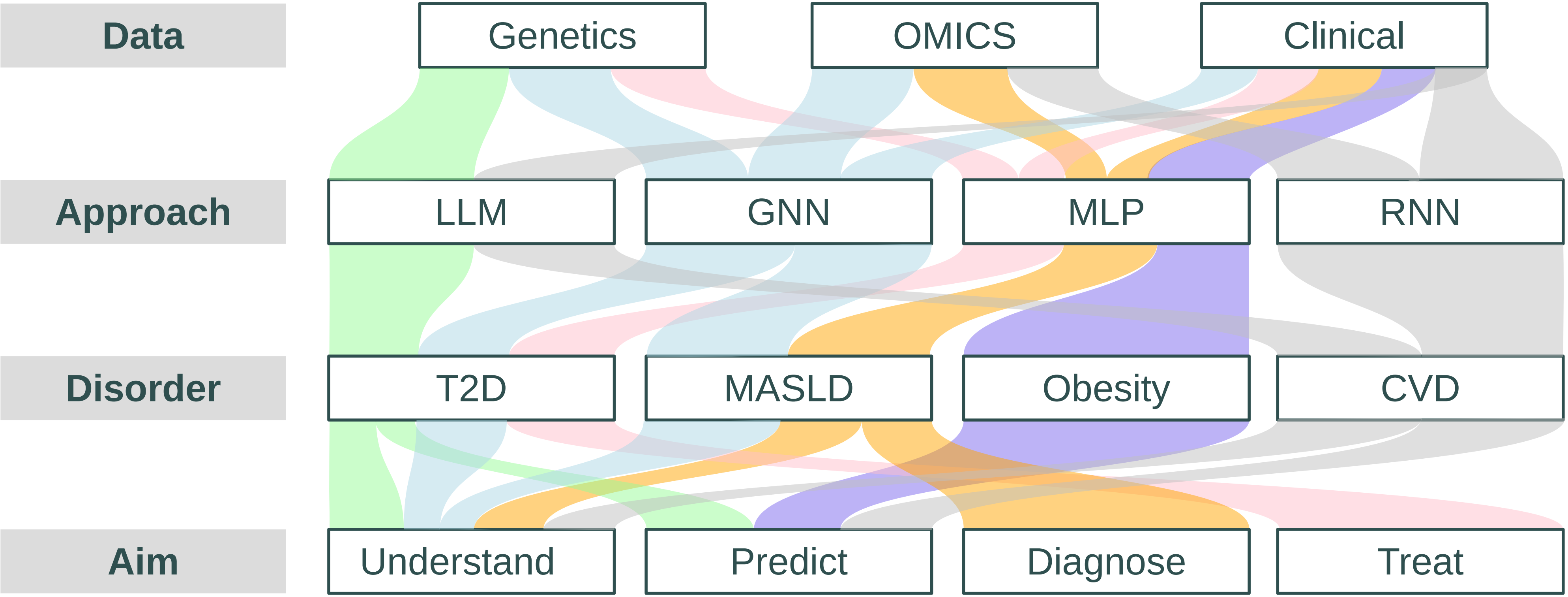
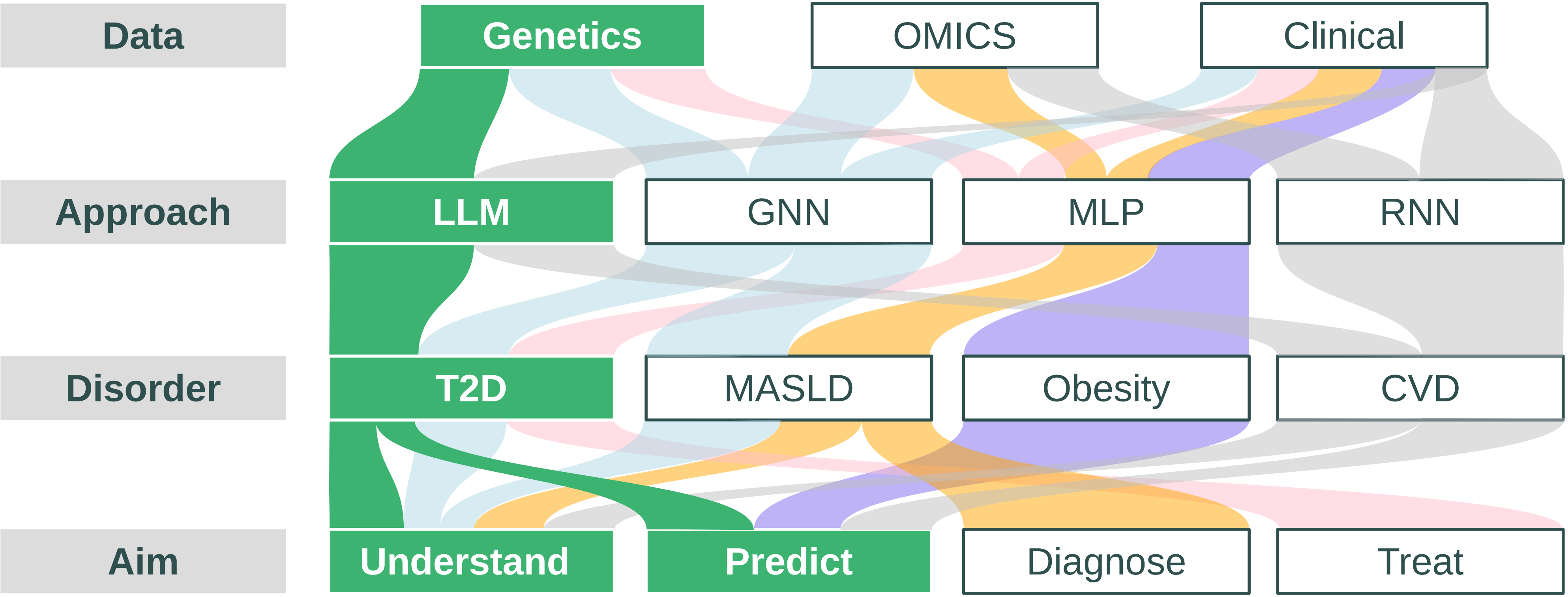
IA @ Egenodia


Mash
Gru
Fullyconnected
Treatment
Lmm
Largelanguagemodel
Rnn
Gemini
Tensor
Steatosis
Encoderdecoder
Vae
Neuralnetwork
Treatment
Lstm
Attention
Gpu
Mash
Diagnosis
Vae
Artificialintelligence
Inflammation
Chatgpt
Diabetes
Prognosis
Diagnosis
Predict
Obesity
Diabetes
Deeplearning
Ssm
Cnn
Dnn
Llama
Chatgpt
Artificialintelligence
Dnn
Cnn
Mash
Predict
Gru
Autoencoder
Inflammation
Transformer
Neuralnetwork
Diabetes
Largelanguagemodel
Convolutional
Chatgpt
Obesity
Diabetes
Autoencoder
Dense
Rnn
Gemini
Artificialintelligence
Claude
Gru
Lstm
Masld
Gpu
Attention
Diagnosis
Obesity
Prognosis
Deeplearning
Artificialintelligence
Llama
Cluster
Neuralnetwork
Tensor
Lmm
Steatosis
Llama
Gpu
Prognosis
Fullyconnected
Largelanguagemodel
Transformer
Encoderdecoder
Tensor
Masld
Cnn
Deeplearning
Gemini
Deeplearning
Ssm
Claude
Ssm
Cluster
Cluster
Vae
Masld
Neuralnetwork
Claude
Steatosis
Encoderdecoder
Obesity
Lstm
Dnn
Inflammation
Dense
Transformer
Attention
Treatment
Autoencoder
Rnn
Fullyconnected
Convolutional
Inflammation


Data
Genetics
OMICS
Clinical
Approach
LLM
GNN
MLP
RNN
Disorder
T2D
MASLD
Obesity
CVD
Aim
Understand
Predict
Diagnose
Treat

