

# Uncovering relationships between stress and psychiatric traits via multi-omics individual-specific networks

Anastasiia Hryhorzhevska

23.03.2023

MAX PLANCK INSTITUTE  
OF PSYCHIATRY



HELMHOLTZ  
MUNICH

LMU

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

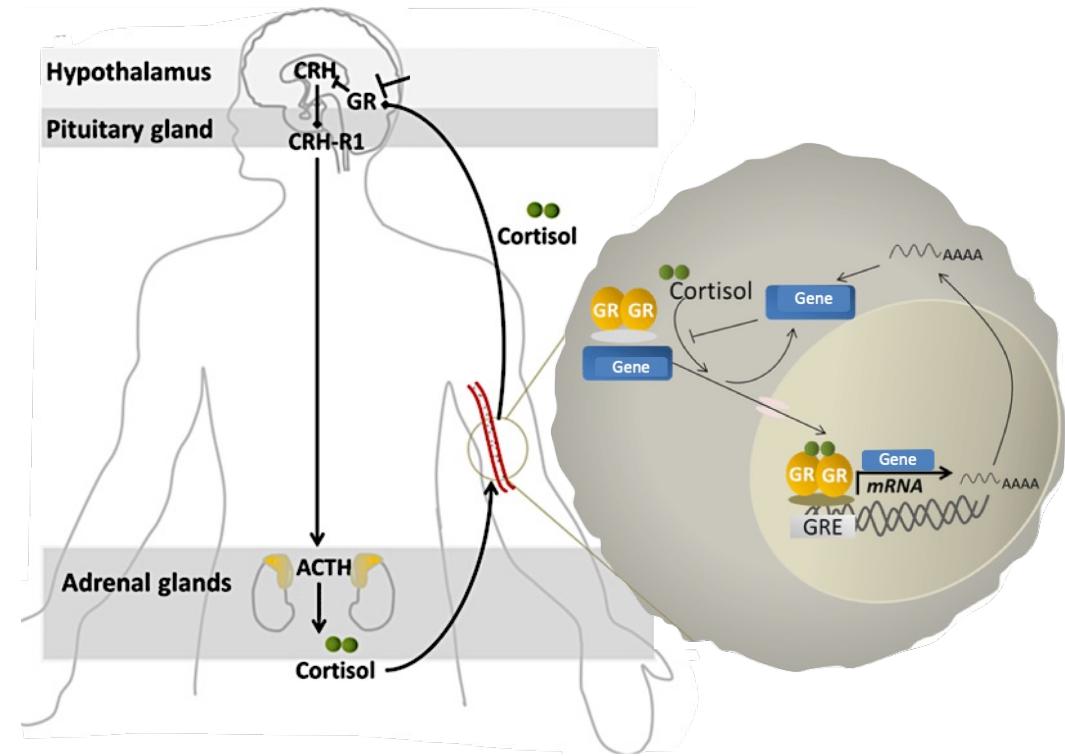
KU LEUVEN

TRANSYS  
PERSONALIZED MEDICINE



# Background

- Stress leads to release of glucocorticoids (cortisol) through activation of the hypothalamic-pituitary-adrenal (**HPA**) axis.
- Upon activation, glucocorticoid receptor(**GR**) translocates from cytoplasm to nucleus and regulates the expression of target genes.
- GR-induced transcriptional activation may mediate risk for psychiatric disorders.
- **Dexamethasone** is a long-lasting corticosteroid.



Source: Ramo-Fernández et al., 2019

# Individual Specific Networks. What?

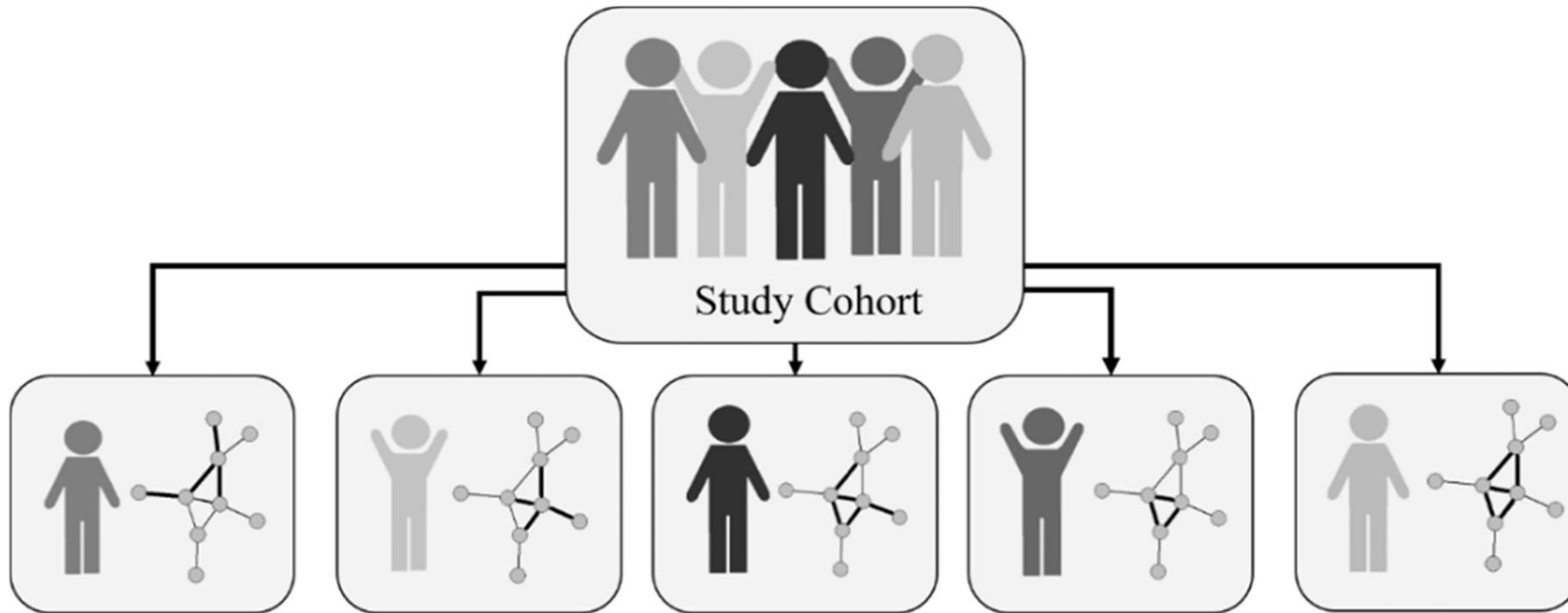


Illustration of the potential heterogeneity inherent in 5 individual-specific networks derived from a study cohort of sample size 5. Line thickness between a pair of nodes indicates the strength of the association (edge weight) within that specific network due to individual-specific variable measurements (Gregorich et al., 2022).

## Individual Specific Networks. Why?

1. A network derived from a collection of individuals can be used as a model for an “average” individual.
2. Individual-specific networks allow focusing on each individual and their specific associations and dynamics over time.

# Individual Specific Networks. How?

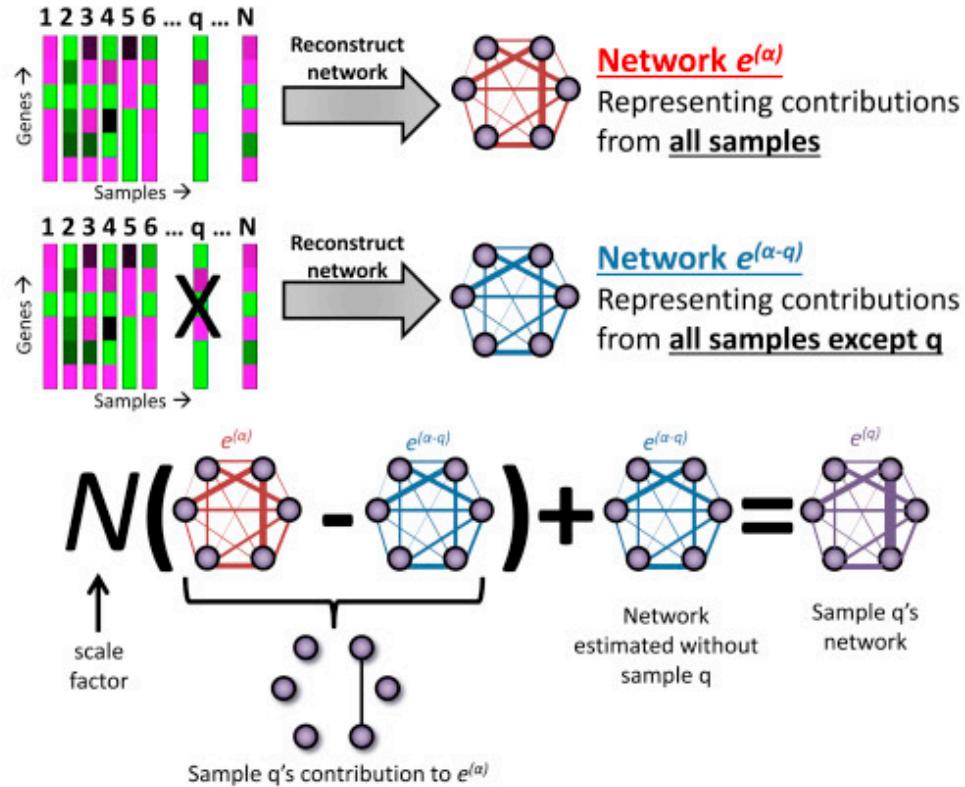


Illustration of the approach to estimate the network for a single sample based on two aggregate network models, one reconstructed using all biological samples in a given dataset and the other using all except the sample of interest, q (Kuijjer et al., 2019).

