Identification of consensus whole blood transcriptomic gene modules in primary Sjogren's Syndrome patients

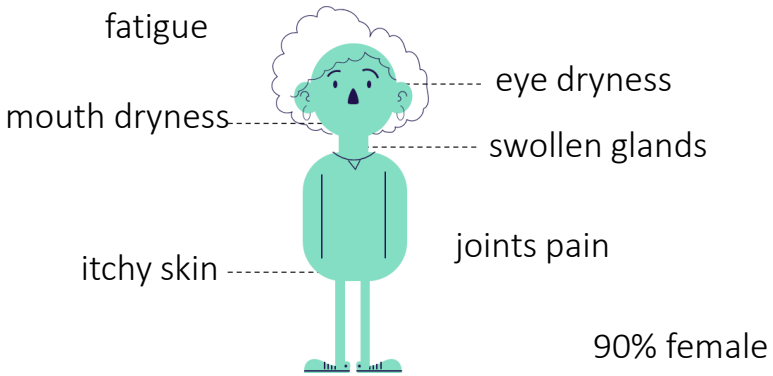
PhD Student :
Cheïma BOUDJENIBA

CIFRE PhD with :
Mickaël GUEDJ and Etienne BECHT (Servier)
Etienne BIRMELE (Université Strasbourg, ex Paris 5)
Benno SCHWIKOWSKI (Institut Pasteur)

Summary

Sjogren's syndrome (SjS)

Autoimmune disease characterized by lymphoid infiltration of the salivary and lacrimal glands



DISEASE DIAGNOSIS

Biopsy **minor salivary glands**
Ultrasonography of **major salivary glands**
Identification of specific **autoantibodies**

Schimer's test
ESSPRI (EULAR Sjögren's Syndrome Patients Reported Index)
ESSDAI (EULAR Sjögren's Syndrome Disease Activity Index)

- **No treatment**, only symptomatic treatments.
- **Multi-organ disease**, clinical manifestations and biological disturbances are considerably heterogeneous among individuals. Such heterogeneity provides a major challenge to designing clinical trials.

IMI project

- **Goals** : identify new sensitive clinical endpoints, define endotype-specific biomarkers and develop an original design for an innovative multi-arm clinical trial.

Currently, involved in WP4 : "Clinical and blood-based biomarkers for patient stratification or as surrogate endpoints"

Stratification of whole blood transcriptome of SjS patients

Soret et al., Nature Communication, 2021. “A new molecular classification to drive precision treatment strategies in primary Sjögren's syndrome”