Smaïn Fettem



Data: **Genetics** OMICS Clinical

Approach: **LLM** GNN MLP RNN

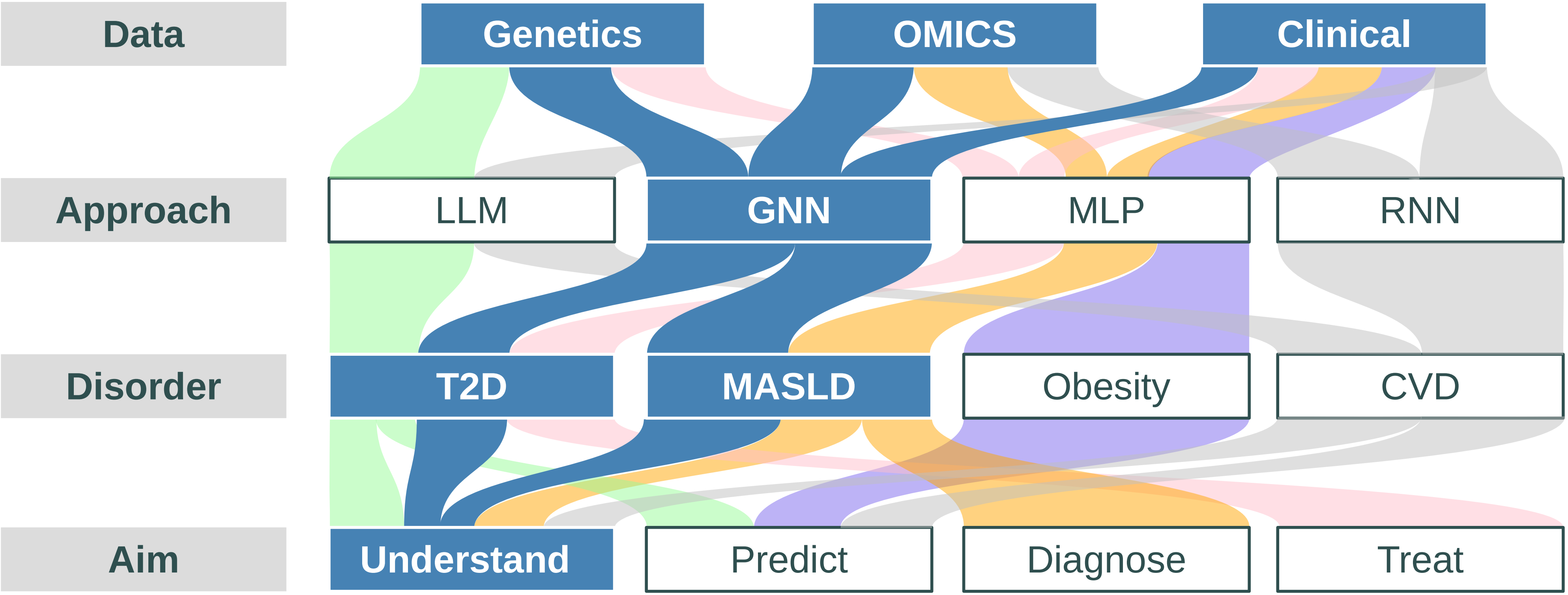
Disorder: **T2D** MASLD Obesity CVD

Aim: **Understand** **Predict** Diagnose Treat

PreciDIAB
National Center for
Precision Diabetic Medicine

Amna Khamis

Omaima Binan



Data: **Genetics** **OMICS** **Clinical**

Approach: LLM **GNN** MLP RNN

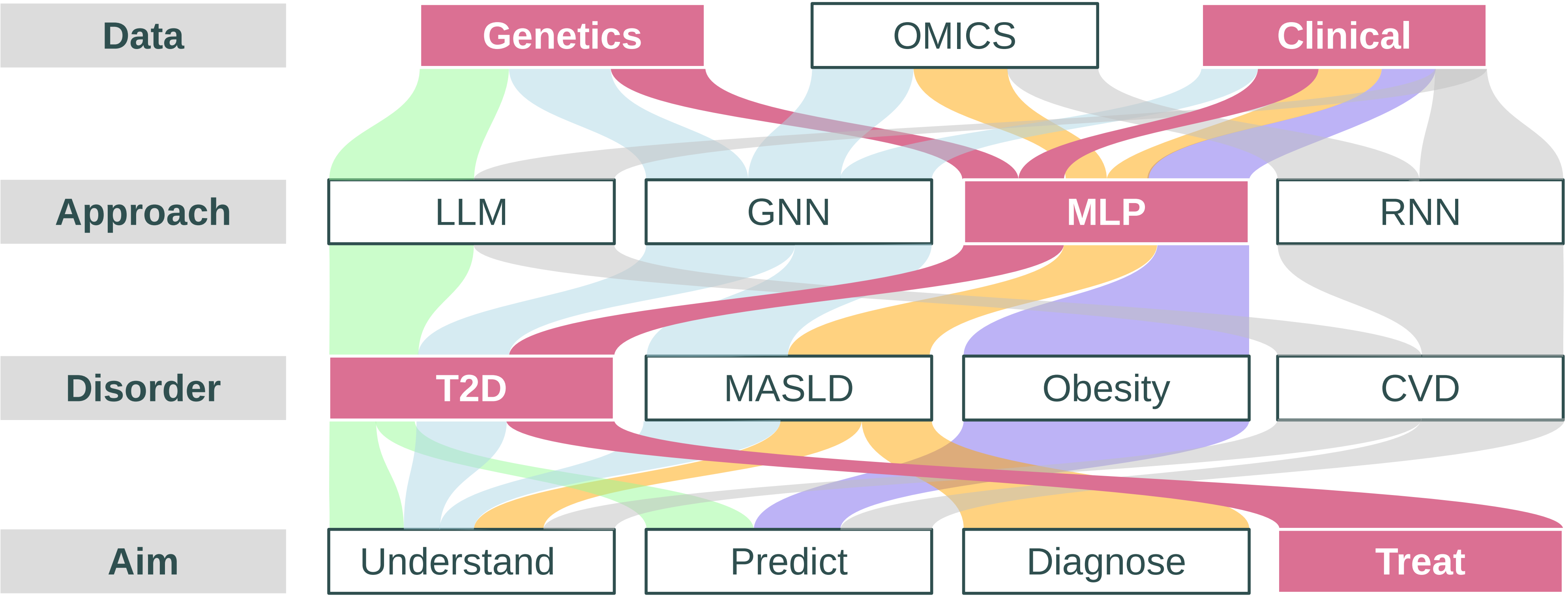
Disorder: **T2D** **MASLD** Obesity CVD

Aim: **Understand** Predict Diagnose Treat

Université
de Lille

Marwa Afnouch

Arthur Six



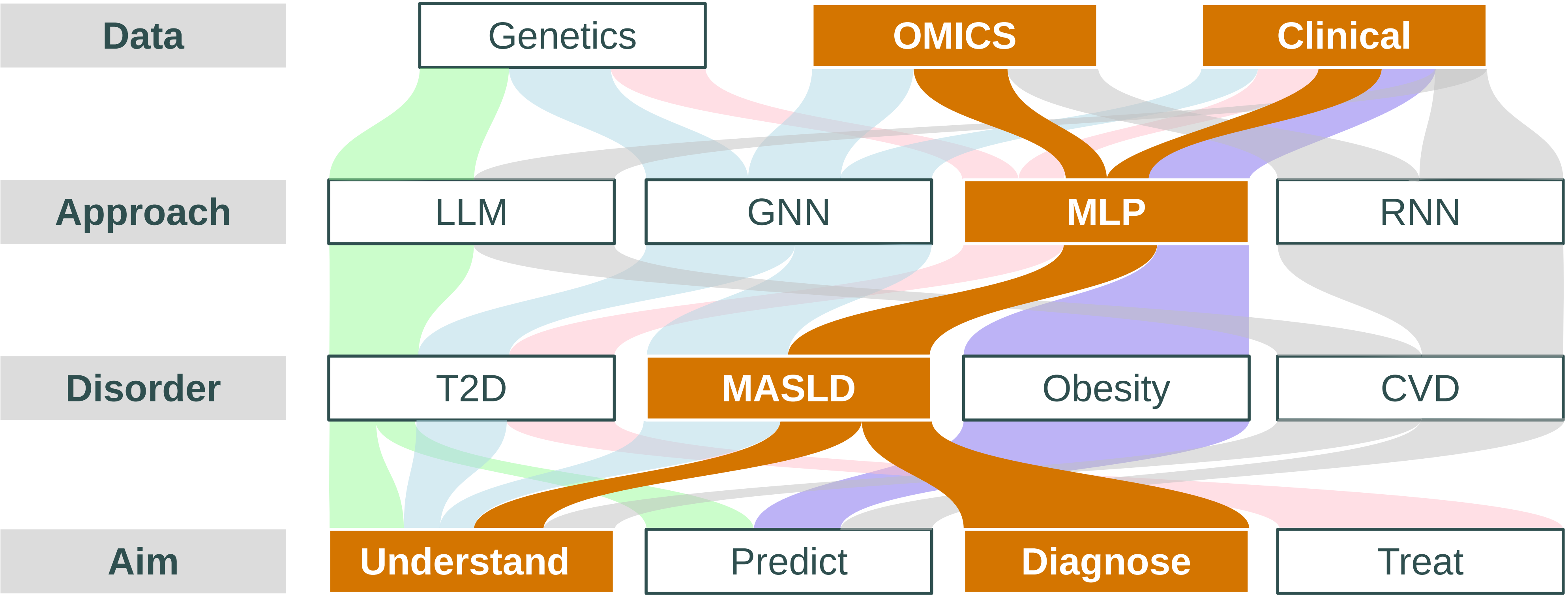
Data
Genetics
OMICS
Clinical

Approach
LLM
GNN
MLP
RNN

Disorder
T2D
MASLD
Obesity
CVD

Aim
Understand
Predict
Diagnose
Treat

PreciDIAG
Precision Diagnosis
for Type 2 Diabetes



Data
Genetics
OMICS
Clinical

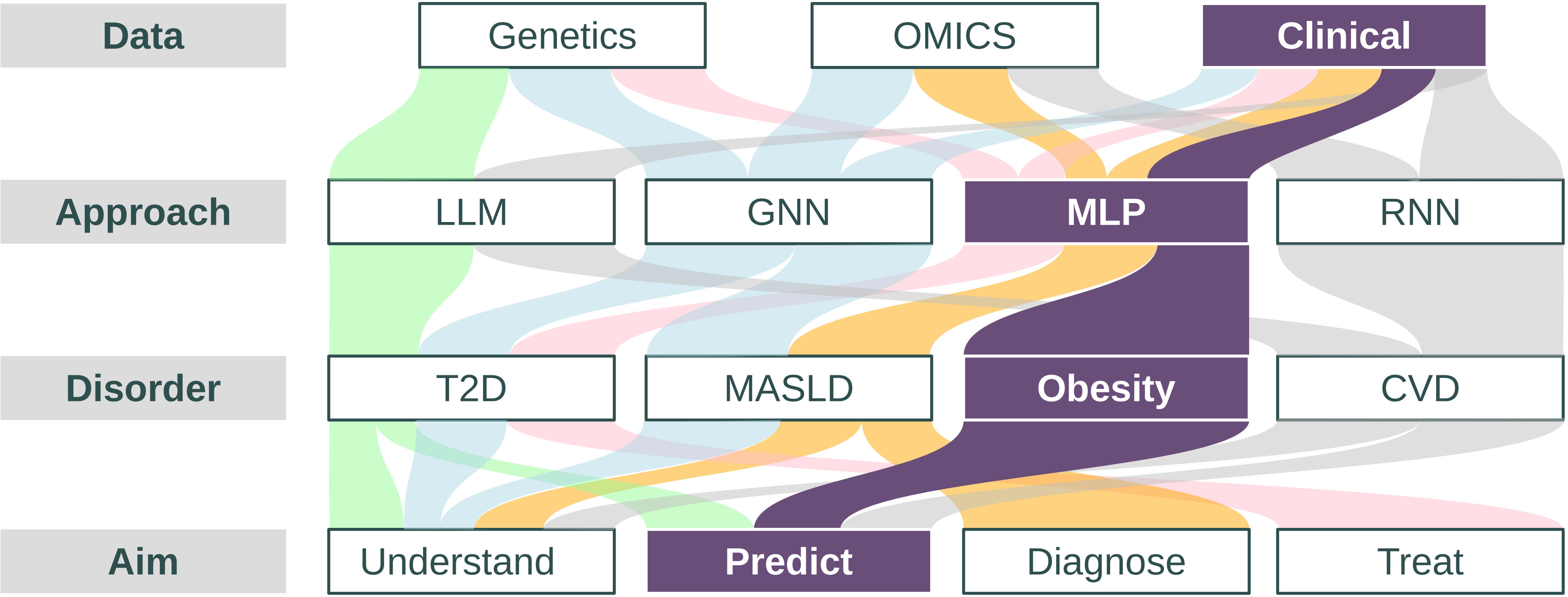
Approach
LLM
GNN
MLP
RNN

Disorder
T2D
MASLD
Obesity
CVD

Aim
Understand
Predict
Diagnose
Treat

PreciNASH
recherche hospitalo-universitaire

Sami Le Guilcher



Data

Genetics

OMICS

Clinical

Approach

LLM

GNN

MLP

RNN

Disorder

T2D

MASLD

Obesity

CVD

Aim

Understand

Predict

Diagnose

Treat

horizon europe

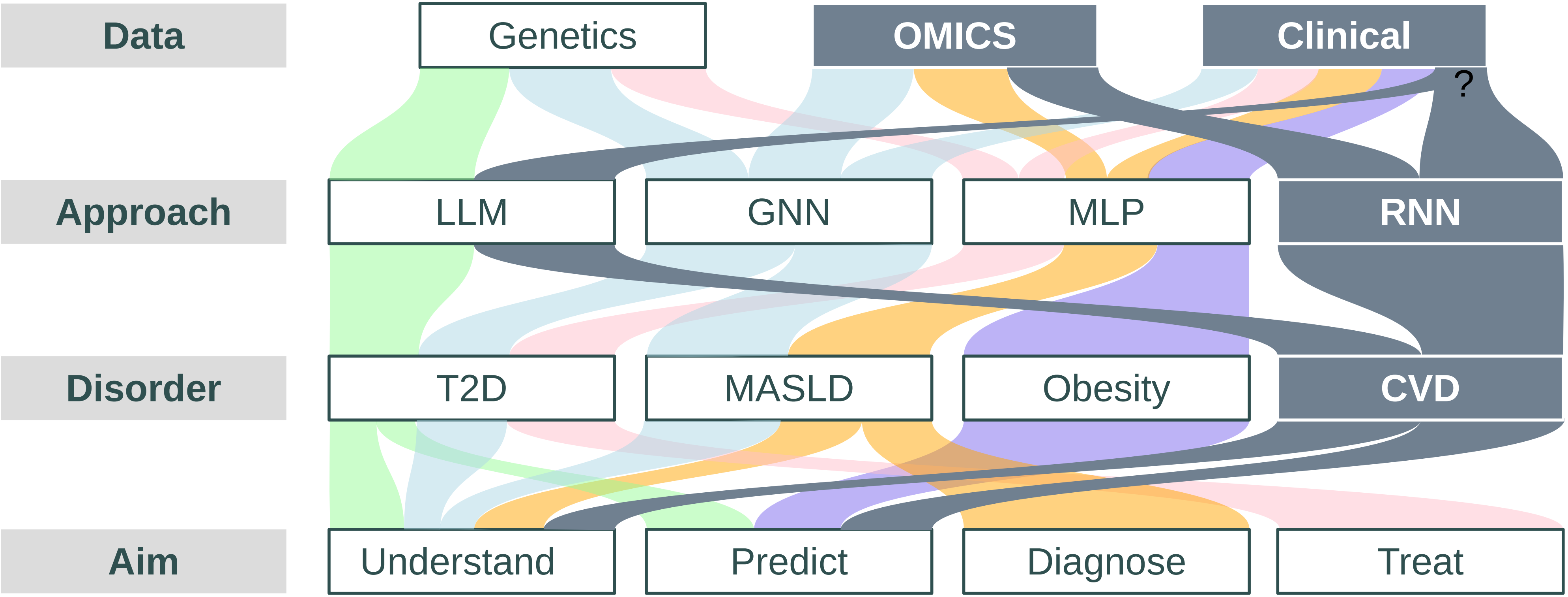
Obelisk

Marc-Emmanuel
Dumas

Salla
Åkerblom

Zahra
Elhamraoui

Francesc
Puig-Castellví



Data: Genetics, **OMICS**, **Clinical**

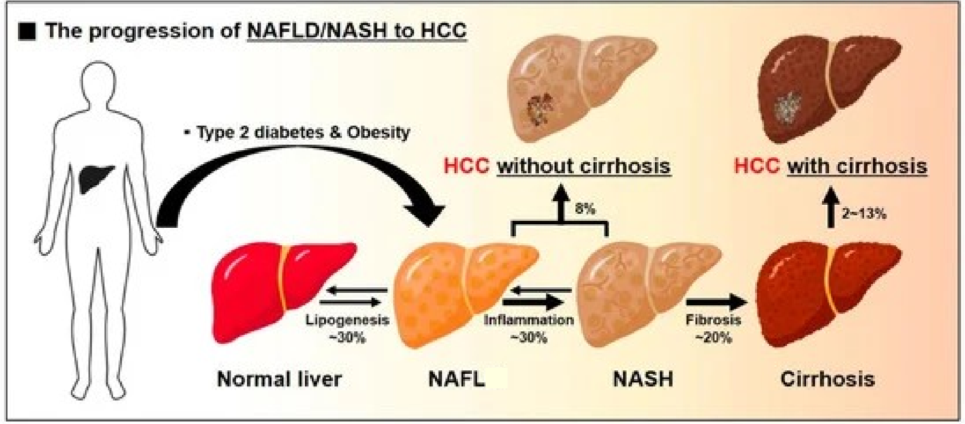
Approach: LLM, GNN, MLP, **RNN**

Disorder: T2D, MASLD, Obesity, **CVD**

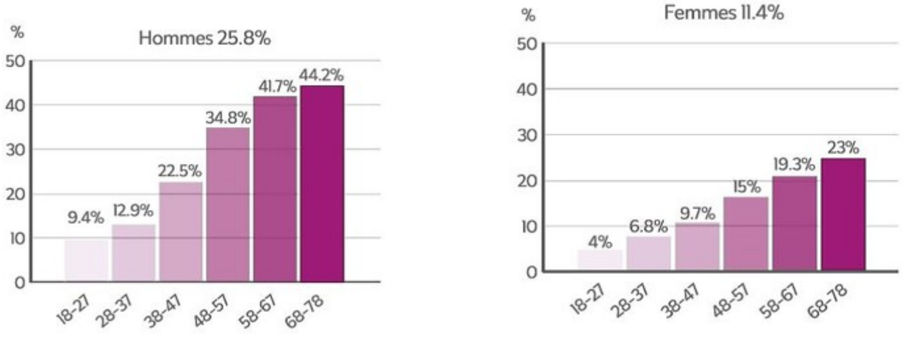
Aim: Understand, Predict, Diagnose, Treat