# Individual Specific Networks. How?

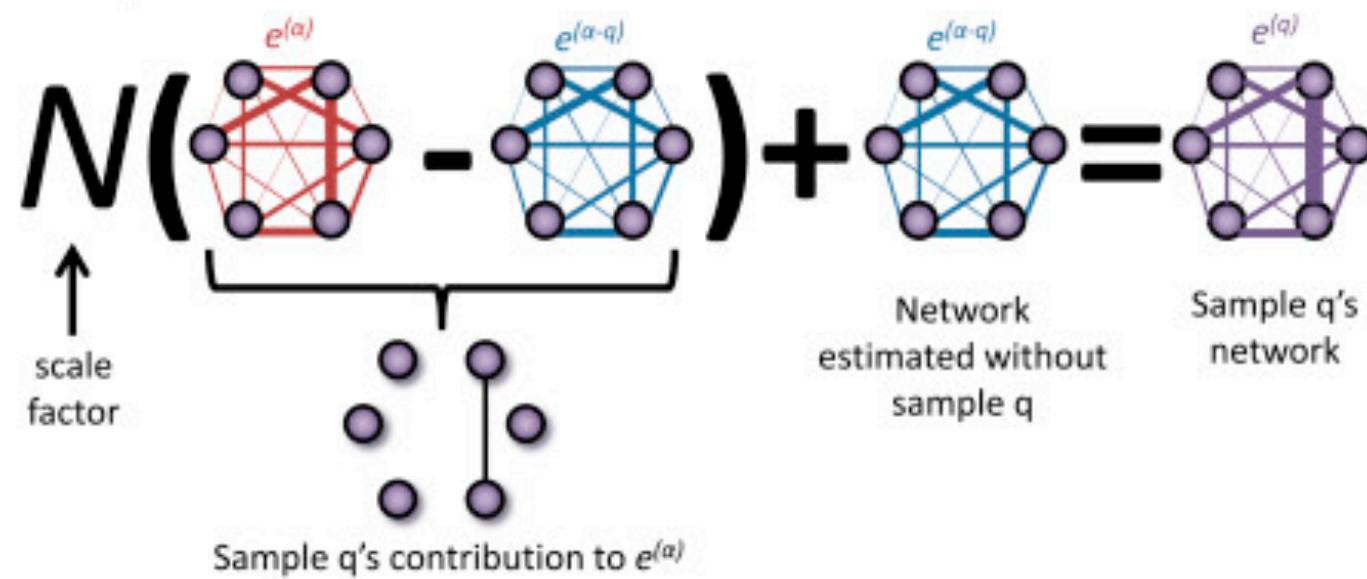


Illustration of the approach to estimate the network for a single sample based on two aggregate network models, one reconstructed using all biological samples in a given dataset and the other using all except the sample of interest,  $q$  (Kuijjer et al., 2019).

# Data

## MPIP Cohort

196 individuals

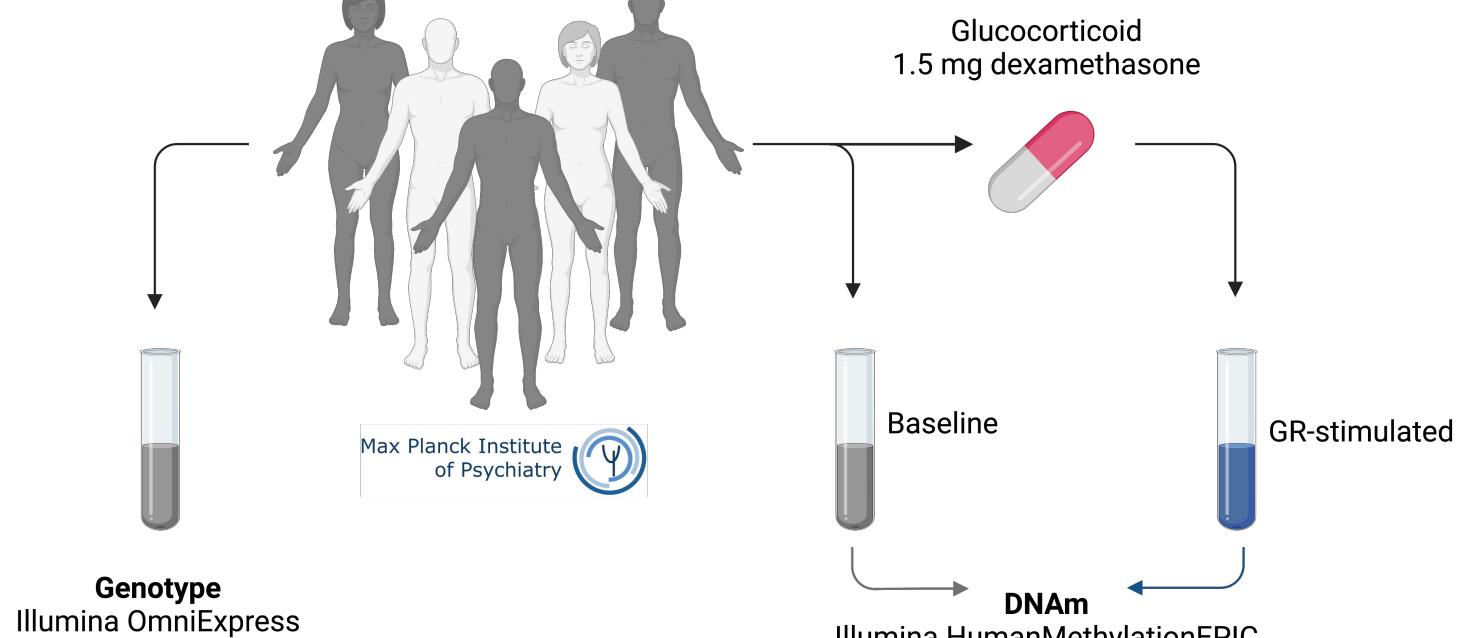
- 66% males
- 44% with MDD

## Genotype data

Illumina OmniExpress &  
genome-wide imputation  
ca. 4 000 000 SNPs

## DNAm data

Illumina HumanMethylationEPIC  
BeadChip  
740 358 CpGs



*Created with BioRender.com*

## Research Aims

1. How are the samples grouped according to the changes before and after stimulation with dexamethasone?
2. What are the main drivers of these changes?