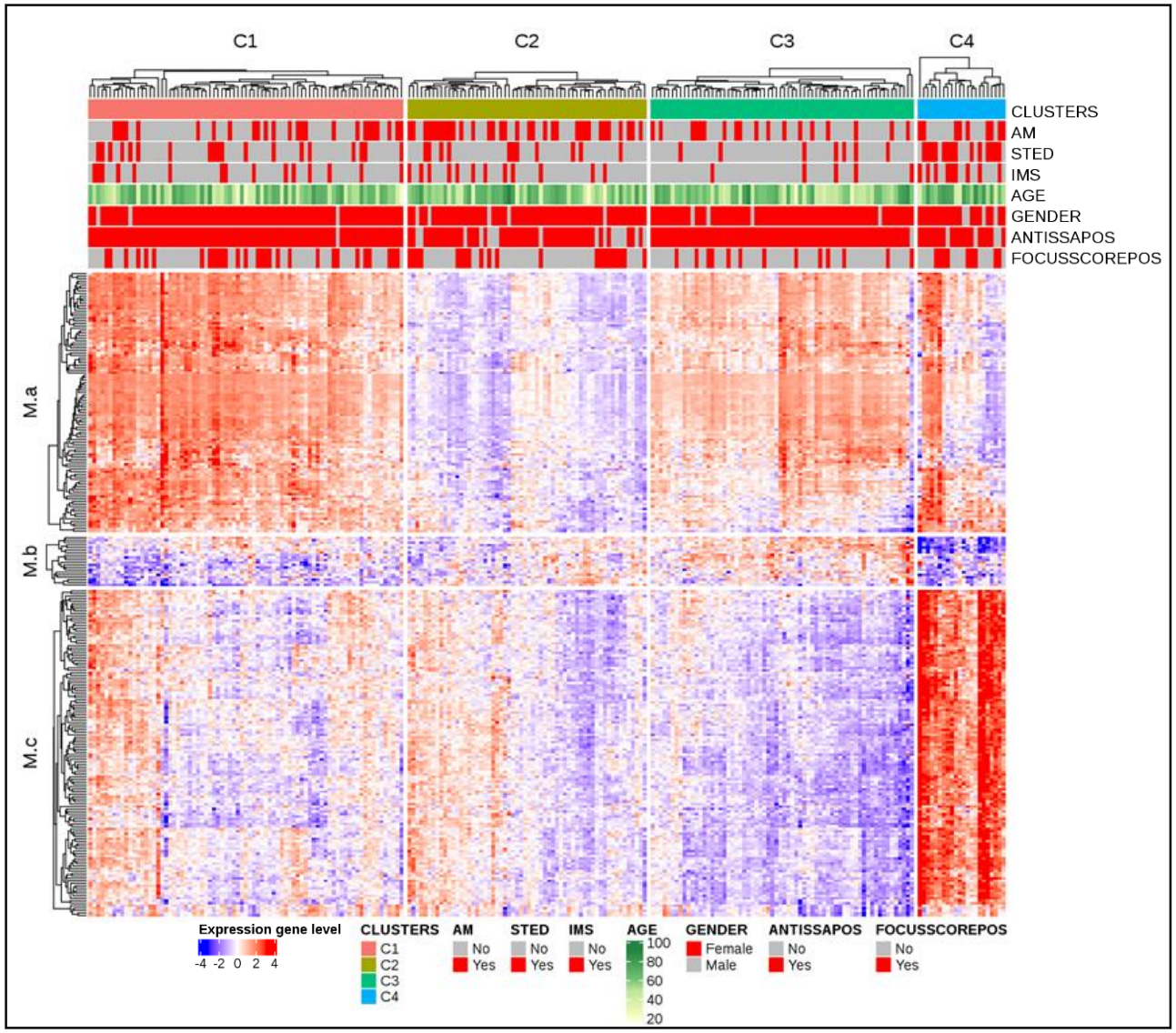


Figure from Soret et al

Identification of 4 molecular subtypes of SjS patients have been described :

C1 : IFN
C2 : Normal-like
C3 : B Cell + IFN
C4 : Inflammatory + IFN

} *Patient clustering*

M.a : Interferon
M.b : Lymphoid Lineage / TCells
M.c : Inflammation

} *Gene modules*

Gene modules identified fairly limited, there is a need of a more granular description of the heterogeinity

Data sources

Name	Number of patients	Types
ASSESS	341	Transcriptomic, Clinical
PreciseSADS	371	Transcriptomic Clinical Biological Antibodies Flow cytometry
UKPSSR	144	Transcriptomic, Clinical
GSE84844	30	Transcriptomic