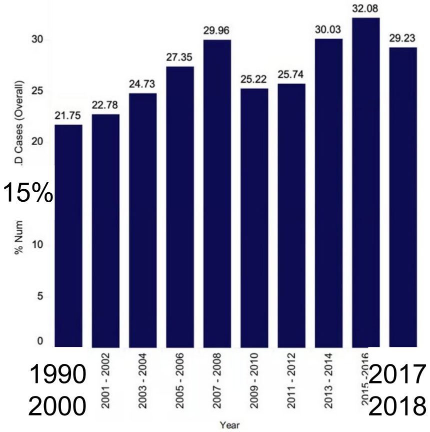
Metabolic dysfunction–associated steatotic liver disease (MASLD)

Kim *et al* (2021) *Int. J. Mol. Sci.*, 22(9): 4495

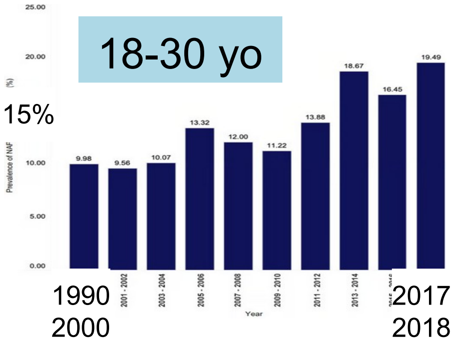


Prévalence en fonction de l'âge et du sexe

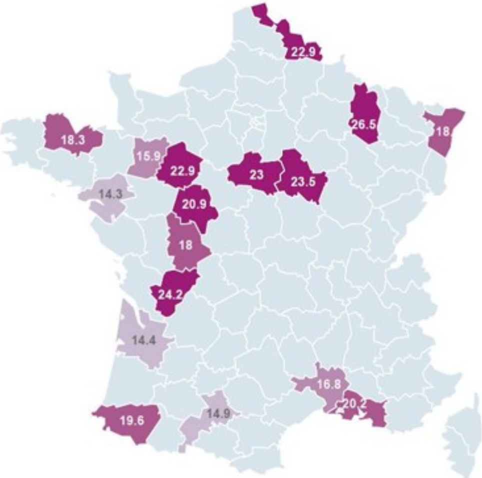




18-30 yo



Kim *et al* (2022) *Met. Target Organ Damage*, 2: 19



NASH
(now MASH)

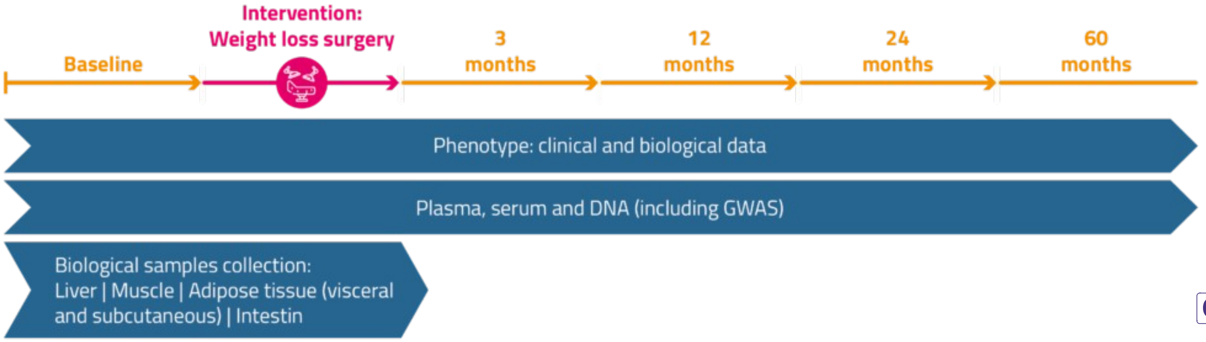
Répartition
par régions

Paris MASH Meeting
(11-12 juillet 2019)

NALFD (now MASLD)

PreciNASH
recherche hospitalo-universitaire

ABOS and PreciNASH



Phenotype: clinical and biological data

Plasma, serum and DNA (including GWAS)

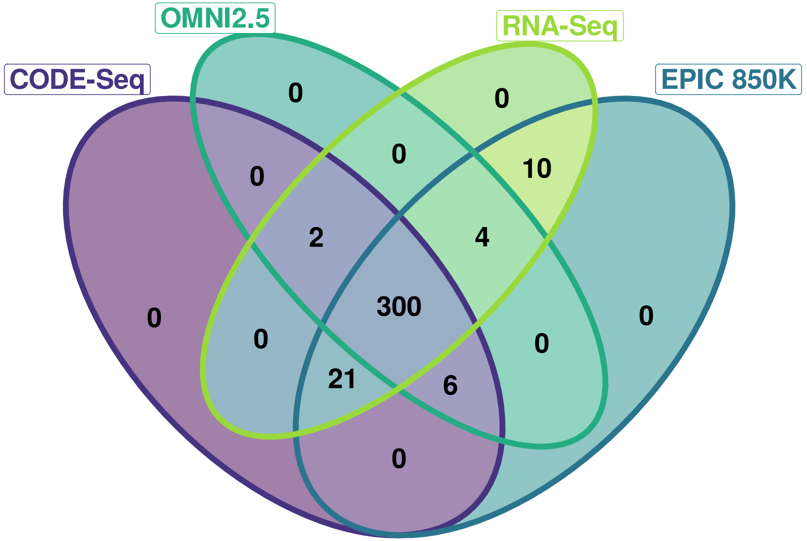
Biological samples collection:
Liver | Muscle | Adipose tissue (visceral and subcutaneous) | Intestin

ABOS (Biological Atlas of Severe Obesity)

All subjects had bariatric surgery

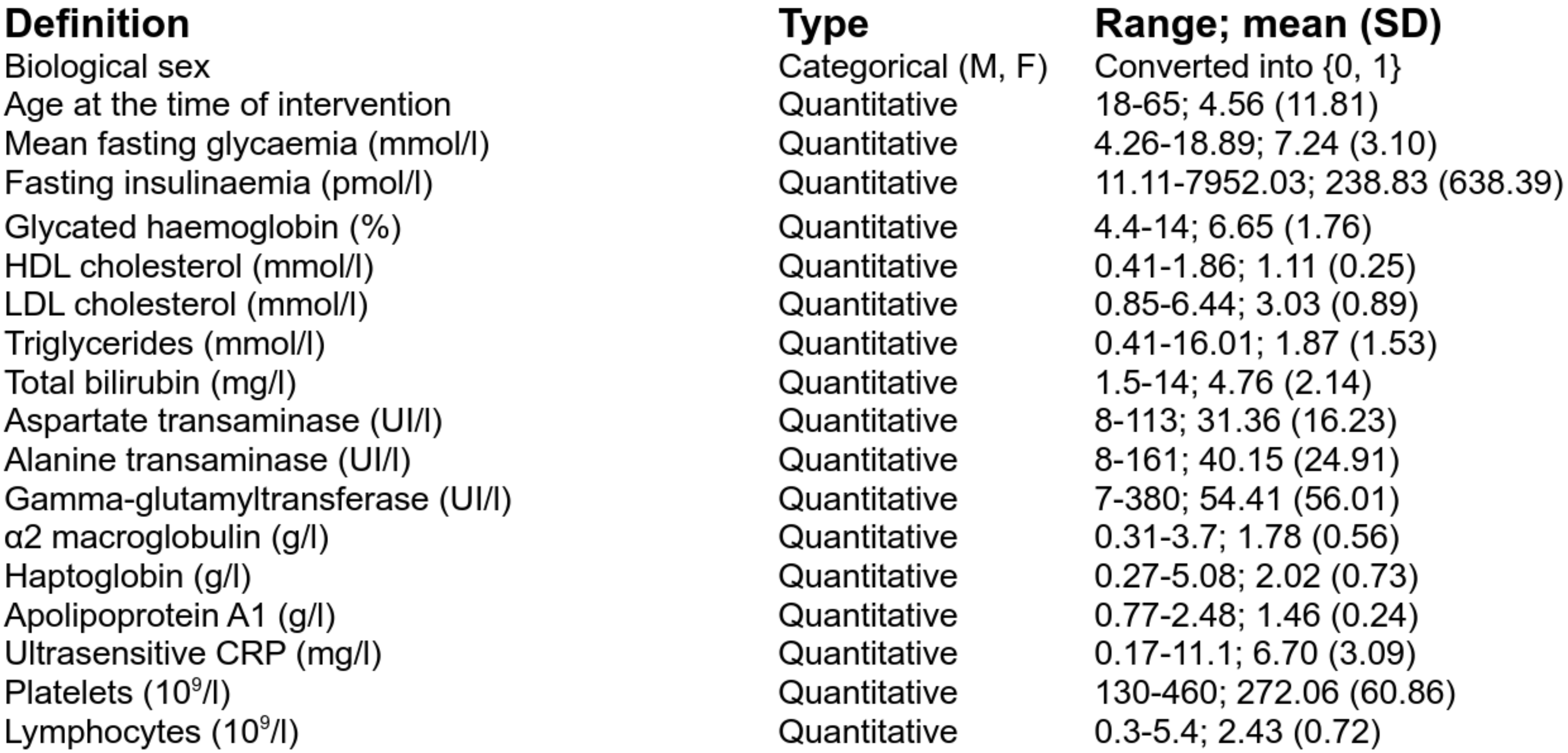
PreciNASH project

ABOS subset: Only European ancestry and unrelated individuals



(+66 clinical and personal data
+1076 identified metabolites
in blood and liver)

Biological and clinical variables



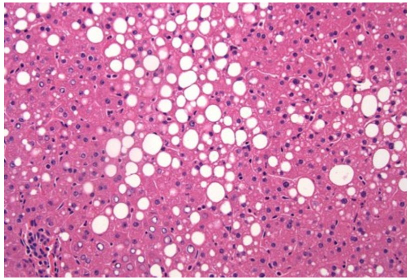
Definition	Type	Range; mean (SD)
Biological sex	Categorical (M, F)	Converted into {0, 1}
Age at the time of intervention	Quantitative	18-65; 4.56 (11.81)
Mean fasting glycaemia (mmol/l)	Quantitative	4.26-18.89; 7.24 (3.10)
Fasting insulinaemia (pmol/l)	Quantitative	11.11-7952.03; 238.83 (638.39)
Glycated haemoglobin (%)	Quantitative	4.4-14; 6.65 (1.76)
HDL cholesterol (mmol/l)	Quantitative	0.41-1.86; 1.11 (0.25)
LDL cholesterol (mmol/l)	Quantitative	0.85-6.44; 3.03 (0.89)
Triglycerides (mmol/l)	Quantitative	0.41-16.01; 1.87 (1.53)
Total bilirubin (mg/l)	Quantitative	1.5-14; 4.76 (2.14)
Aspartate transaminase (UI/l)	Quantitative	8-113; 31.36 (16.23)
Alanine transaminase (UI/l)	Quantitative	8-161; 40.15 (24.91)
Gamma-glutamyltransferase (UI/l)	Quantitative	7-380; 54.41 (56.01)
α2 macroglobulin (g/l)	Quantitative	0.31-3.7; 1.78 (0.56)
Haptoglobin (g/l)	Quantitative	0.27-5.08; 2.02 (0.73)
Apolipoprotein A1 (g/l)	Quantitative	0.77-2.48; 1.46 (0.24)
Ultrasensitive CRP (mg/l)	Quantitative	0.17-11.1; 6.70 (3.09)
Platelets (10^9/l)	Quantitative	130-460; 272.06 (60.86)
Lymphocytes (10^9/l)	Quantitative	0.3-5.4; 2.43 (0.72)