# Overview of the general study design

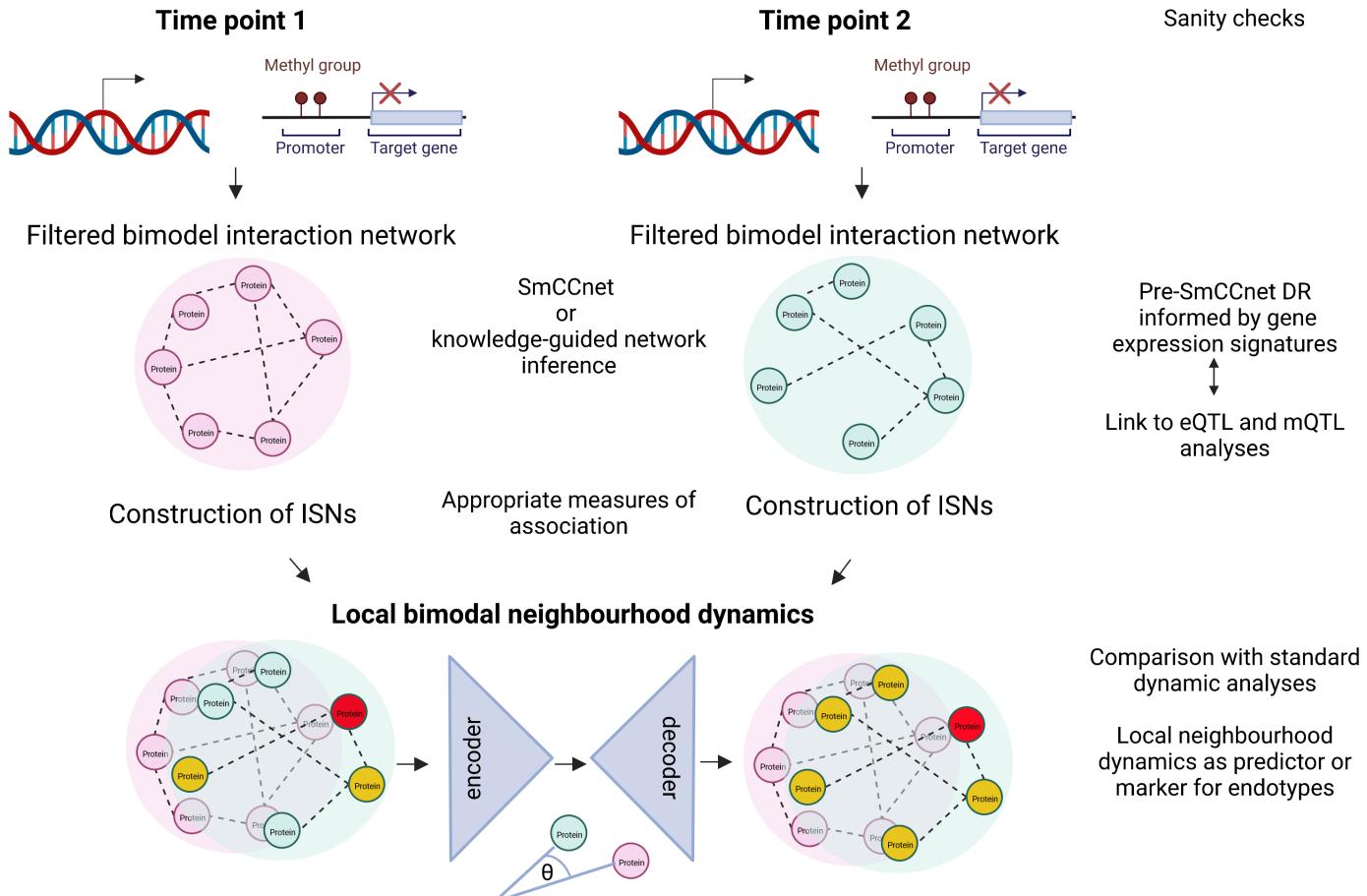
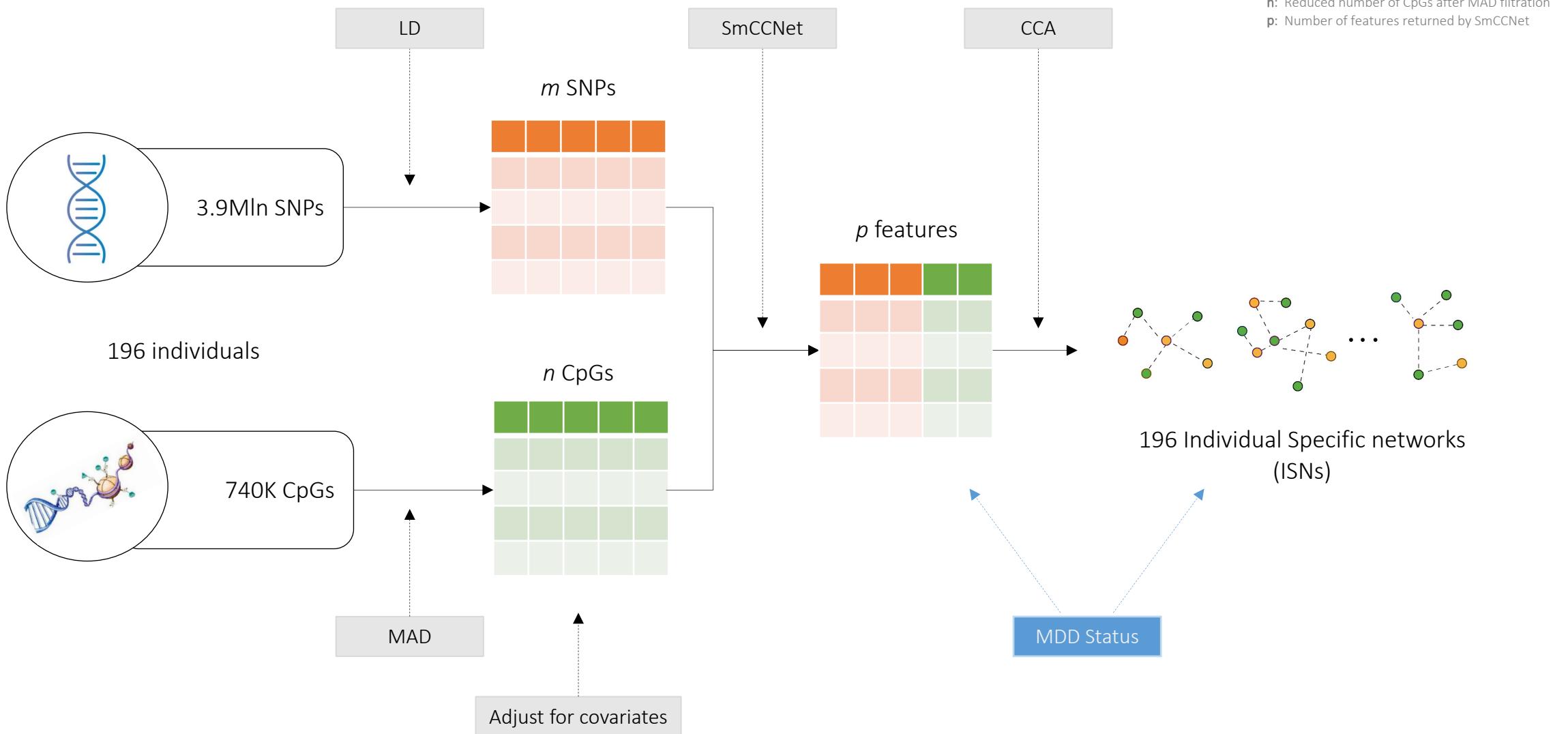


Illustration of the workflow of the general study design by Prof. Dr. Dr. Kristel Van Steen

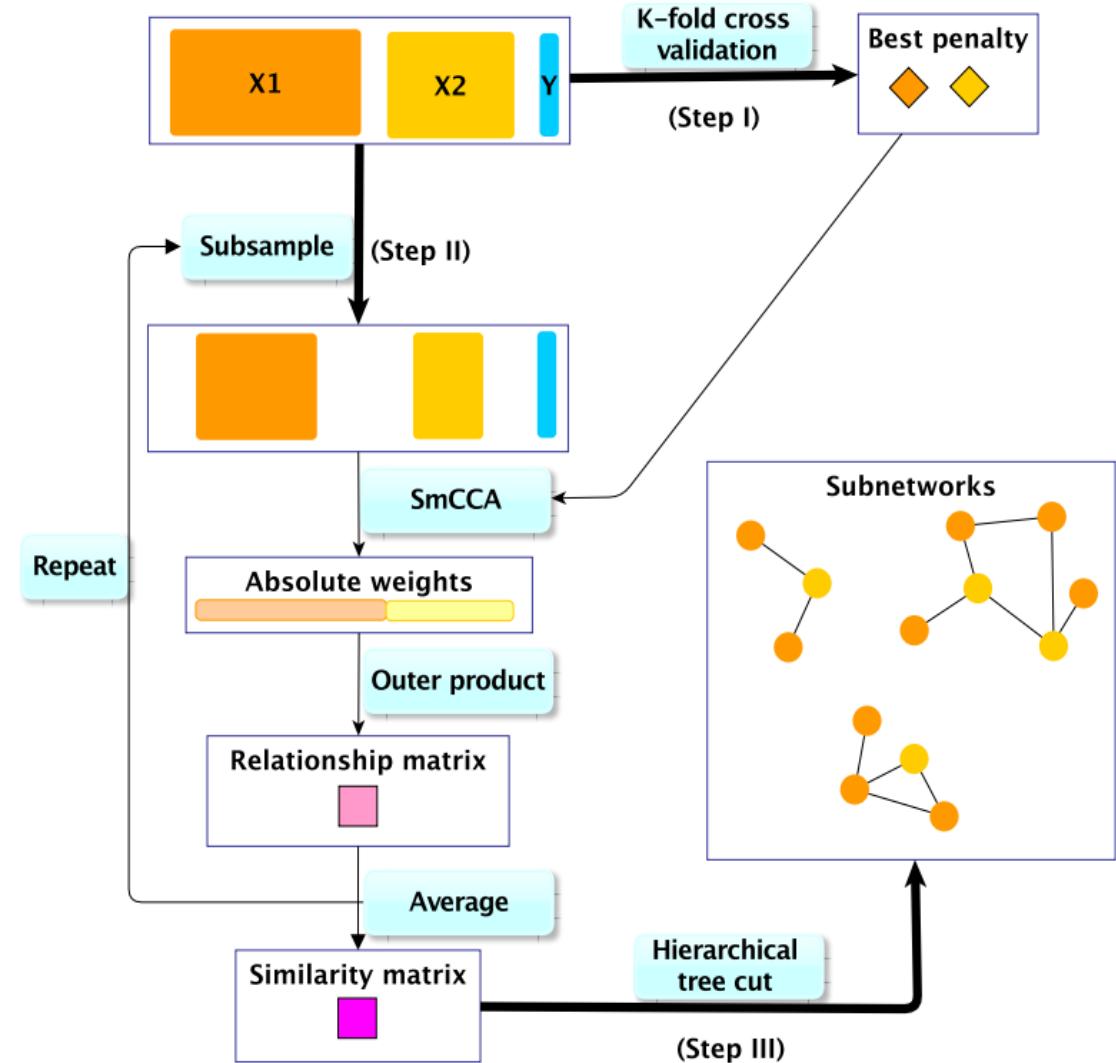
Created with [BioRender.com](#)

# Workflow. Construction of ISNs



# Workflow. SmCCNet

- Sparse multiple Canonical Correlation Network analysis is a method which integrates multiple omics data and phenotype information to build interpretable networks that model the underlying mechanisms.
- With preselected sparsity penalties, it finds multi-omics modules which are relevant for the phenotype.
- Workflow
  1. Determine sparsity penalties
    - prior knowledge or
    - cross-validation
  2. Define similarity matrix by using features subsampling
  3. Hierarchical clustering to the multi-omics networks



SmCCNet workflow overview.  $X_1$  and  $X_2$  are two omics data types for the same set of  $N$  samples.  $Y$  indicates a quantitative phenotype measure for those  $N$  samples (Shi et al. 2019)