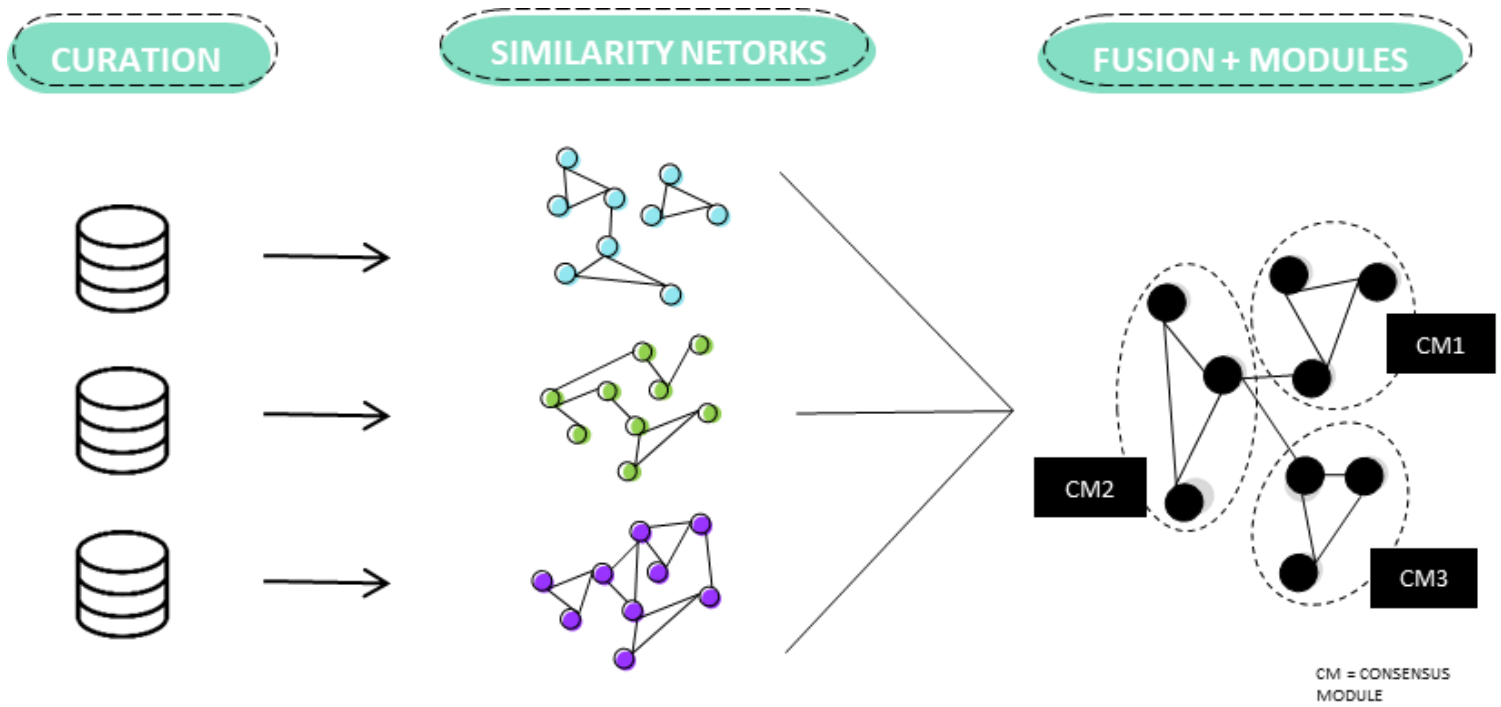
Methods

Pseudo Code
Inputs:
X: transcriptomic matrix

Step 1 : Gene selection
For each X, select most varying genes : g_{list_X}
Keep the intersection of genes from g_{list_X}

Step 2 : Similarity and fusion *(Wang et al., 2014)*
For each X, $C_X = \ cor(X, X)$, genes correlation matrix
Then, $A_X = exp(0{,}5 * (1 - C_X))$, gene affinity matrix
Finally apply SNF(A)

Step 3 : Spectral Clustering *(Luxburg et al., 2009)*
Spectral transformation
K-means clustering on first principal components



Identification of 11 co-expressed and consensus gene modules (CM)

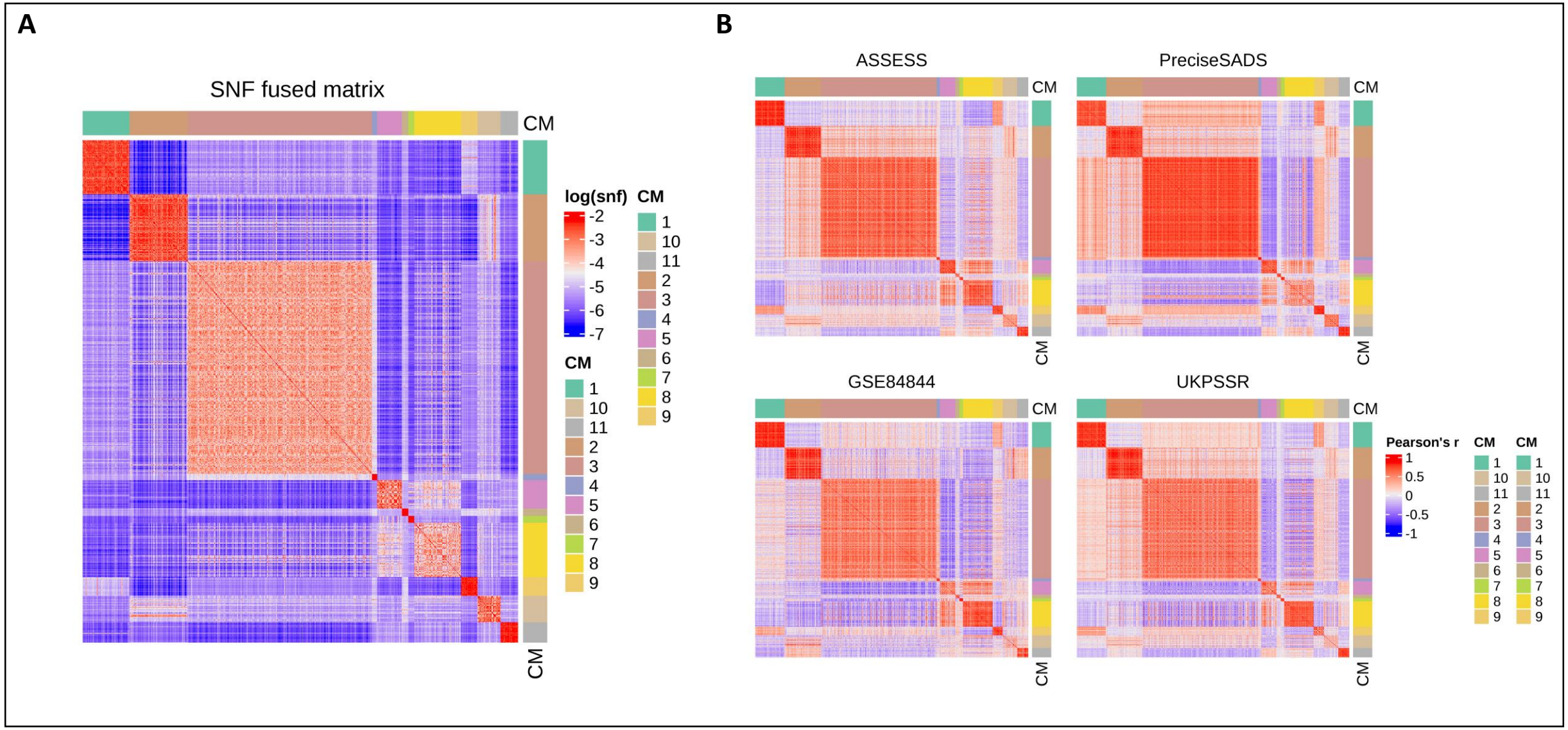


Figure 1 – A) Heatmap of the fused graph B) Heatmap of affinity matrices of the four inputs dataset, with genes grouped by their consensus gene module