Among 66 available, ignoring redundant, overlapping, and irrelevant ones

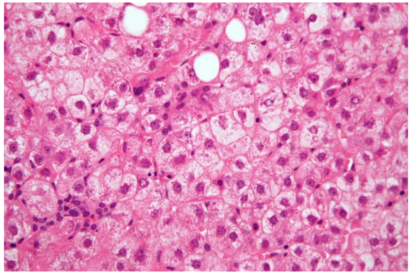
Subject grouping

Scoring on liver biopsy with the method from Kleiner and Brunt 2005

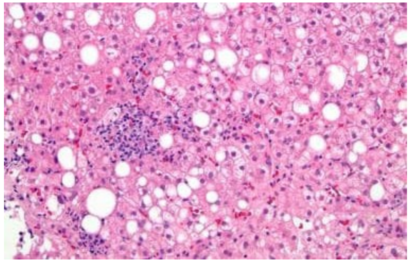
Steatosis
Categorical [0-3] from
quantitative measurement



Ballooning
Categorical [0-2]
= {none, some, much}



Inflammation
Categorical [0-3] from
number of foci



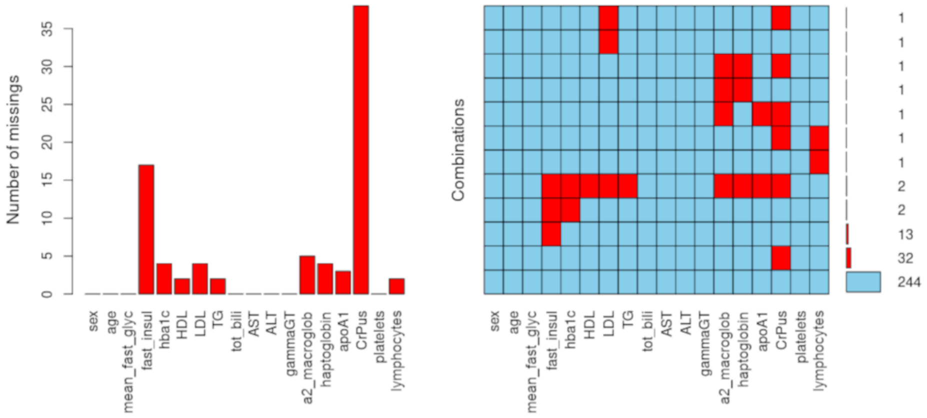
Final score:

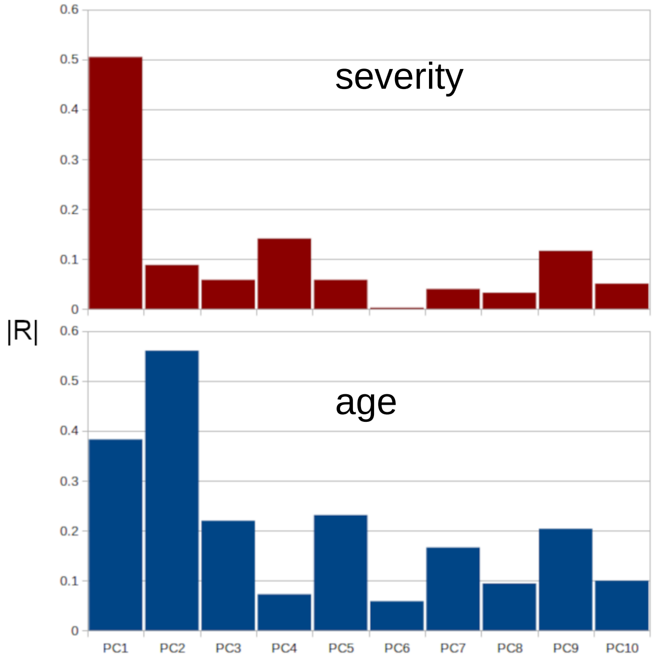
Healthy: $S = 0, B = 0, I = 0$ $n = 80$

NAFL: $S > 1, B = 0, I \geq 1$ $n = 137$
$S > 1, B > 1, I = 0$

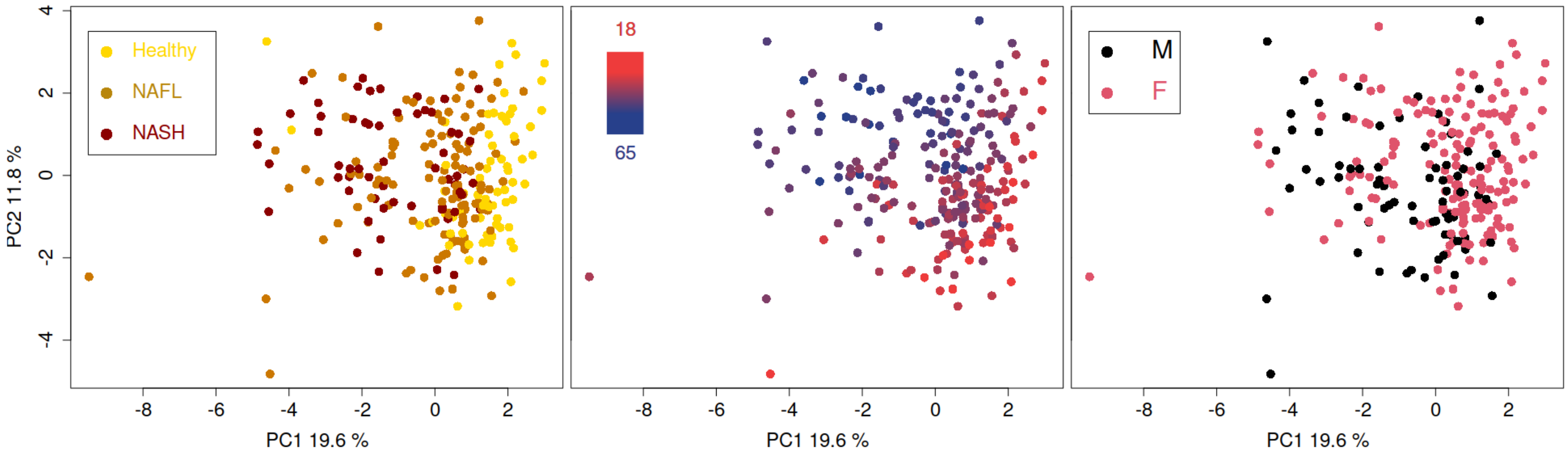
NASH: $S > 0, B > 0, I > 0$ $n = 83$

Clinical data

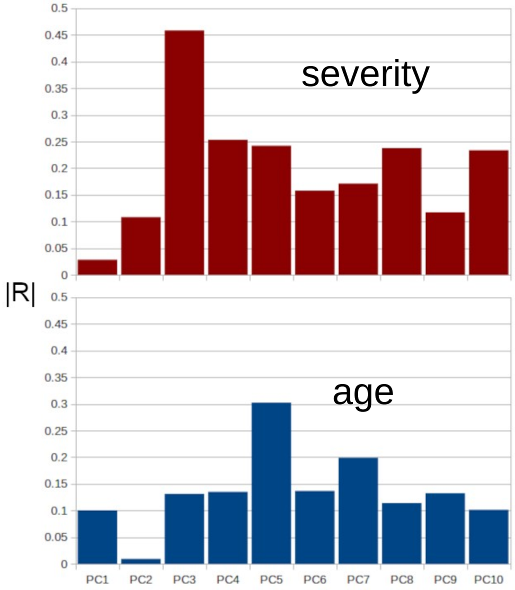


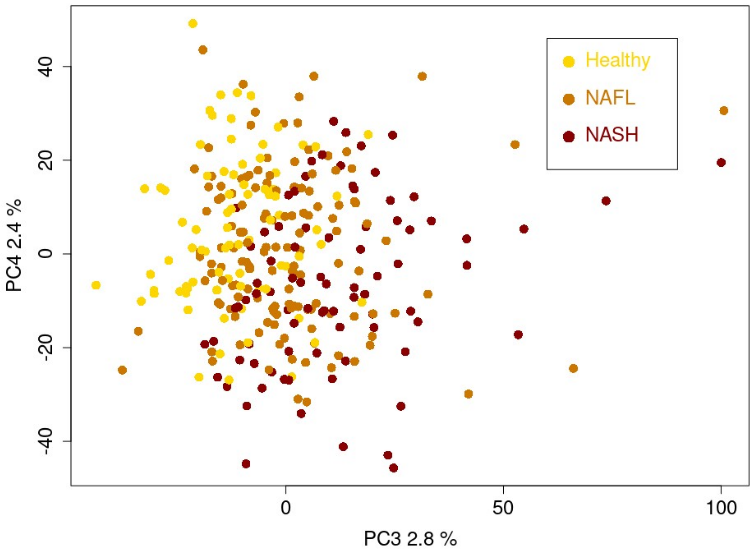


16 clinical variables
+ sex
+ age



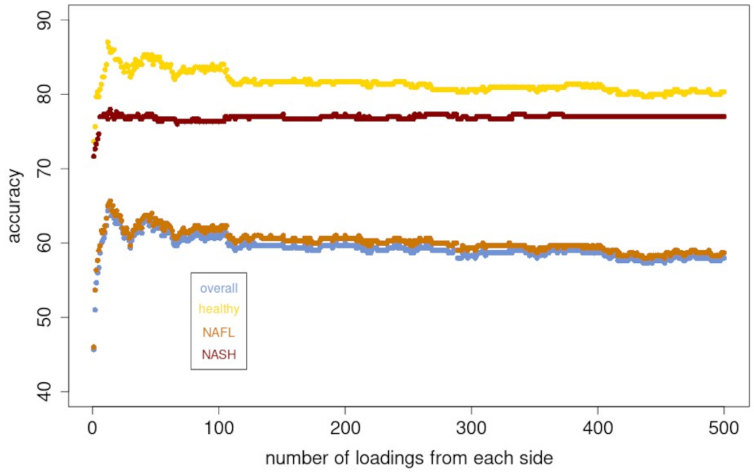
RNA-seq

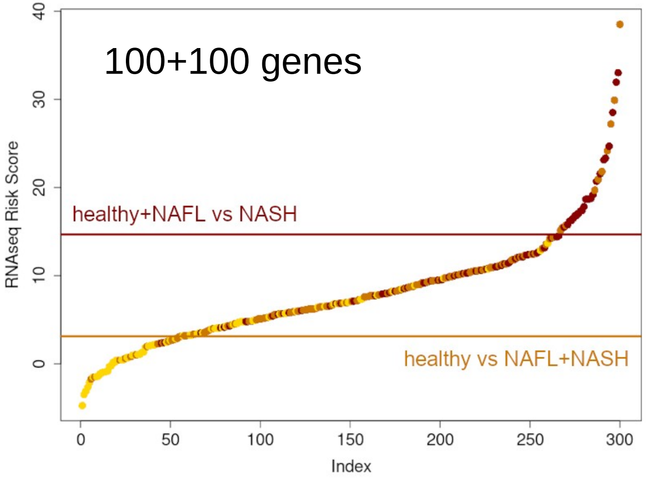




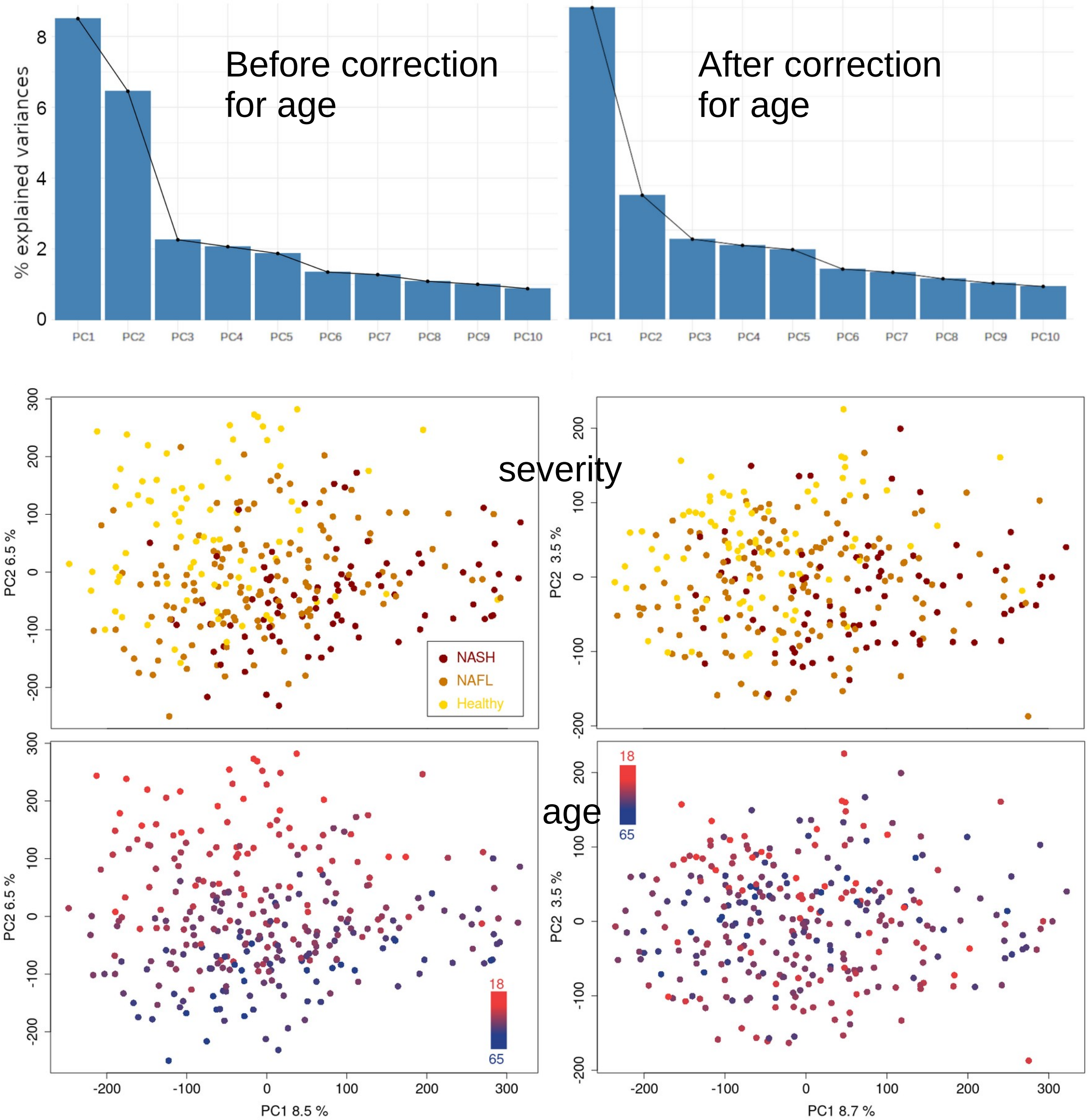
Score based on gene expression and gene “loadings” (impact of a gene on a given principal component)