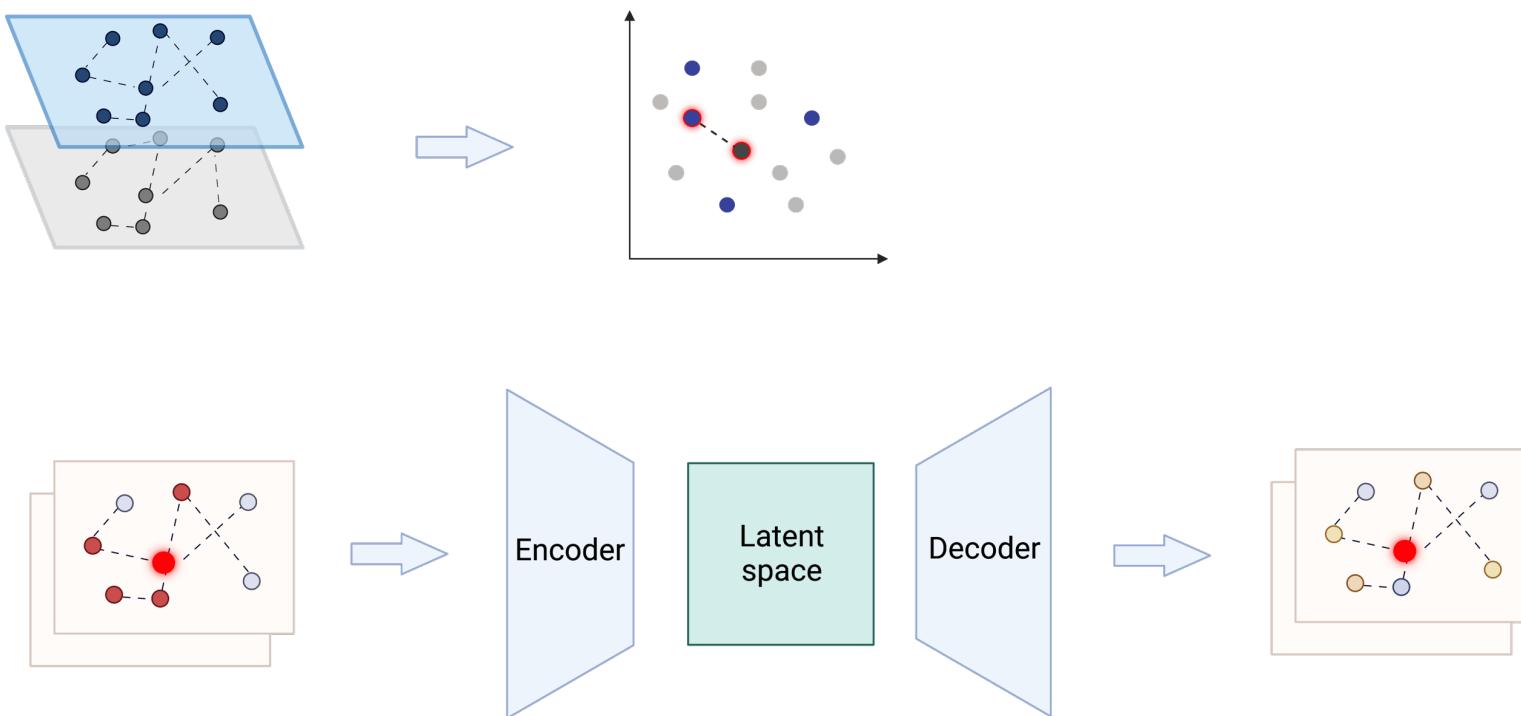
## SmCCNet: Sparse multiple Canonical Correlation Network discovery

- Apart from  $X_1, X_2$ , there's also a phenotype of interest  $Y$  that has been measured for the same  $n$  subjects.
- One way to take phenotype into account is to treat it as 3rd data type.
- The weights are to prioritize correlations with the phenotype rather than between omics data types.

$$(w_1, w_2) = \arg \max_{\tilde{w}_1, \tilde{w}_2} (a \tilde{w}_1^T X_1^T X_2 \tilde{w}_2 + b \tilde{w}_1^T X_1^T Y + c \tilde{w}_2^T X_2^T Y)$$

subject to  $\|\tilde{w}_s\|^2 = 1, P_s(\tilde{w}_s) \leq c_s, s = 1, 2.$

# Workflow. Multiplex Network Differential Analysis (MNDA)



Created with [BioRender.com](https://biorender.com)

# Results. Part I

# Attempts

## 1. All features:

DNAm: 196 x 740K

SNPs: 196 x 3.9Mln

45 cores and 900 Gb => memory issue

## 2. All features by chromosome:

### a) Chr1 was tested:

DNAm: 196 x 72'020

SNPs: 196 x 302'458

### b) 45 cores and 900 Gb

⇒ memory issue after 35 hr

## 3. Reduced matrices:

### a) DNAm

The 80% percentile was taken as the cut-off of the Median Absolute Deviation, MAD, score. The intersection of baseline and post-dexamethasone CpGs were taken

=> CpGs: 196 x 48'753

### b) SNPs

LD pruning with the parameters:

- window size in SNPs = 100

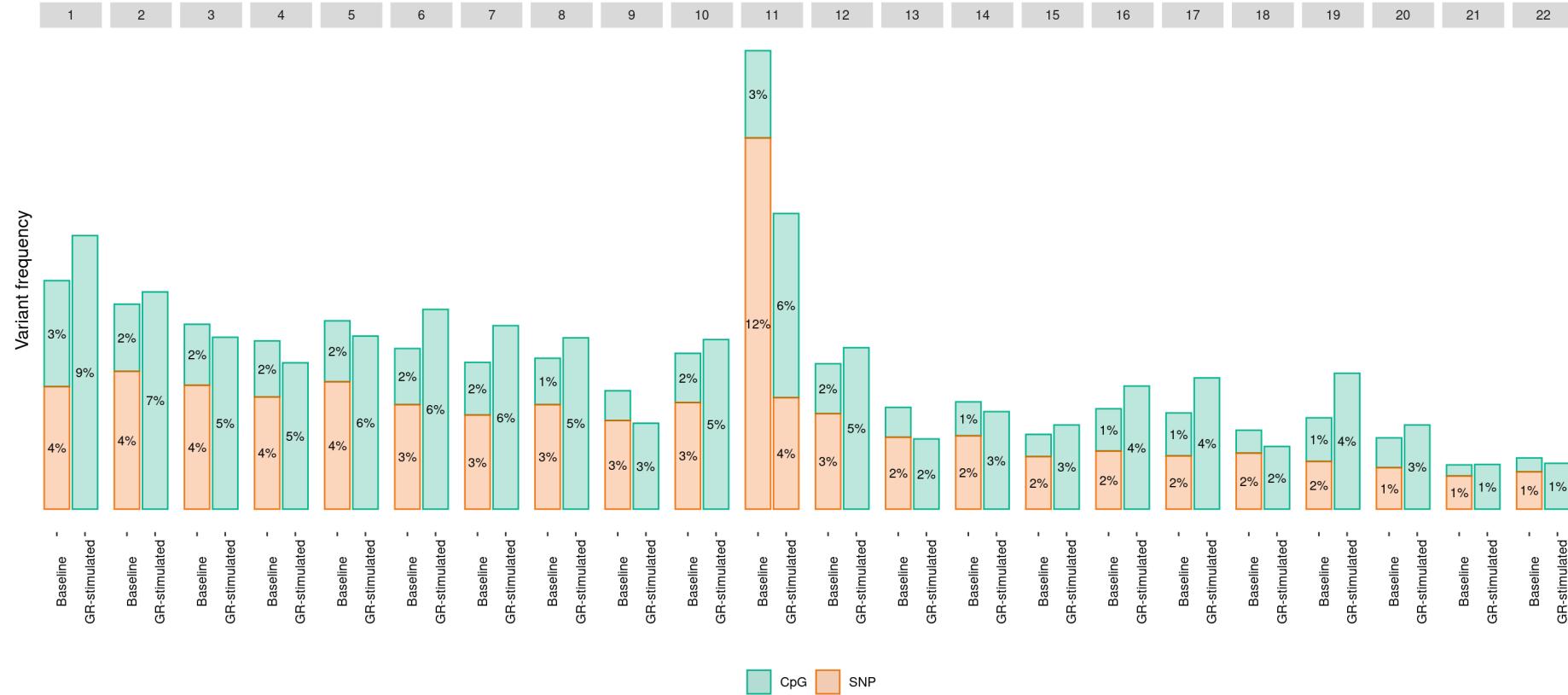
- the number of SNPs to shift the window at each step = 50