Enrichment analysis for CM annotation

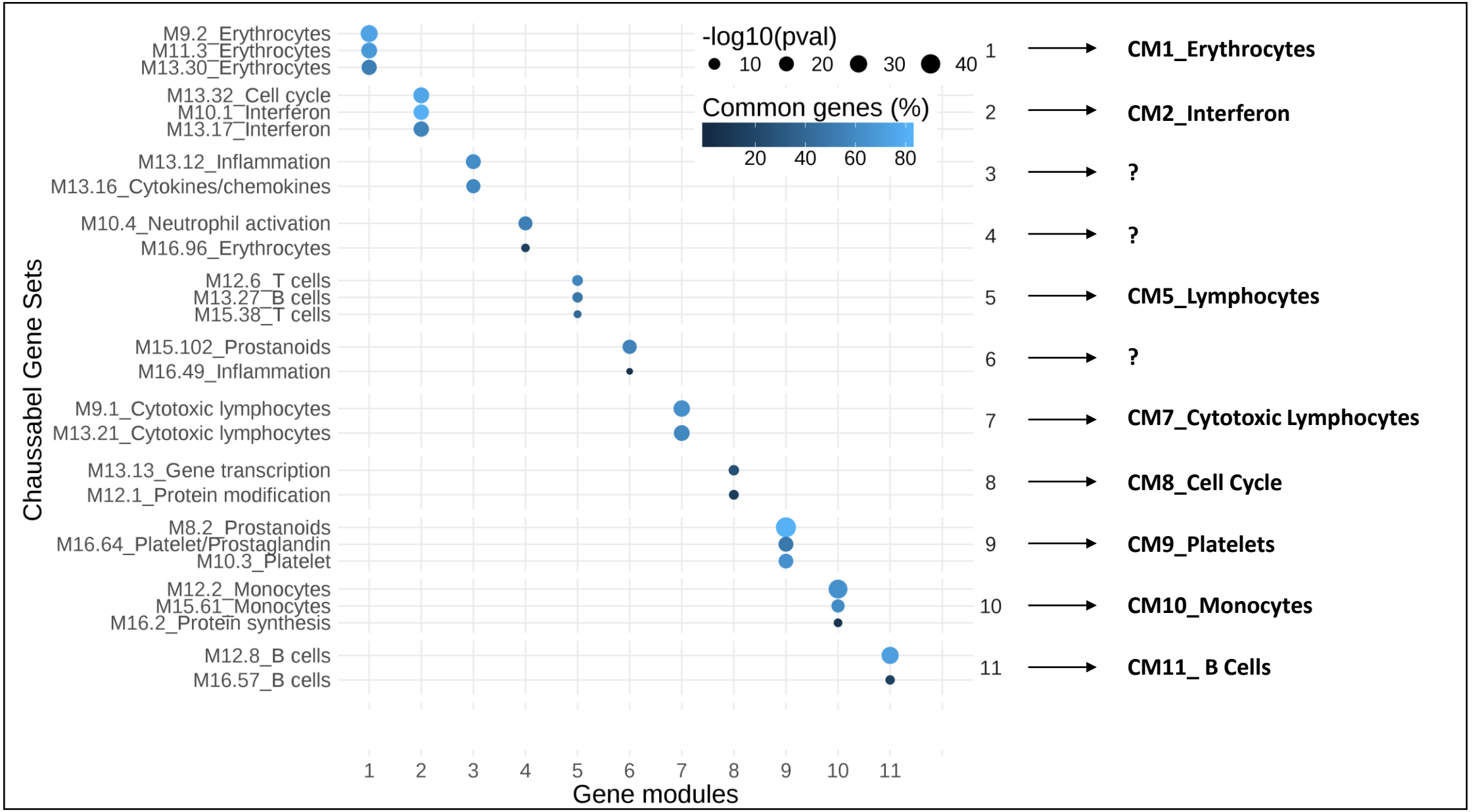


Figure 2 – Enrichment analysis of gene modules, pval.th = 10e-4, perc.th = 20

Altman, Chaussabel, Banchereau et al, 2021, "Development of a fixed module repertoire for the analysis and interpretation of blood transcriptome data"