Logistic regression to find the thresholds best separating the severity groups





DNA Methylation

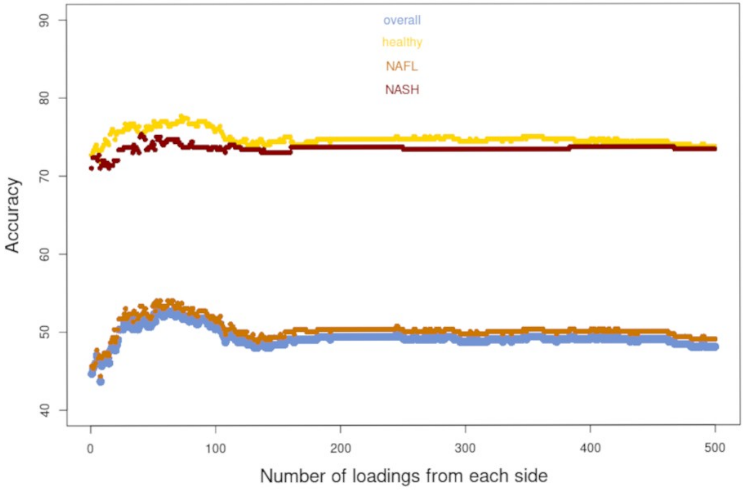


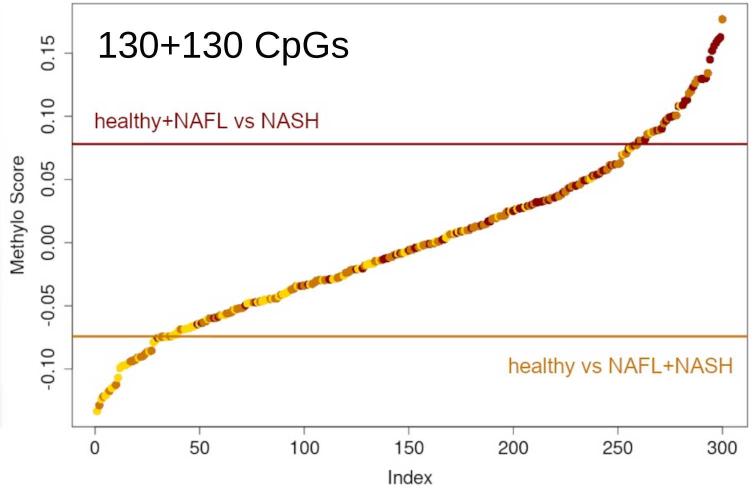
Before correction for age

After correction for age

severity

age





130+130 CpGs

healthy+NAFL vs NASH

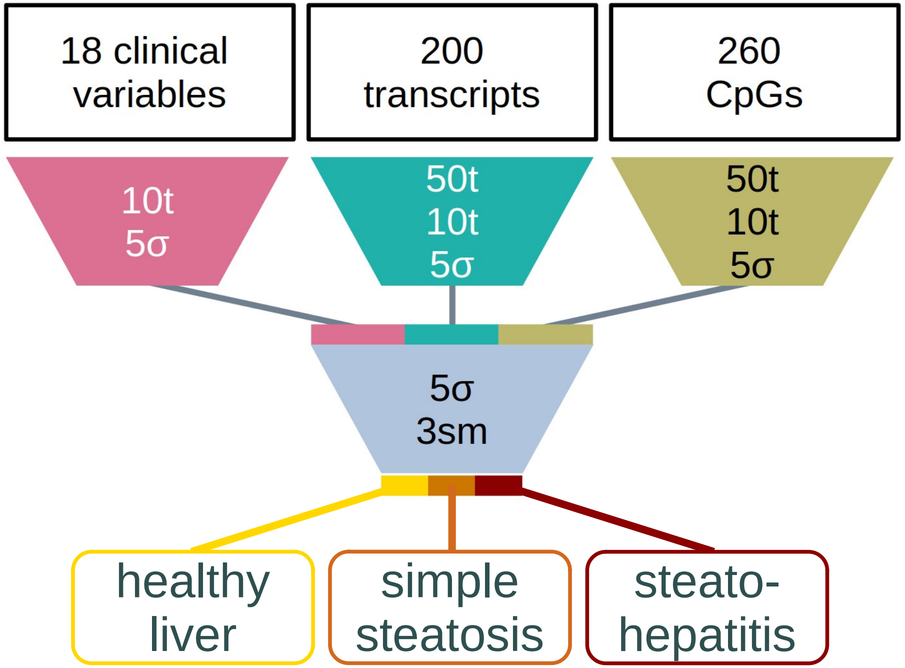
healthy vs NAFL+NASH

Multi-module MLP

Gene expression

Clinical data

DNA methylation



18 clinical variables

200 transcripts

260 CpGs

10t
5σ

50t
10t
5σ

50t
10t
5σ

5σ
3sm

healthy liver

simple steatosis

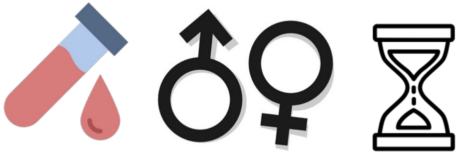
steato-hepatitis

13 layers; 25 000 parameters

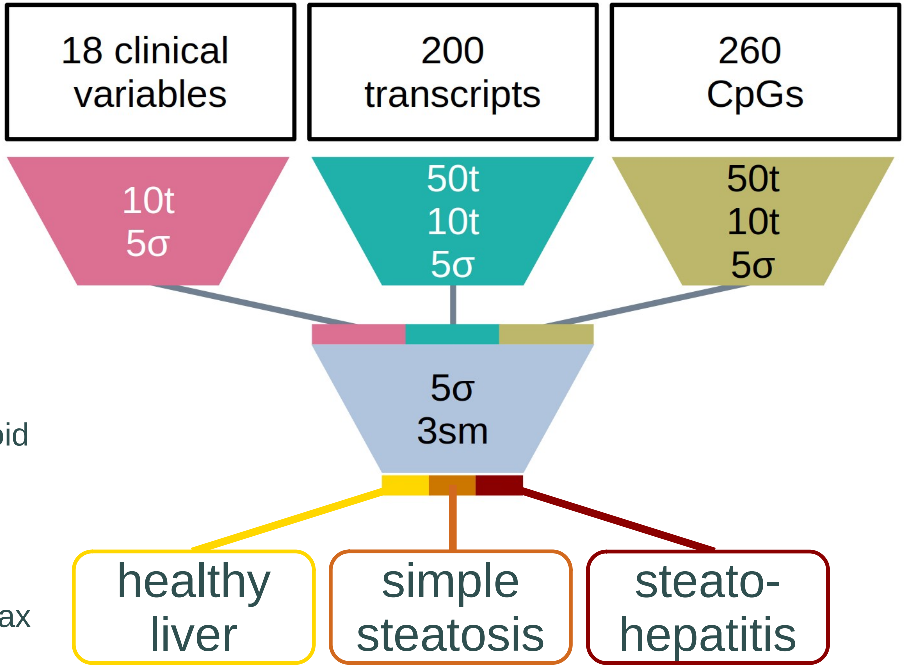
Multi-module MLP

Gene expression

Clinical data



DNA methylation



18 clinical variables

200 transcripts

260 CpGs

10t
5σ

50t
10t
5σ

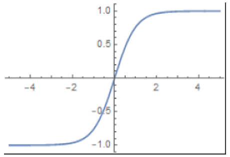
50t
10t
5σ

5σ
3sm

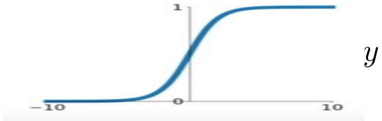
healthy liver

simple steatosis

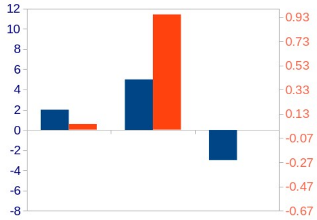
steato-hepatitis



$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$ Tanh



$$y = \frac{1}{1 + e^{-x}}$$ Sigmoid



$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$$ Softmax

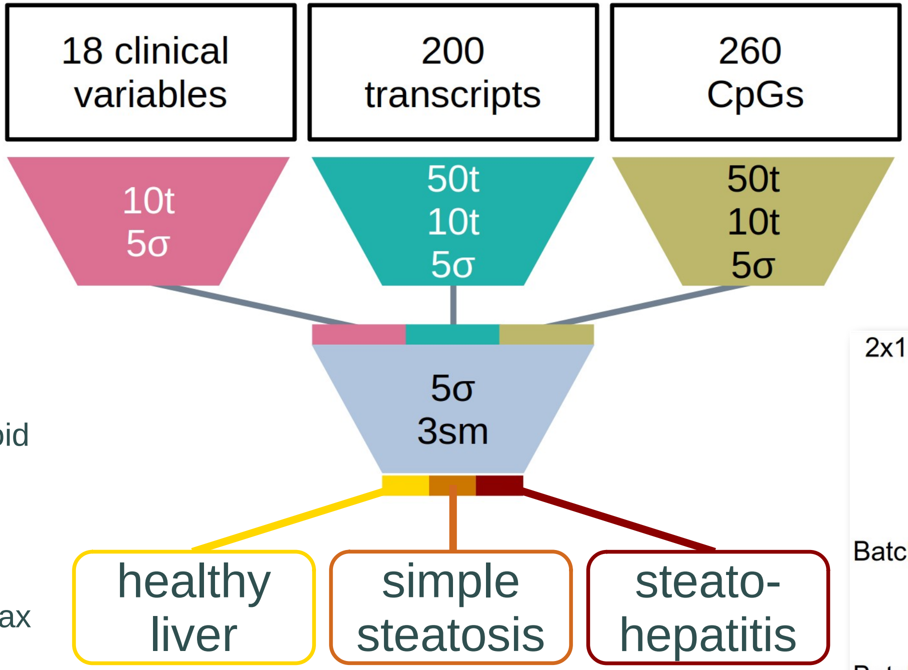
13 layers; 25 000 parameters

Multi-module MLP

Gene expression

Clinical data

DNA methylation



18 clinical variables

200 transcripts

260 CpGs

10t
5σ

50t
10t
5σ

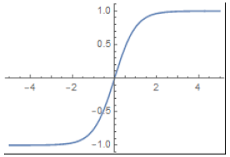
50t
10t
5σ

5σ
3sm

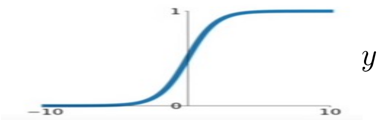
healthy liver

simple steatosis

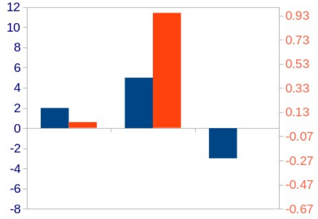
steato-hepatitis



$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$ Tanh

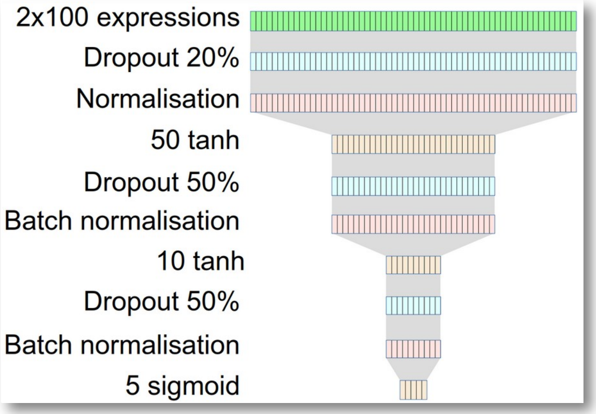


$$y = \frac{1}{1 + e^{-x}}$$ Sigmoid



$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$$ Softmax

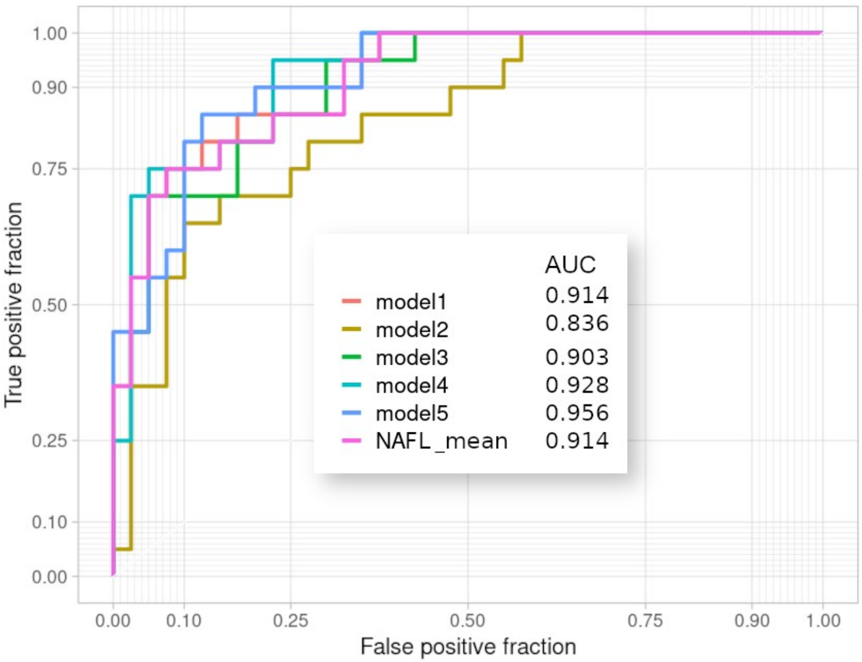
13 layers; 25 000 parameters



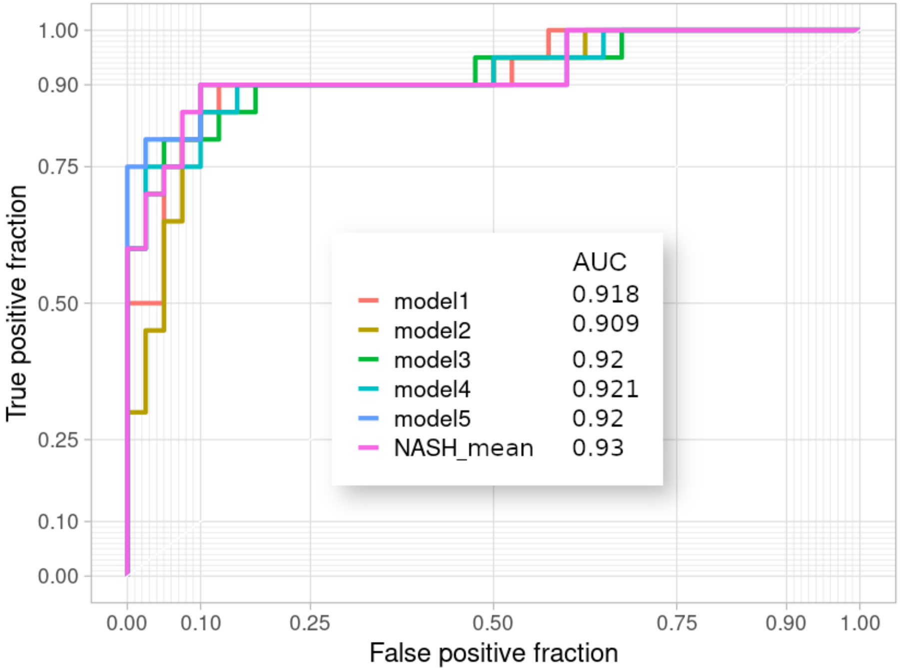
2x100 expressions
Dropout 20%
Normalisation
50 tanh
Dropout 50%
Batch normalisation
10 tanh
Dropout 50%
Batch normalisation
5 sigmoid