-  $r^2$  threshold = 0.2

⇒ SNPs: 196 x 124'611

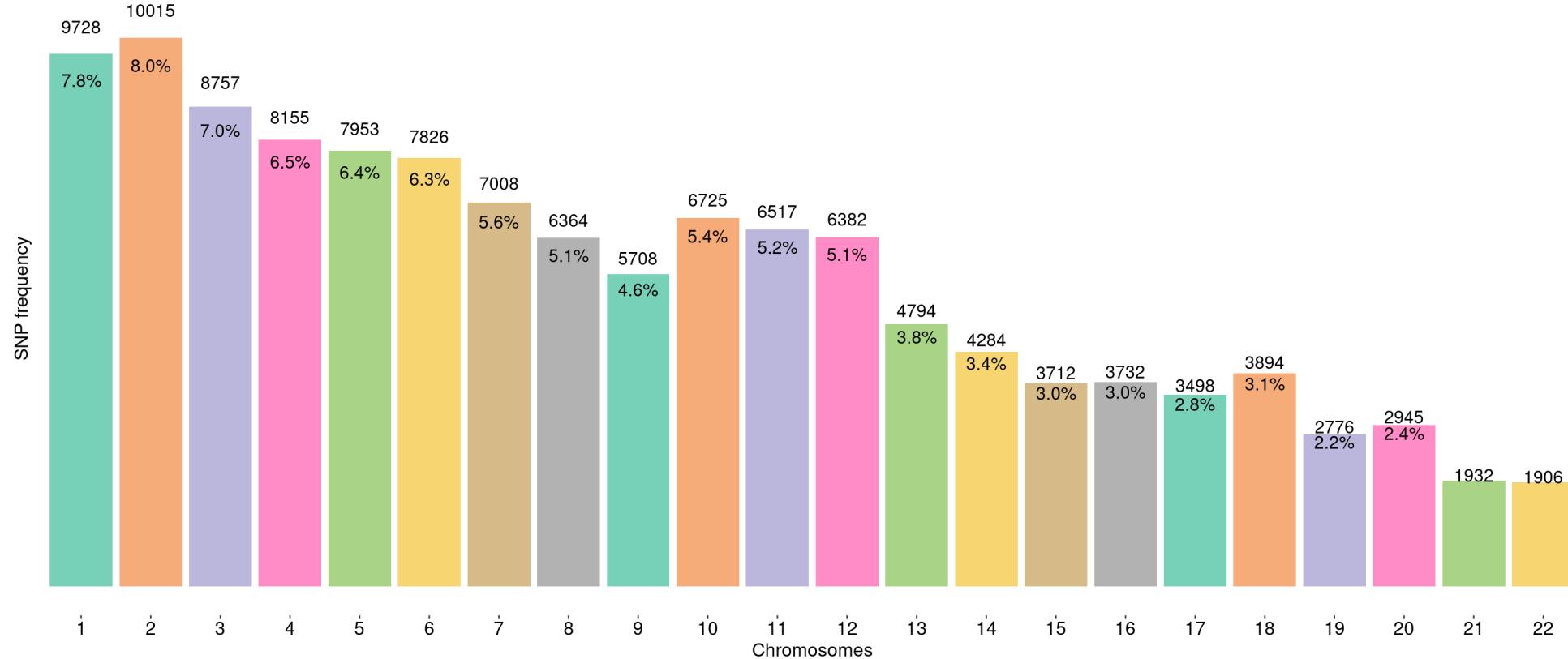
⇒ success

# Global network analysis



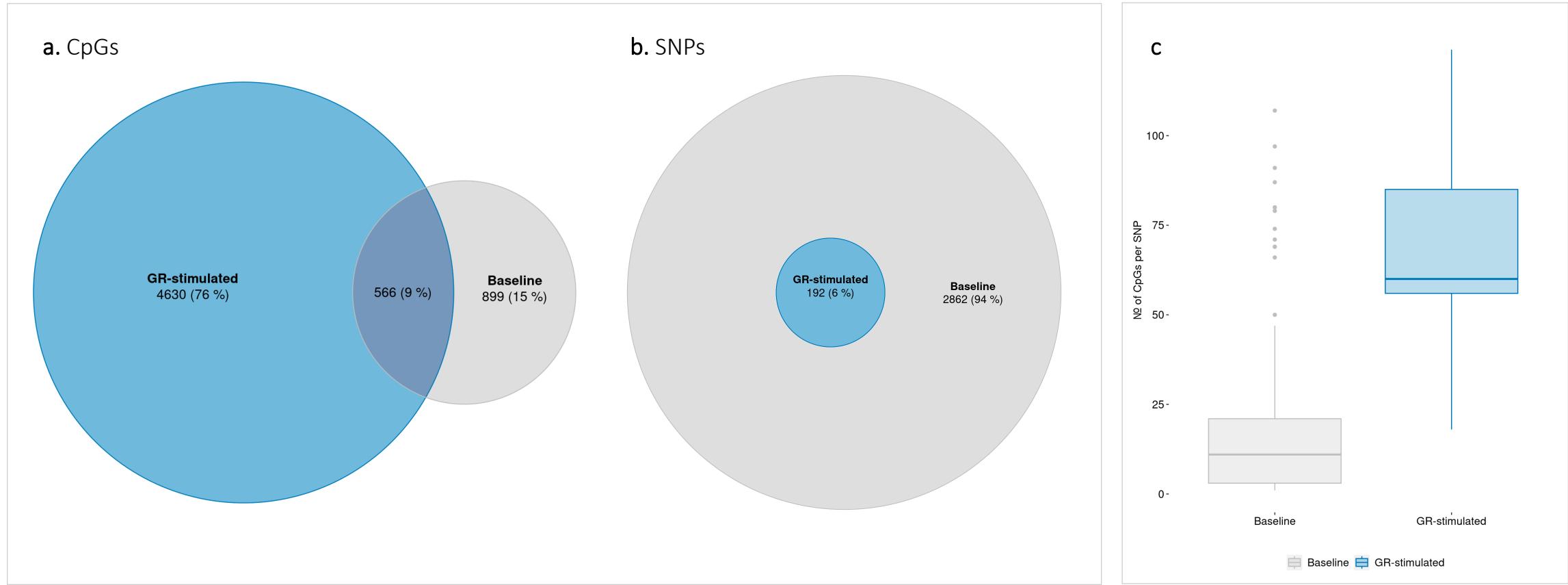
SmCCNet results. Frequency distribution of CpG sites and SNPs selected by SmCCNet across 22 chromosomes. SmCCNet penalty parameters were obtained using a 5-fold CV resulting in  $\text{I1} = 0.05$  and  $\text{I2} = 0.05$  for baseline and  $\text{I1} = 0.1$  and  $\text{I2} = 0.02$  for post-dexamethasone network analysis.

# Global network analysis



Frequency distribution of 124'611 SNPs after QC and LD-pruning (input for SmCCNet) across 22 chromosomes.

# Global network analysis



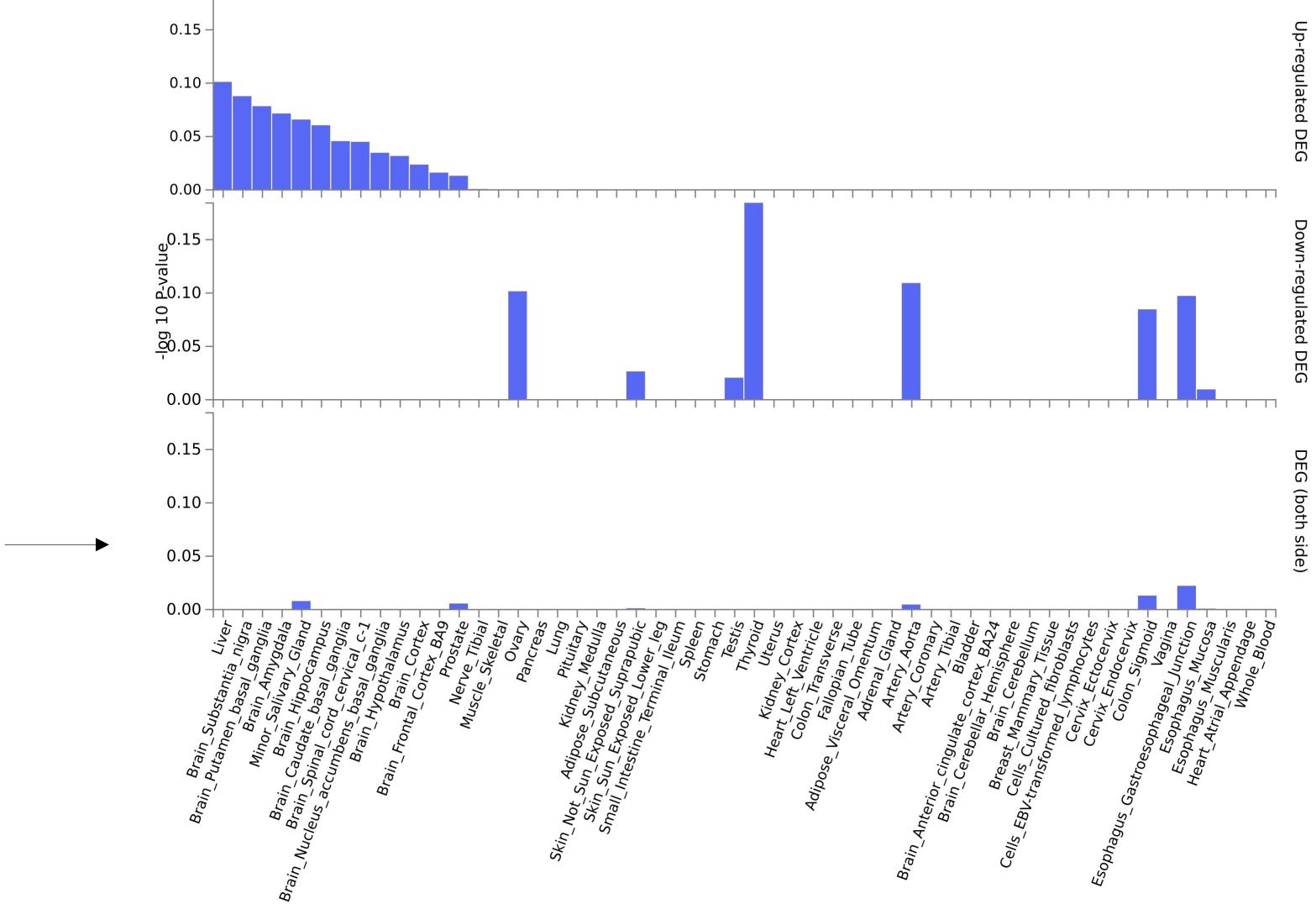
Number of intersection between post-dexamethasone and baseline (a) CpGs and (b) SNPs. (c) Box plot of the number of CpG sites per SNP by treatment. The features selected by SmCCNet and similarity matrix obtained by CCA. The results are shown for the pairs of features with the similarity coefficient  $\geq 0.7$ . On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points.