Transcriptome of purified cell types for CM annotation

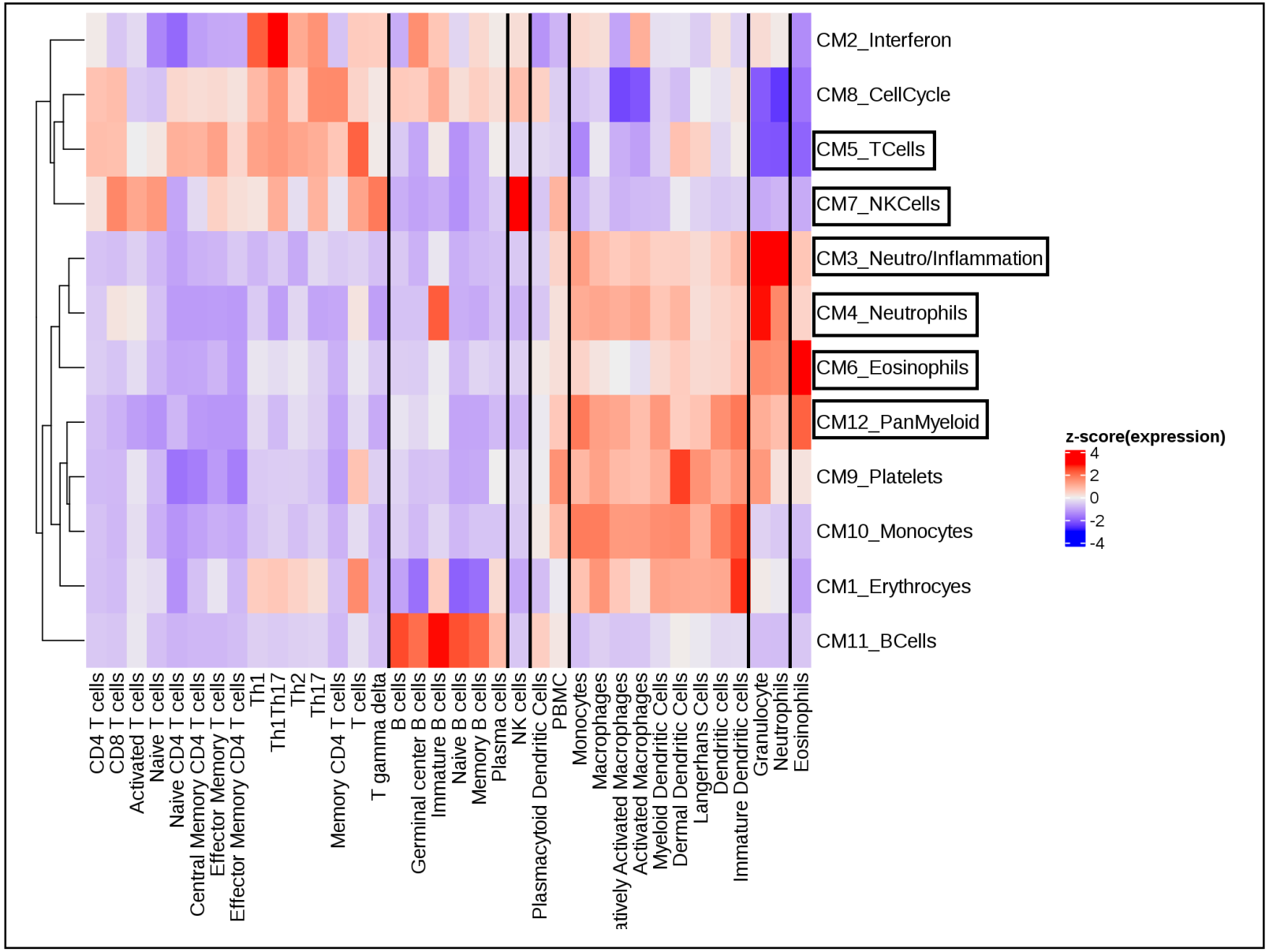


Figure 3 - Average of cell types and modules mapped with flow cytometry sorted purified cell types

Becht et al., 2016, "Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression"

Validation of cellular modules with cell types measured by flow cytometry

For the same patients we look at the **correlation** between :
Gene expression and cell types measured by Flow