Performance on validation set: AUC > 0.9 for NAFL and NASH

sensitivity

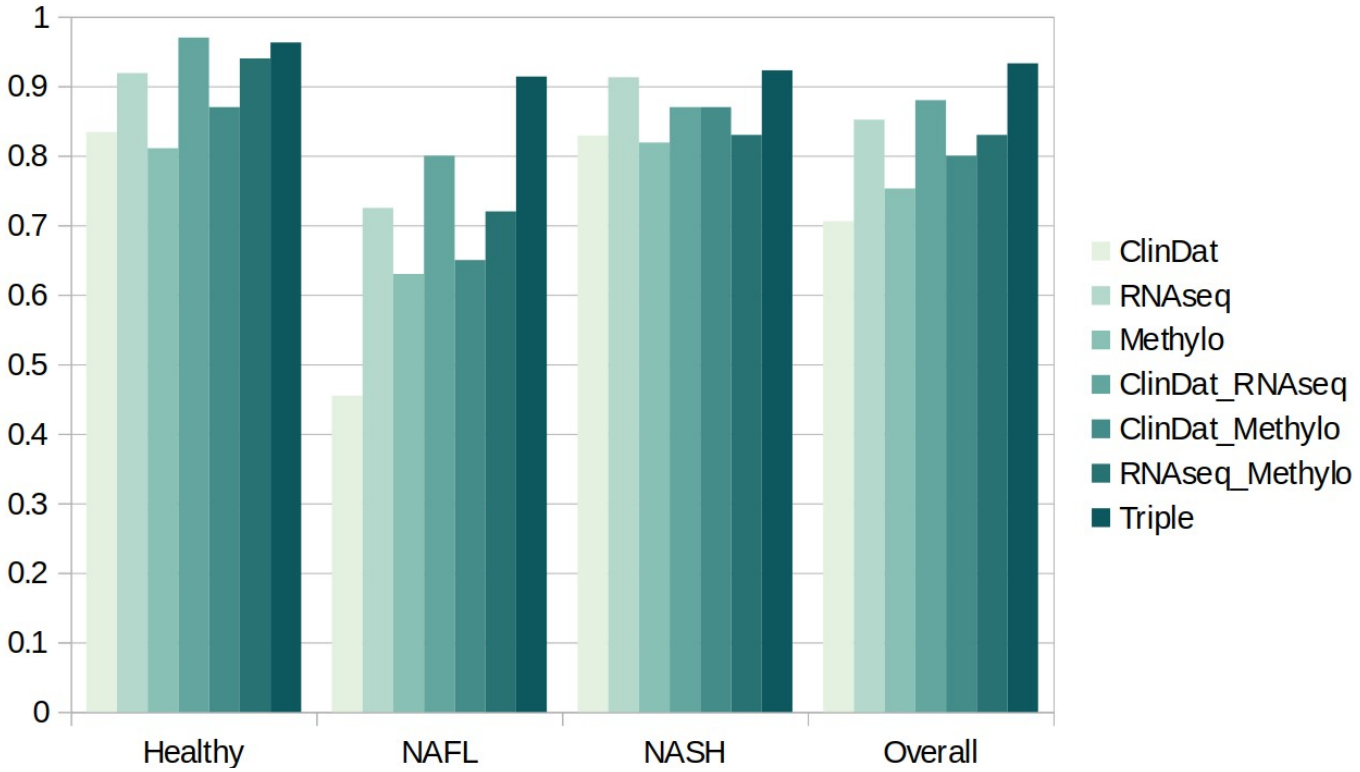


1 - specificity

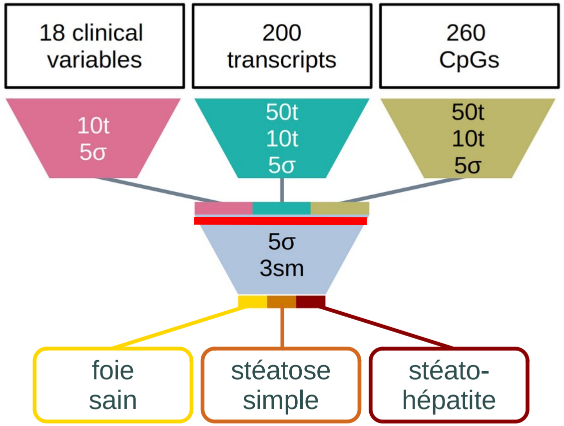


1 - specificity

Performance of the different modules



Peeping in the black box: RNA-seq is key


18 clinical variables

200 transcripts

260 CpGs

10t
5σ

50t
10t
5σ

50t
10t
5σ

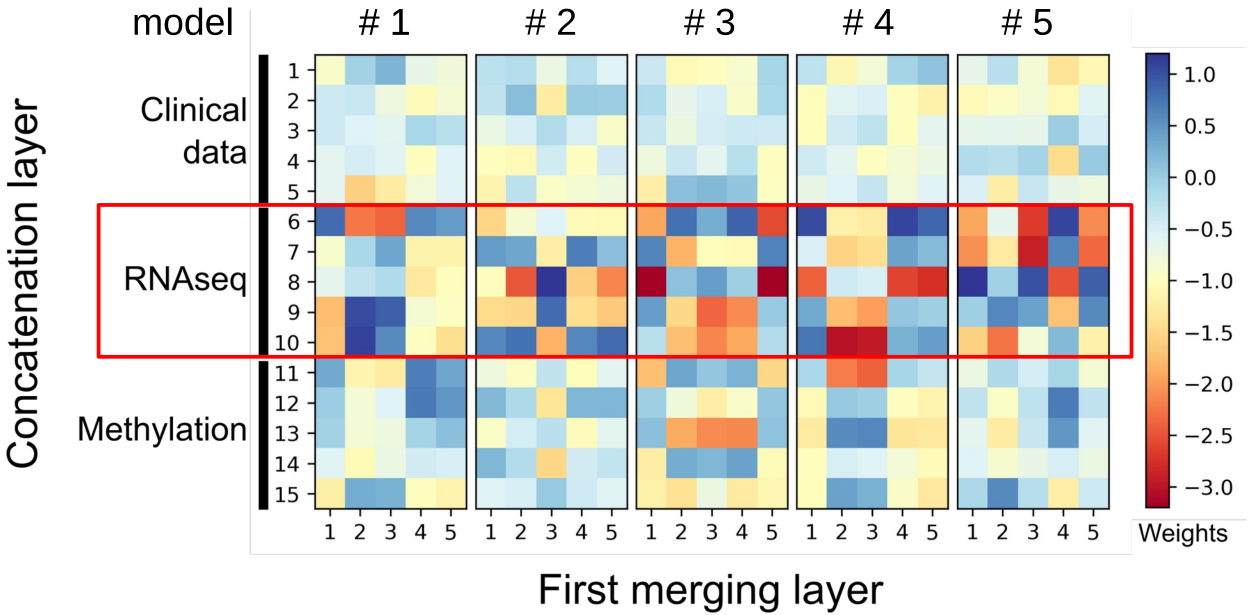
5σ
3sm

foie sain

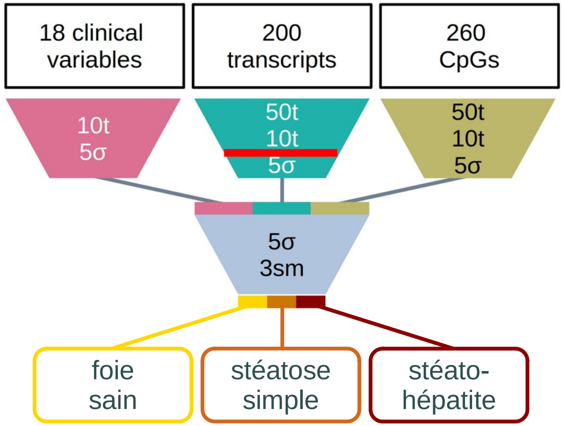
stéatose simple

stéato-hépatite

The weights reading the RNAseq module are larger → most impact on output

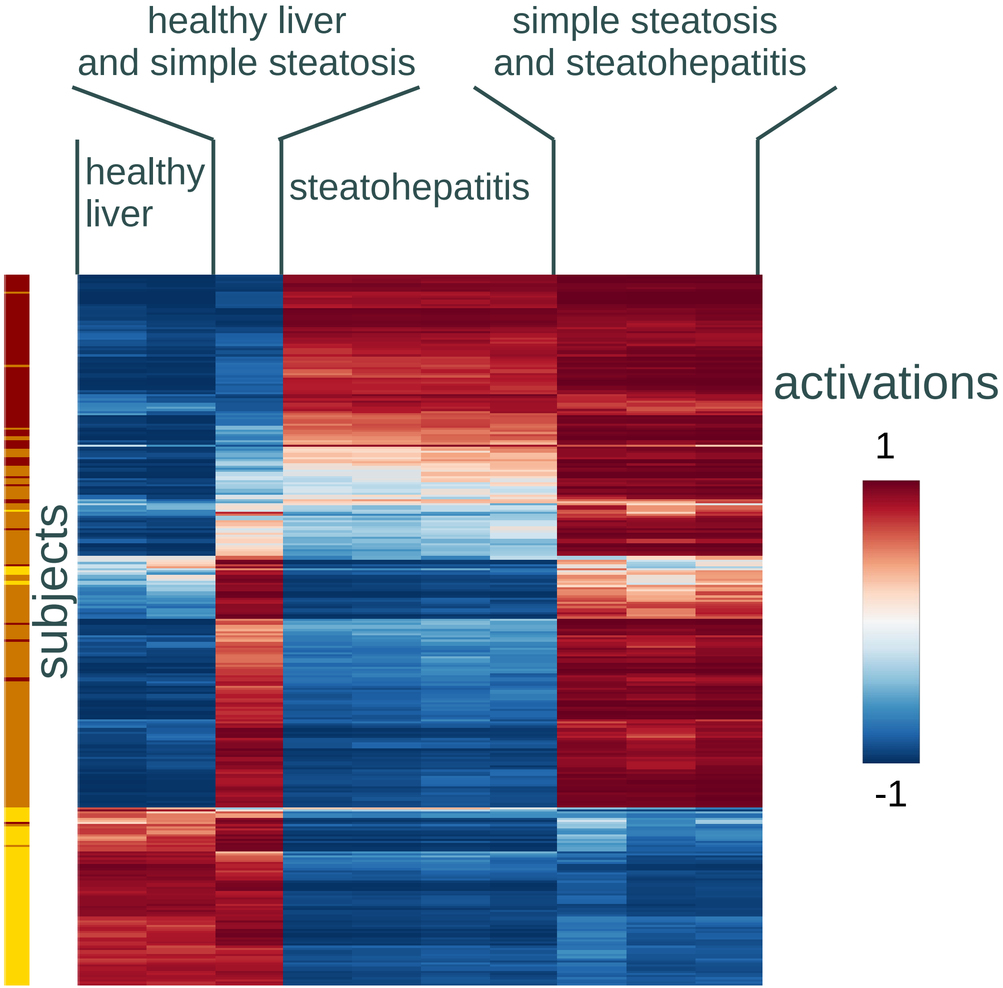


Peeping in the black box: RNA-seq is key

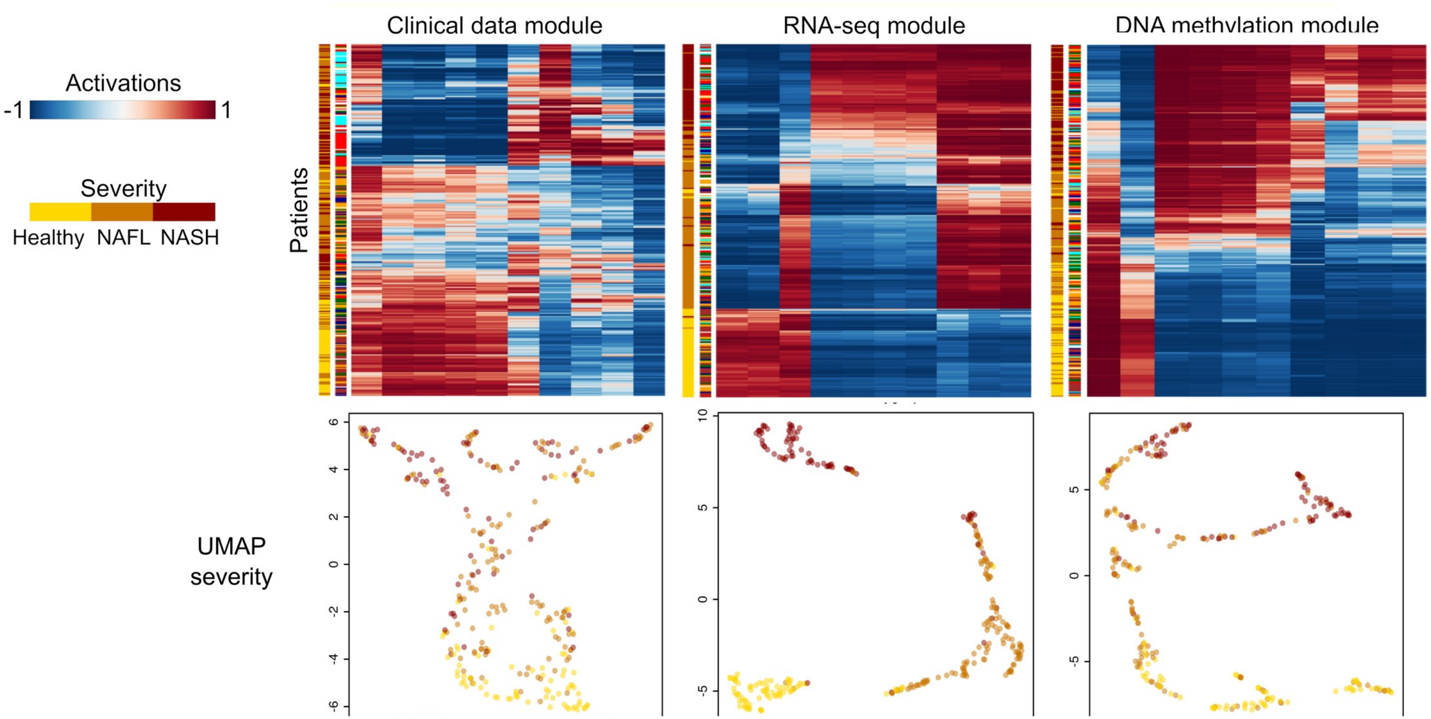


severity

healthy liver	🟨
simple steatosis	🟧
steato-hepatitis	🟥



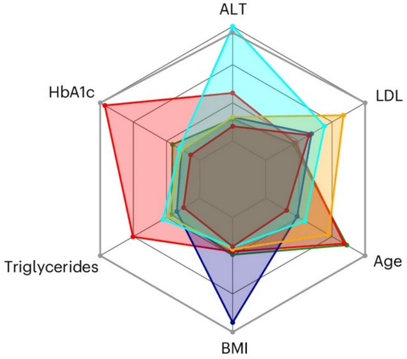
Exploring latent spaces of each module

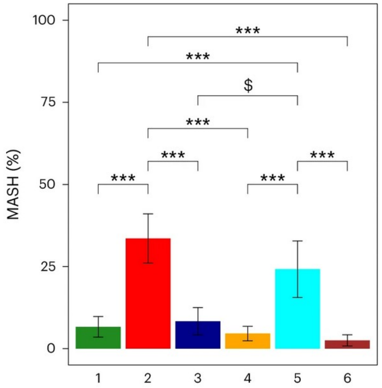


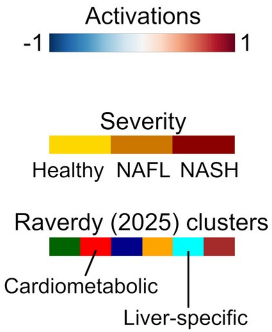
RNA-seq: recognises the 3 conditions

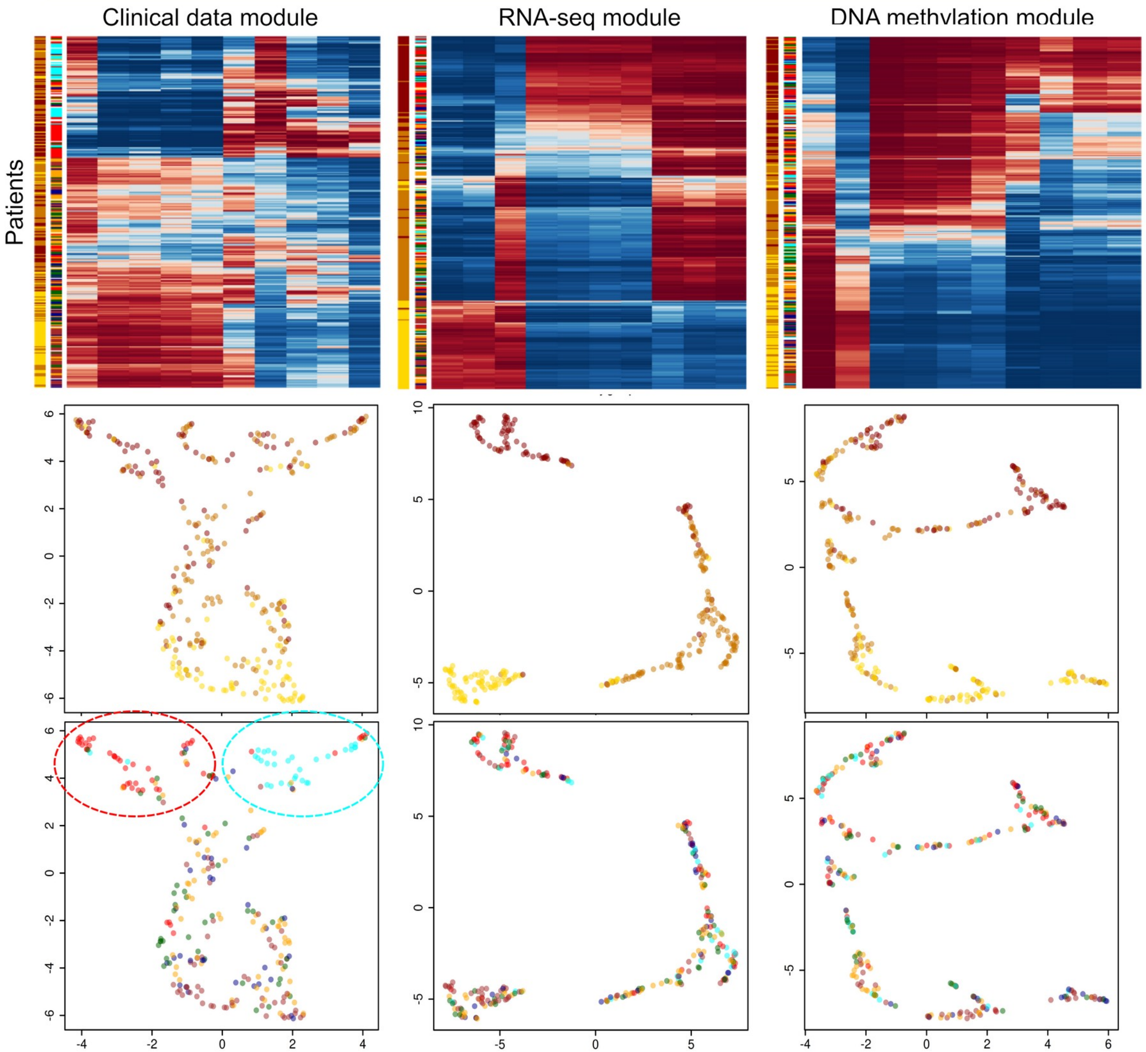
Methylation: continuum of severity

Exploring latent spaces









UMAP severity

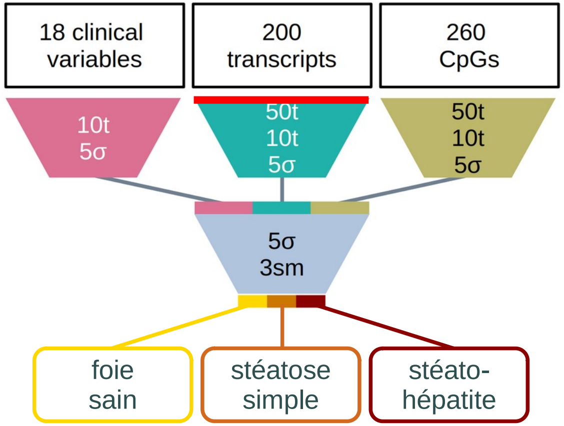
UMAP Raverdy clusters

Clinical data: several populations
See Raverdy *et al. Nat Med* 30:3624–3633 (2024)