# Global network analysis. FUMA analysis of post-dex SNPs

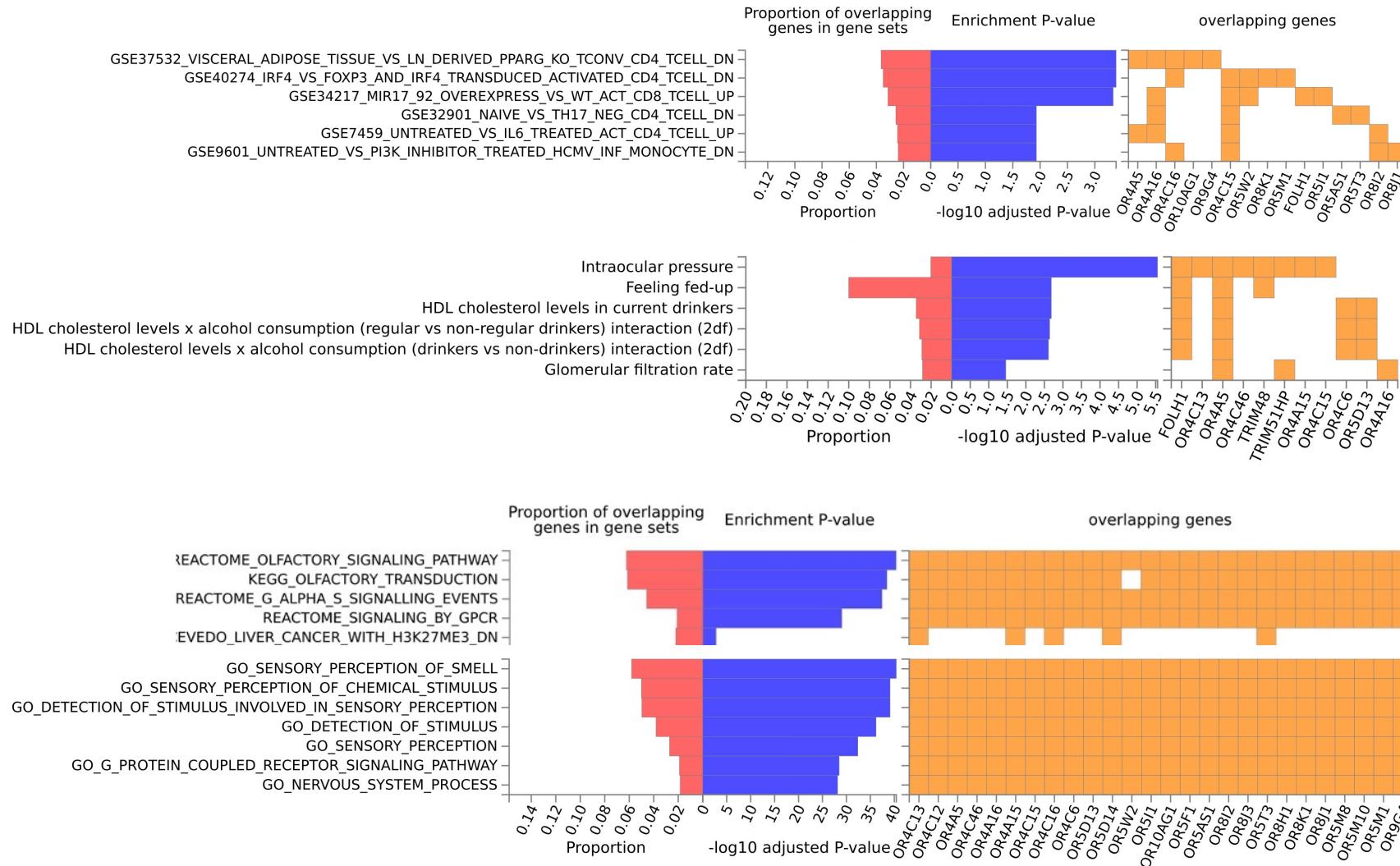
192 post-dexamethasone SNPs were mapped to 37 gene symbols using the R-package ChiPSeeker.

37 gene symbols were annotated in a biological context using the FUMA platform.

Tissue specificity related to the 37 differentially expressed genes (DEG) derived from the 192 post-dexamethasone SNPs selected by SmCCNet. A distinction is made between upregulated DEG (top), downregulated DEG (middle), and bidirectional DEG (bottom). The p values represent the probability from the hypergeometric test.



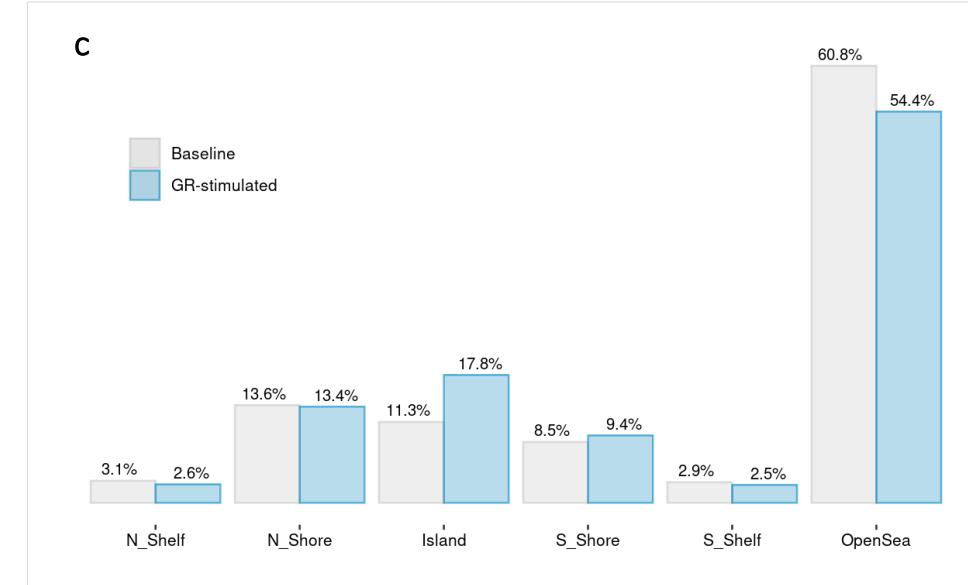
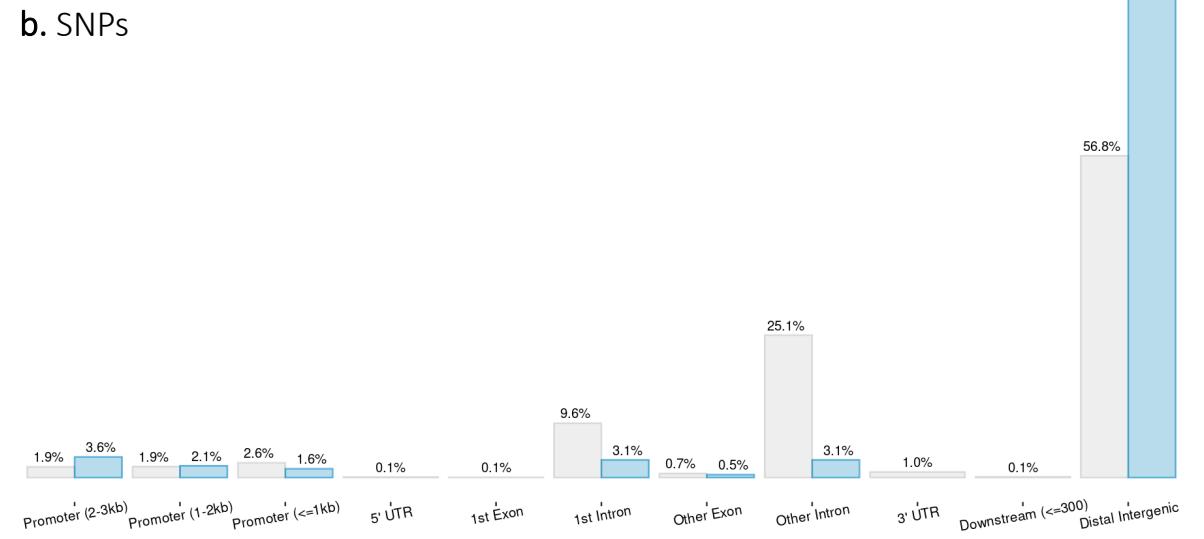
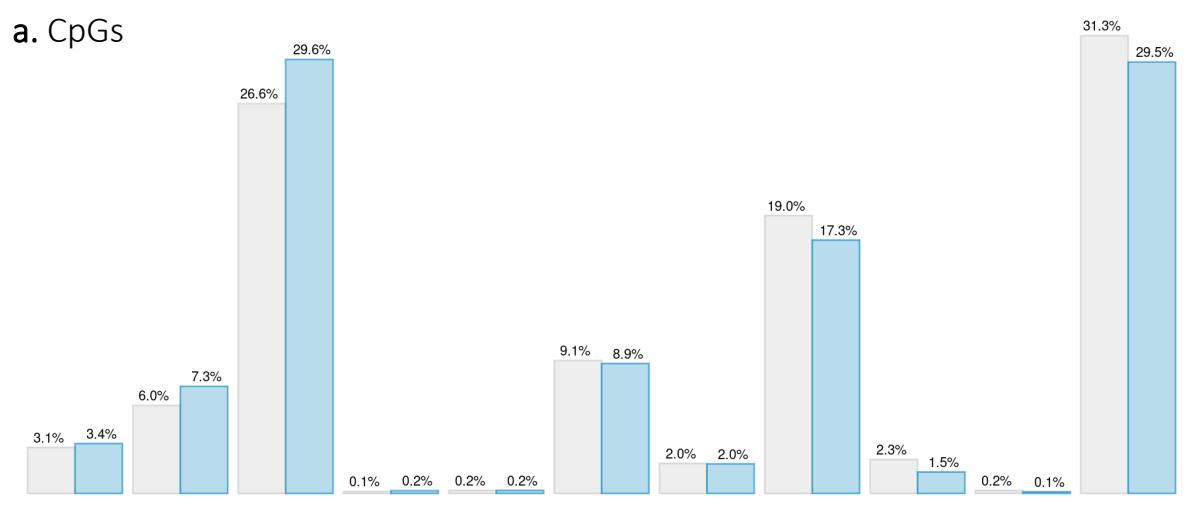
# Global network analysis. FUMA analysis of post-dex SNPs



**Gene Set Enrichment and GWAS enrichment analysis** of 37 genes corresponding to the 192 post-dexamethasone SNPs selected by SmCCNet.