Average expression of modules

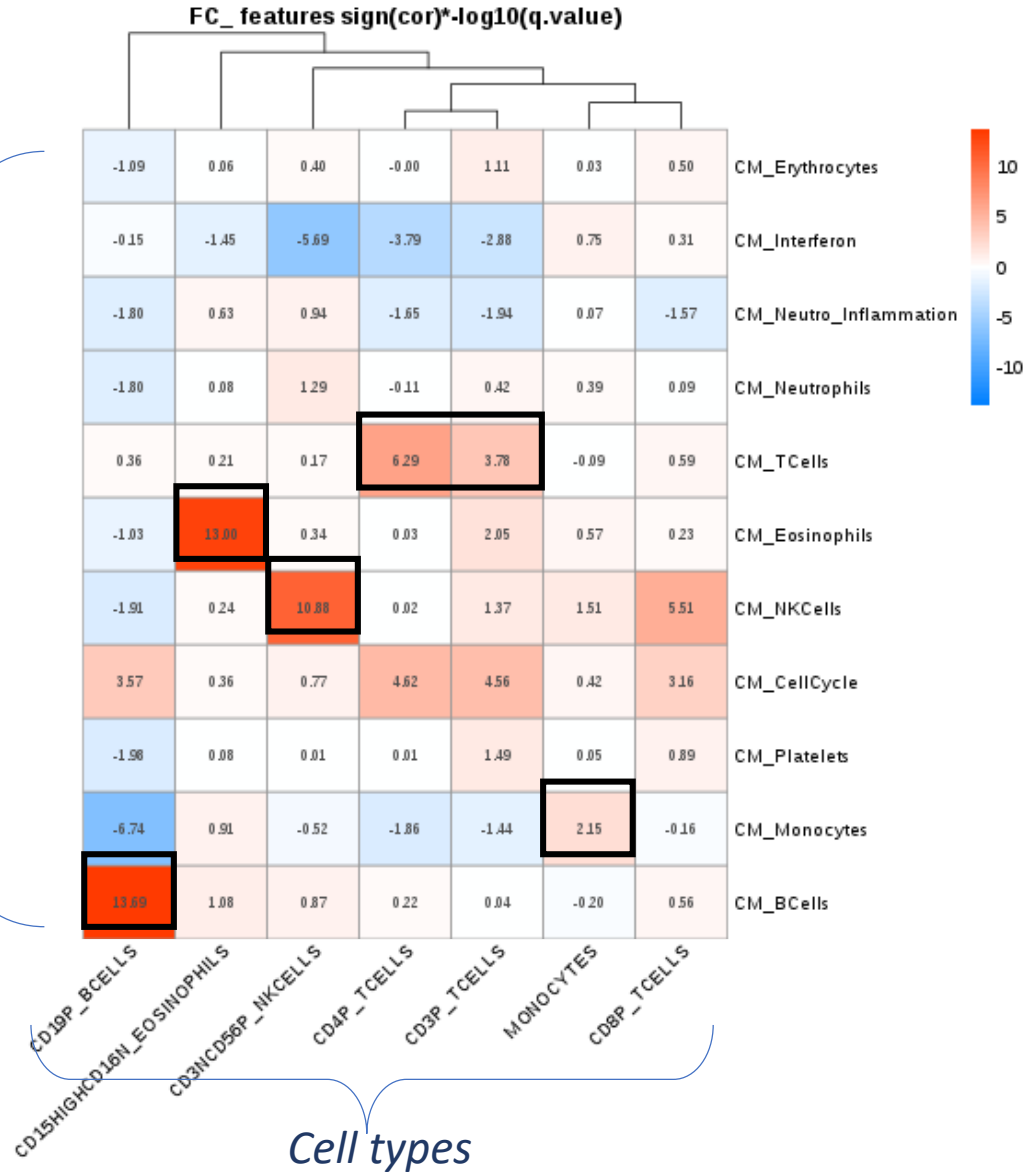


Figure 4 - Heatmap of q.value. Correlation with flow cytometry features

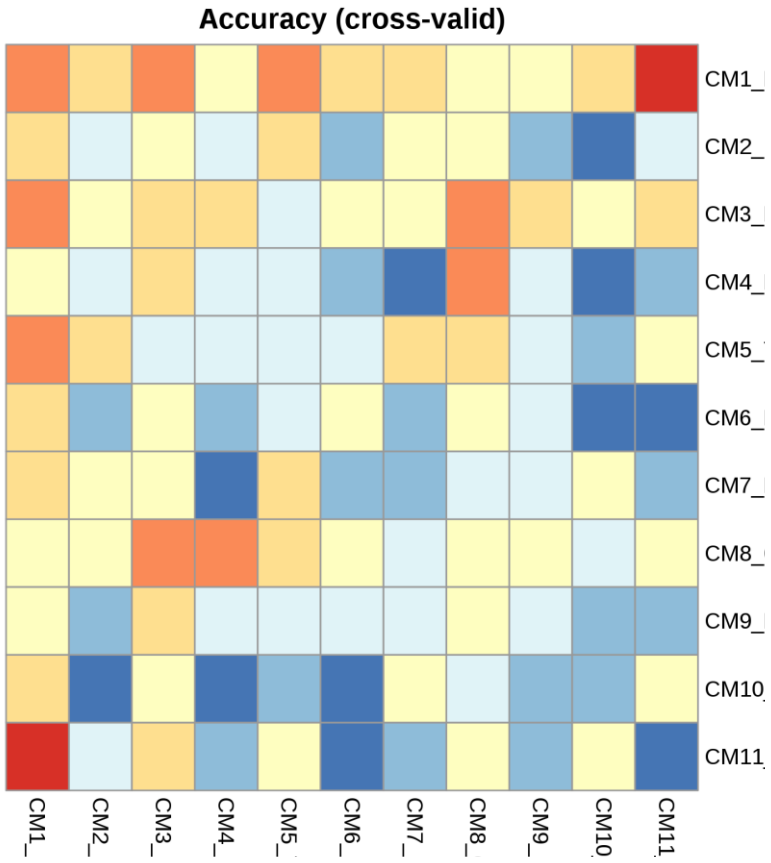
Summary

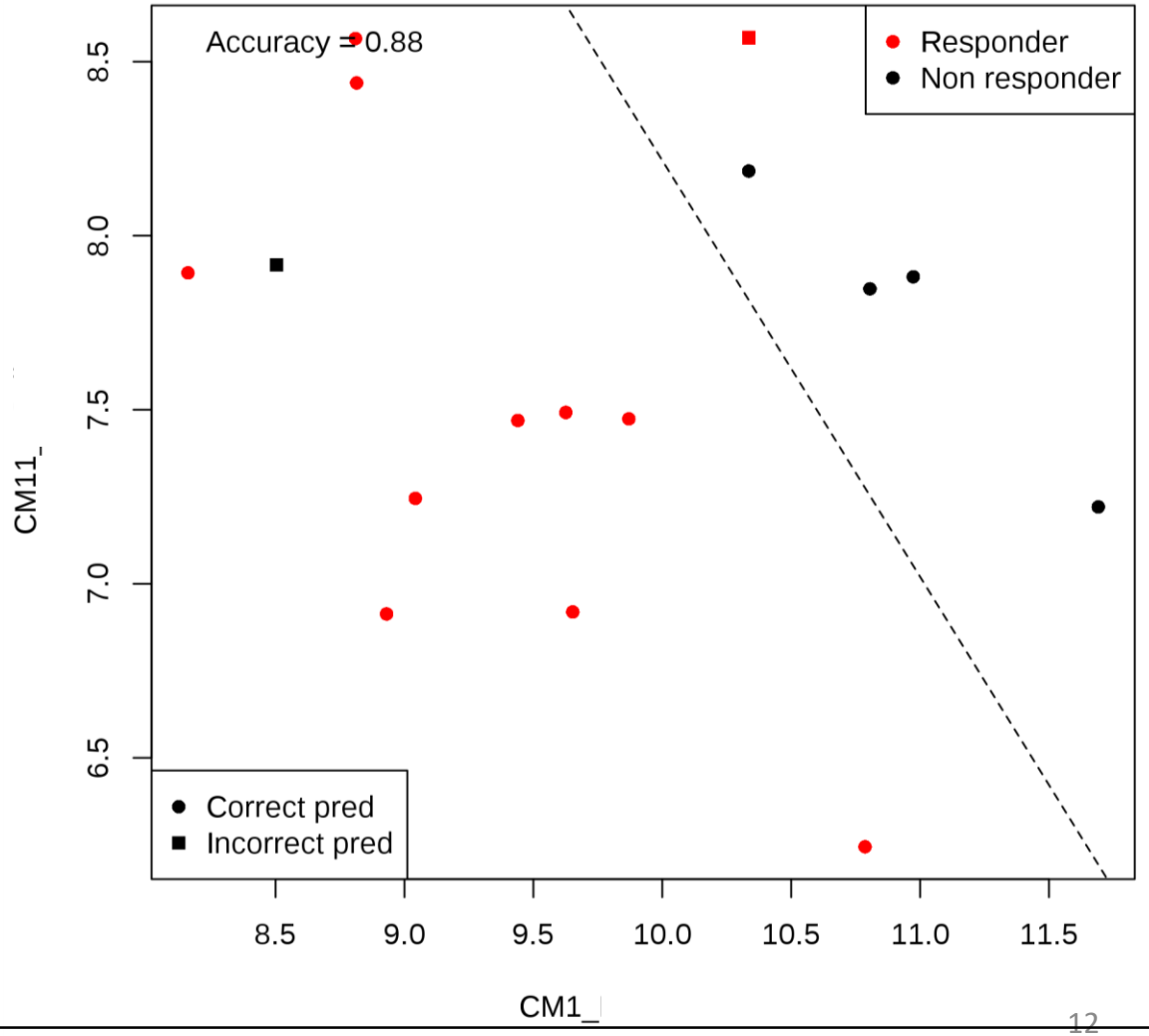
	CM1 Erythrocytes (n=253)	CM2 Interferon (n=359)	CM3 Neutro/ Infla (n=809)	CM4 Neutrophils (n=37)	CM5 T Cells (n=1449)	CM6 Eosinophils (n=65)	CM7 NK (n=192)	CM8 Cell Cycle (n=658)	CM9 Platelets (n=155)	CM10 Monocytes (n=329)	CM11 B Cells (n=156)
Enrichment	Erythrocytes (M9.2, M11.3, M13.30)	Interferon (M10.1, M13.17)	Inflammation (M13.12) Neutrophil degranulation	Neutrophils (M10.4) Neutrophil degranulation	T Cells (M12.6)	Prostanoids (M15.102)	Cytotoxic Lymphocytes (M9.1, M13.21)	Gene transcription (M13.13)	Prostanoids, Platelets (M8.2, M16.64,M10.3)	Monocytes (M12.2, M15.61)	B Cells (M12.8, M16.57)
Cellular		Th17, Macrophage, Dendritic.	Granulo, Neutro, Myeloid lineage.	Neutro, Myeloid lineage.	T Cells	Eosinophils	NK Cells	Lymphoid lineage	Myeloid lineage	Myeloid lineage	B Cells

Table 1 – Results of characterisation of the 11 identified modules. Enrichment is obtain by selecting top enriched pathway (Chaussabel, GO, MsigDB). Cellular row is from analysis of purified cells association (GSE86362) and flow cytometry dataset.