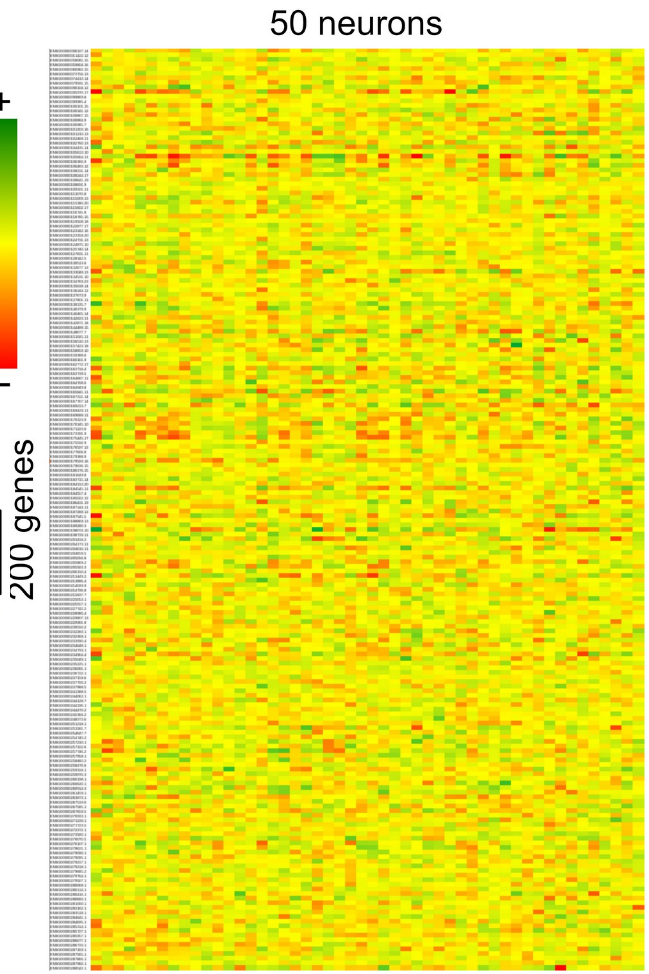
RNA-seq: recognises the 3 conditions

Methylation: continuum of severity

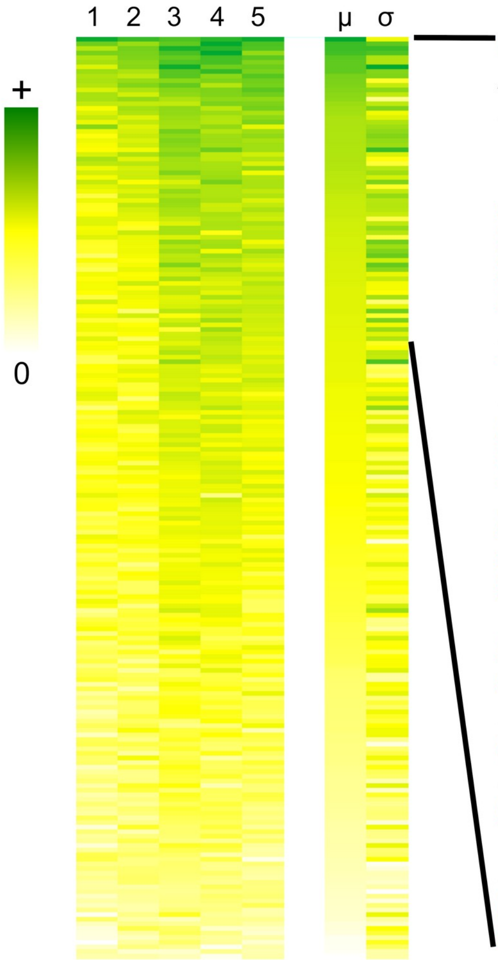
What are the genes taken into account by the models?



“synaptic weights”



|Weight|

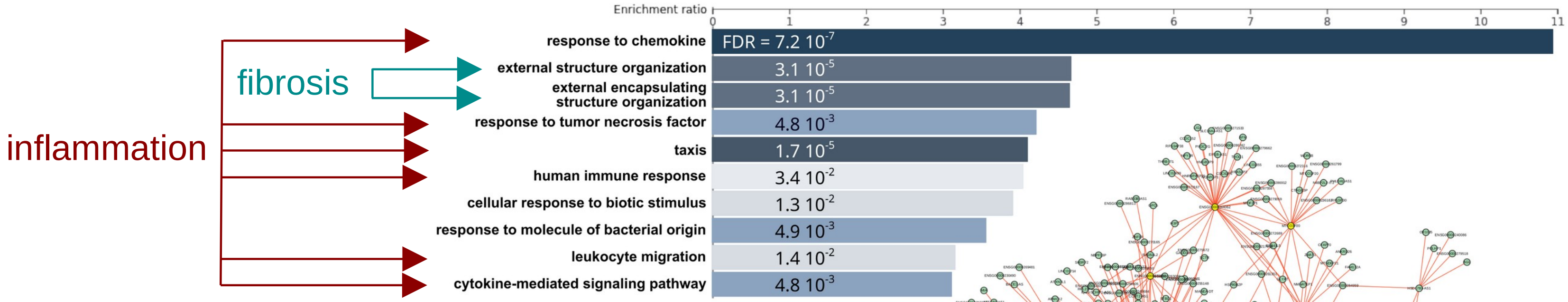


Known involvment in fatty liver

New genes

- *COMP*
- *ANKRD1*
- *PRAMEF10*
- *SFRP4*
- *CHI3L1*
- *unknown*
- *RAB3B*
- *FABP5P7*
- *KRTAP5-1*
- *PGAM2*
- *THBS1-AS1*
- *FABP4*
- *PADI1*
- *CXCL3*
- *THY1-AS1*
- *RGS1*
- *unknown*
- *PRAMEF33*
- *GDF15*
- *PNPLA5*
- *GPR158*
- *FCAR*
- *LINC02348*
- *MMP7*
- *ESPNL*
- *CYP1A1*
- *MT1B*
- *KRTAP5-AS1*
- *SDHAP2*
- *TREM2*
- *AKR1B10*
- *LPL*
- *STMN2*
- *DUSP8*
- *unknown*
- *GAPDHP28*
- *CA12*
- *unknown*
- *CYP2C19*
- *SPP1*
- *ART5*
- *EEF1A2*
- *unknown*
- *CH25H*
- *OLR1*
- *THBS1-IT1*
- *CEMP1*
- *DHRS2*
- *ALOX15B*
- *CCL20*
- *HKDC1*
- *MMP9*
- *unknown*
- *KRT80*
- *TRIM31*
- *BCL2A1*
- *LINC00940*
- *FOS*

Inferred gene network

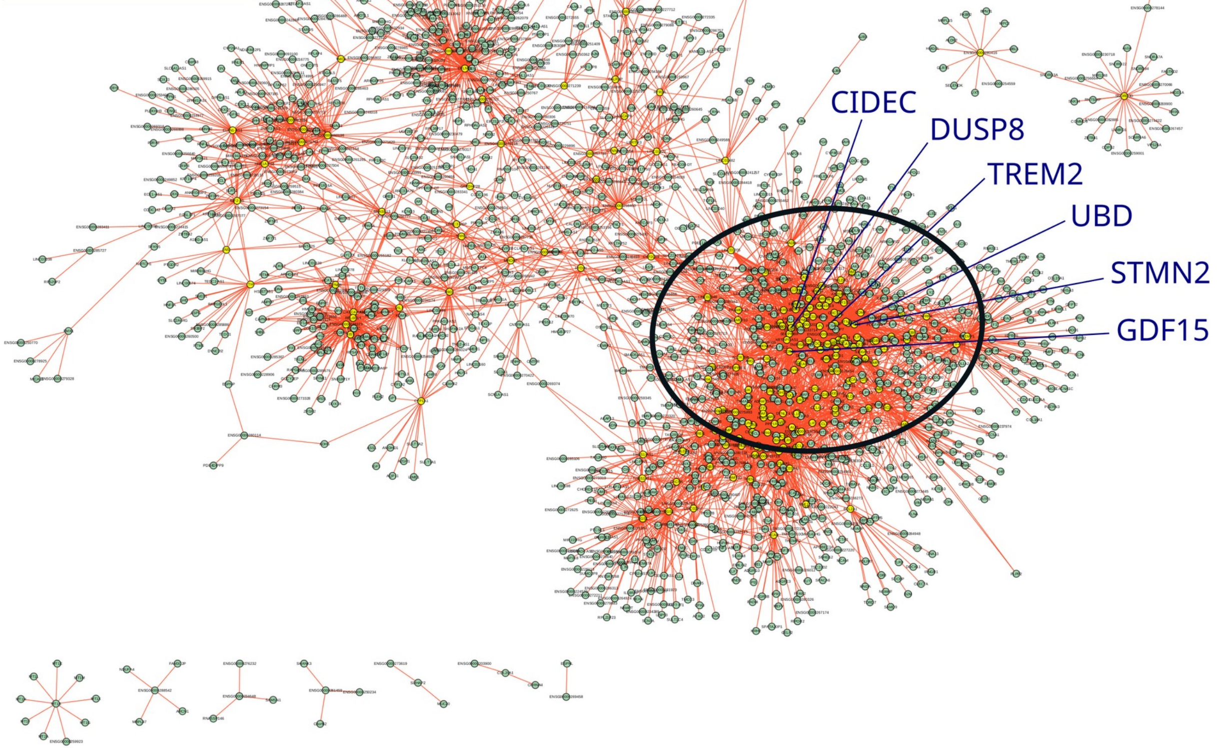


RNA-seq dataset

Inference with CLR+GENIE3
(584 358 193 non-zero edges)

Only interactions involving
our 200 genes
(3 162 288 interactions)

Knee method
(1 698 inteaction)



Acknowledgments

The group

Smaïn Fettem
Marwa Afnouch
Omaima Binan
Salla Åkerblom
Zahra Elhamraoui
Sami Le Guilcher
Arthur Six

The tools

R package developers
(*VIM*, *MICE*, *DESEq2*, *ChAMP*)

Python package developers
(*Tensorflow/Keras*, *Numpy*,
Pandas, *Scikit-learn*)

PreciDIAB
National Center for
Precision Diabetic Medicine

cnrs

Université
de Lille

UMR 1283 8199
Multi-omique et physiopathologie des maladies métaboliques

Metab-Omics

Philippe Froguel
Amélie Bonnefond

Mathilde Boissel
Mehdi Derhourhi
Emma Henriques
Stefan Gaget
ShuangShuang Geng
Anne-Sophie Ledoux
Vincent Massy
Lijiao Ning

Marc-Emmanuel Dumas
Amna Khamis
Francesc Puig-Castellvi

Hélène de Gavre
Mélanie Hocquet

PreciNASH

François Pattou
Violetta Raverdy
Bart Staels
Philippe Lefebvre

TRD
Translational Research for Diabetes

UMR 1011
Nuclear Receptors, Metabolic
& Cardiovascular Diseases

anr

Région
Hauts-de-France