**Pathway analysis and gene ontology analysis** of 37 genes corresponding to the 192 post-dexamethasone SNPs selected by SmCCNet.

# Global network analysis. Functional annotation

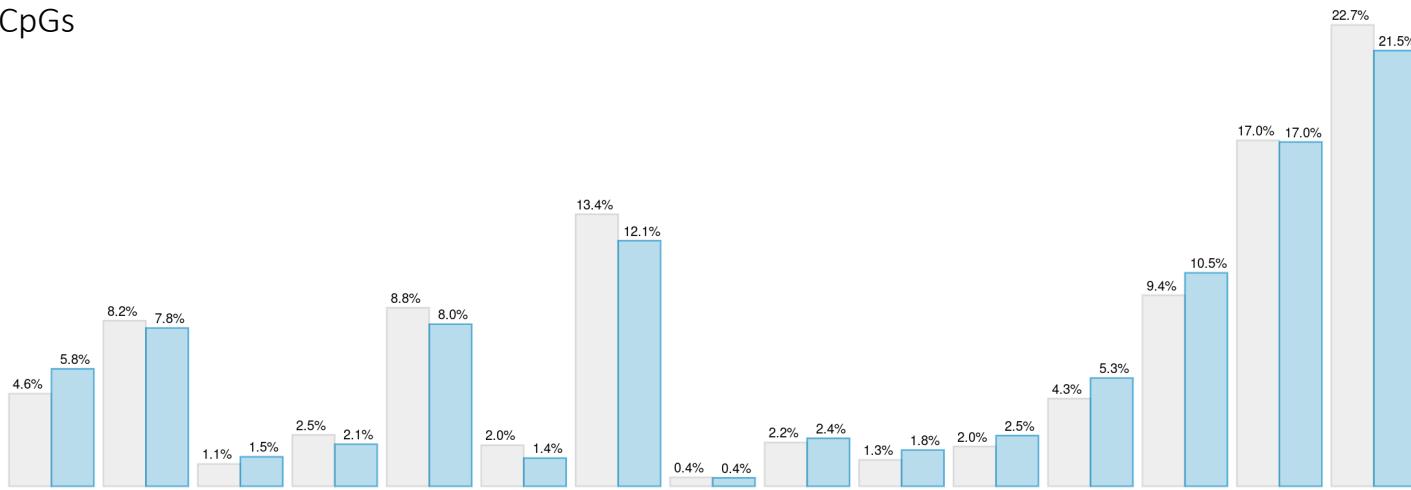


(a, b) Characterisation of CpGs and SNP variants derived from SmCCNet at baseline and after stimulation with dexamethasone in relation to nearby gene regions.

(c) Characterisation of CpGs sites derived from SmCCNet at baseline and after stimulation with dexamethasone in relation to CpG islands.

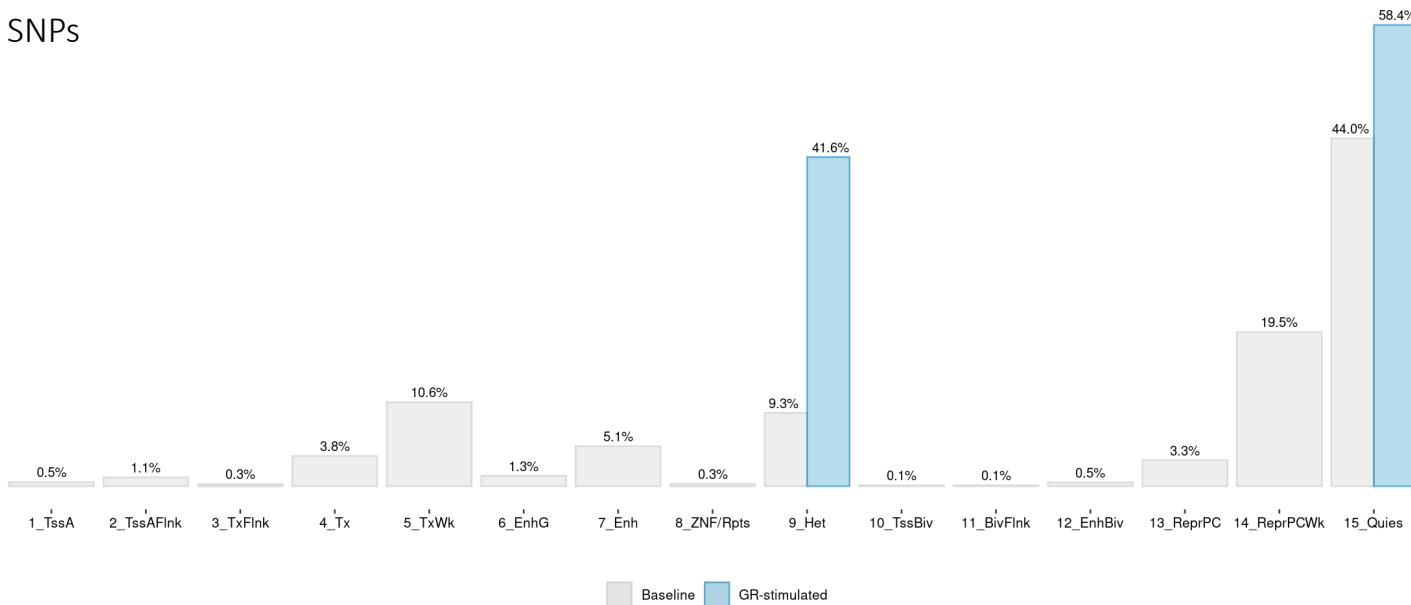
# Global network analysis. ChromHMM annotation

a. CpGs



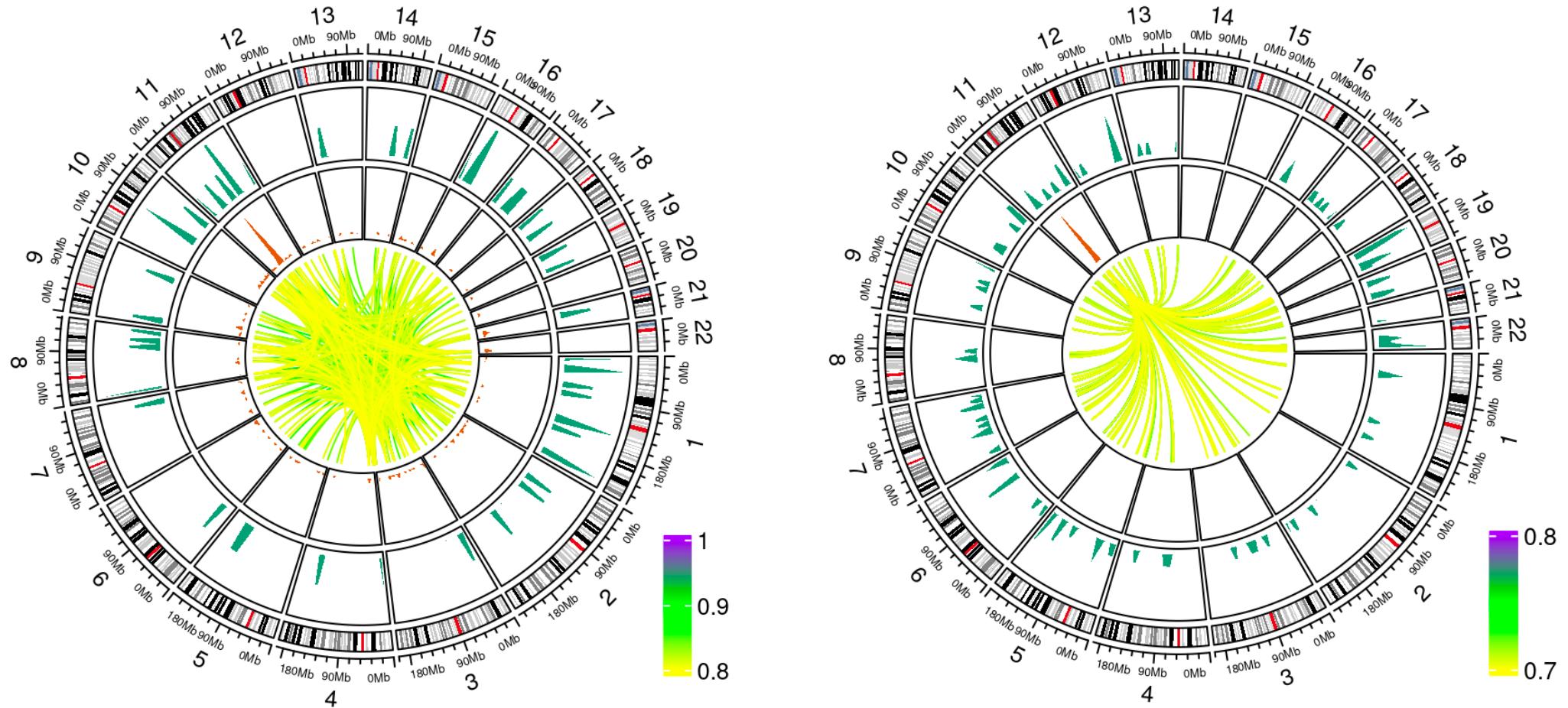
(a) Characterisation of CpGs variants derived from SmCCNet at baseline and after stimulation with dexamethasone in relation to 15 chromatin states from 18 blood and T- & B-cells types.