Application in clinical trial : identification of biomarker for response to treatment

1) Average expression of each module.
2) Pairwise logistic regression (LeaveOneOutCrossValidation) to predict response to treatment.
3) Select best model based on accuracy





Conclusion

- Identification of 11 co-expressed gene modules reproducible across multiple independant cohorts

- Functional and cellular interpretation of the modules

- Validation by flow cytometry

- Ongoing application in clinical trials :
  - Prediction of response at baseline
  - Correlation with other biomarkers for translability

Acknowledgements

Etienne Becht
Mickael Guedj
Antoine Hamon
Laurence Laigle
Yann Bachelot
Bastien Chassagnol
Antoine Bichat
Emiko Desvaux
Perrine Soret
Sandra Hubert
Céline Lefebvre
Audrey Aussy
Phillipe Moingeon
Jack Swindle
Anne-Sophie Pulcrano

Etienne Birmele

Diana Trutschel
Benno Schwikowski
Federico Melograna
Pierre Bost

Xavier Mariette
Fan Wai Ng
Jessica Tarn
Joel Von Roon
JE Gottenberg
JO Pers
Marta Alarcon-Riquelme
Michael Barnes

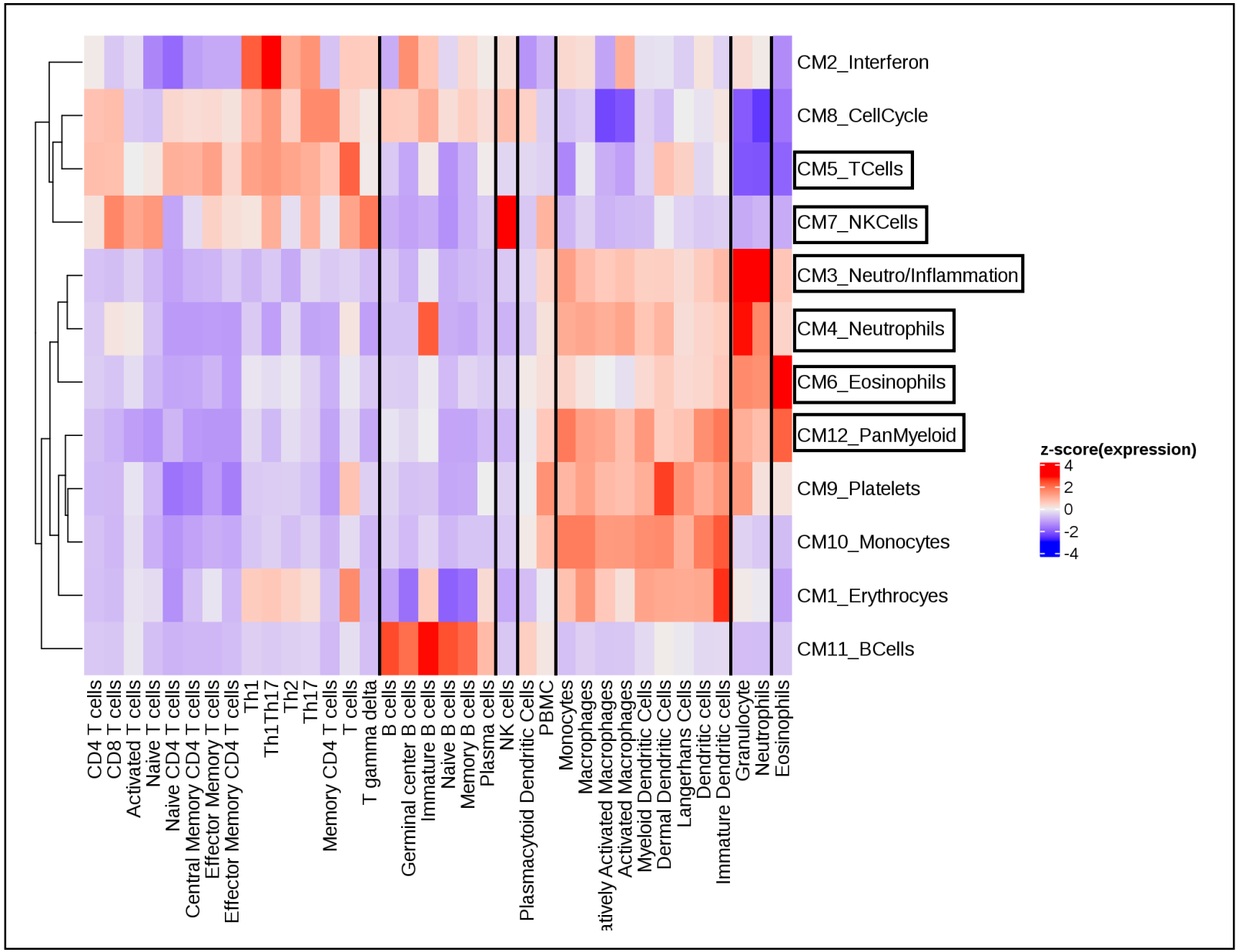


Figure 3 - Average of cell types and modules mapped with flow cytometry sorted purified cell types