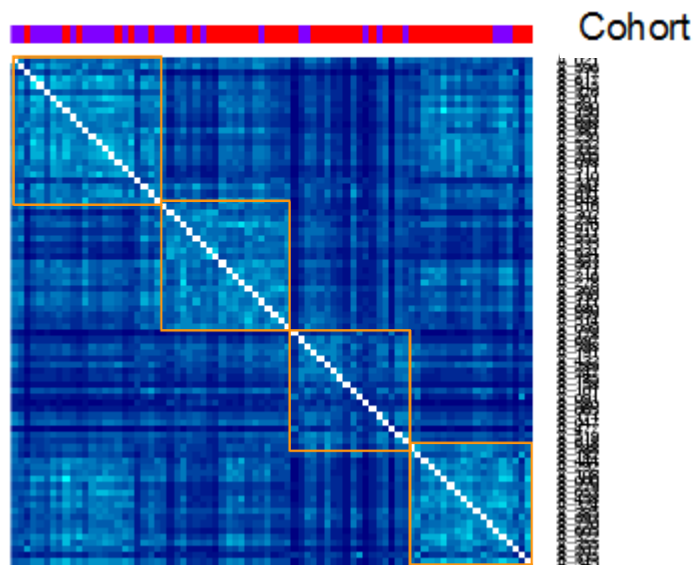Clean data → Handprint discovery

Similarity Network Fusion (SNF)

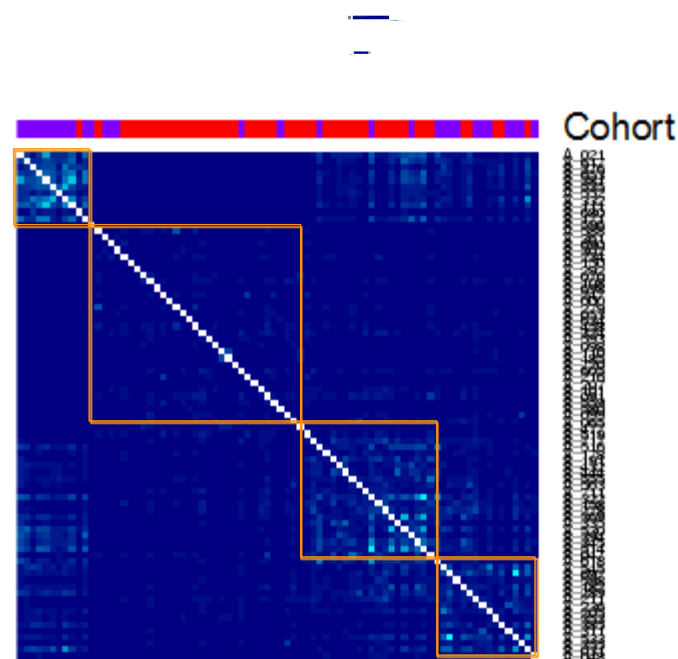# "Handprint" generation

Each platform alone (Severe asthmatics and Healthy controls)
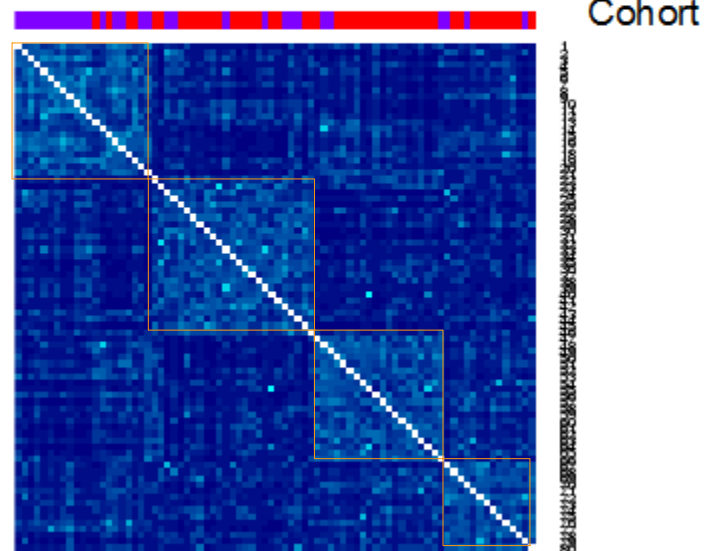

Peripheral Blood Transcriptomics

Serum Proteomic

"Handprint" generation

Similarity Network Fusion (Severe Asthmatics and healthy controls)

Peripheral Blood Transcriptomics and Serum Proteomics

Conclusion

Flexible Workflow to generate "fingerprints" and "handprints"

From biased to completely unbiased analysis

Need in any case validation of results

**Thank you,
Questions ?**