b. SNPs



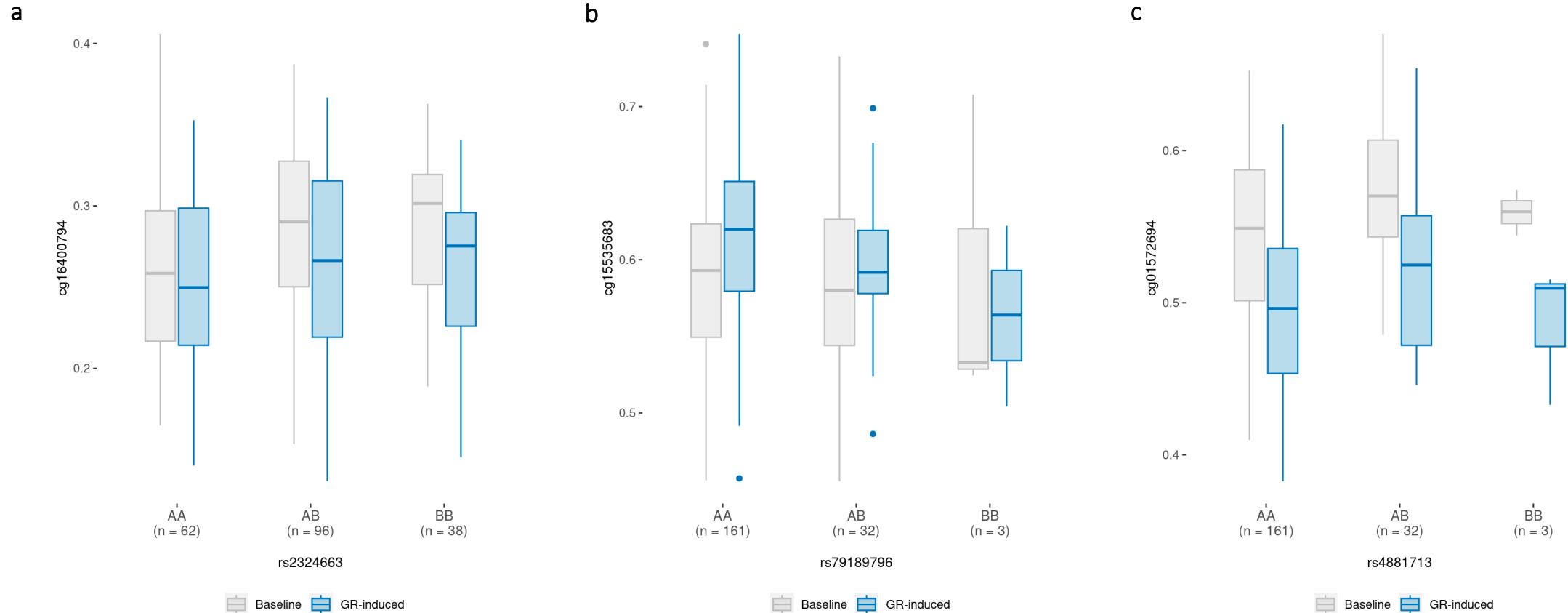
(b) Characterisation of SNPs variants derived from SmCCNet at baseline and after stimulation with dexamethasone in relation to 15 chromatin states from 18 blood and T- & B-cells types.

# Global network analysis



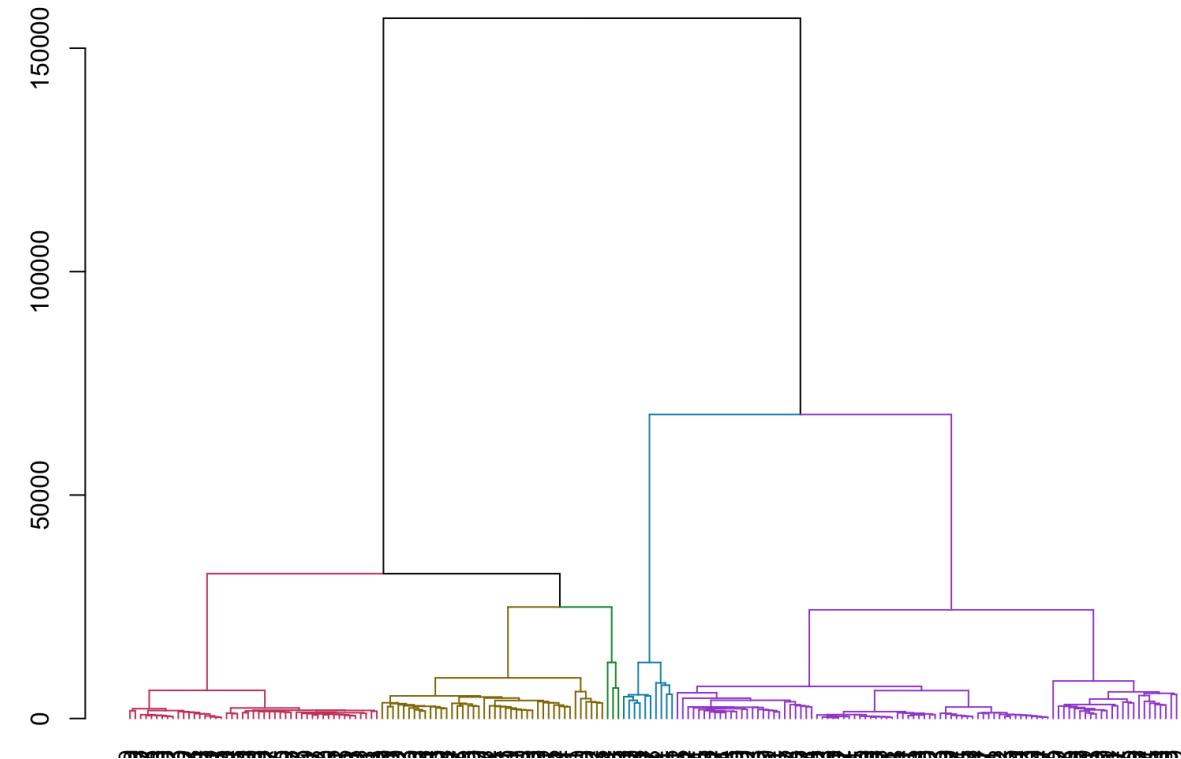
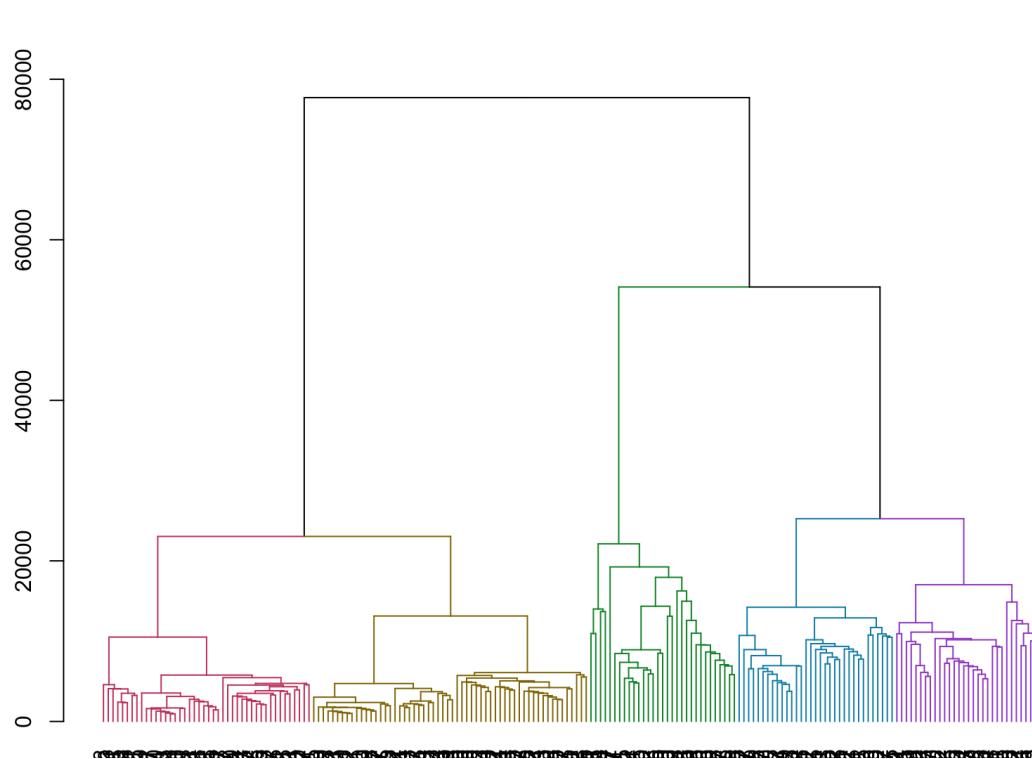
Circos plot of the top 500 meQTLs with the highest similarity coefficient for baseline (left) and post-dexamethasone (right). Tracks from outside to inside are: density of CpGs (green) and SNPs (red). Arrows are pointing from SNPs to the CpG sites they are associated with, and are coloured according to the similarity coefficient level.

# Global network analysis



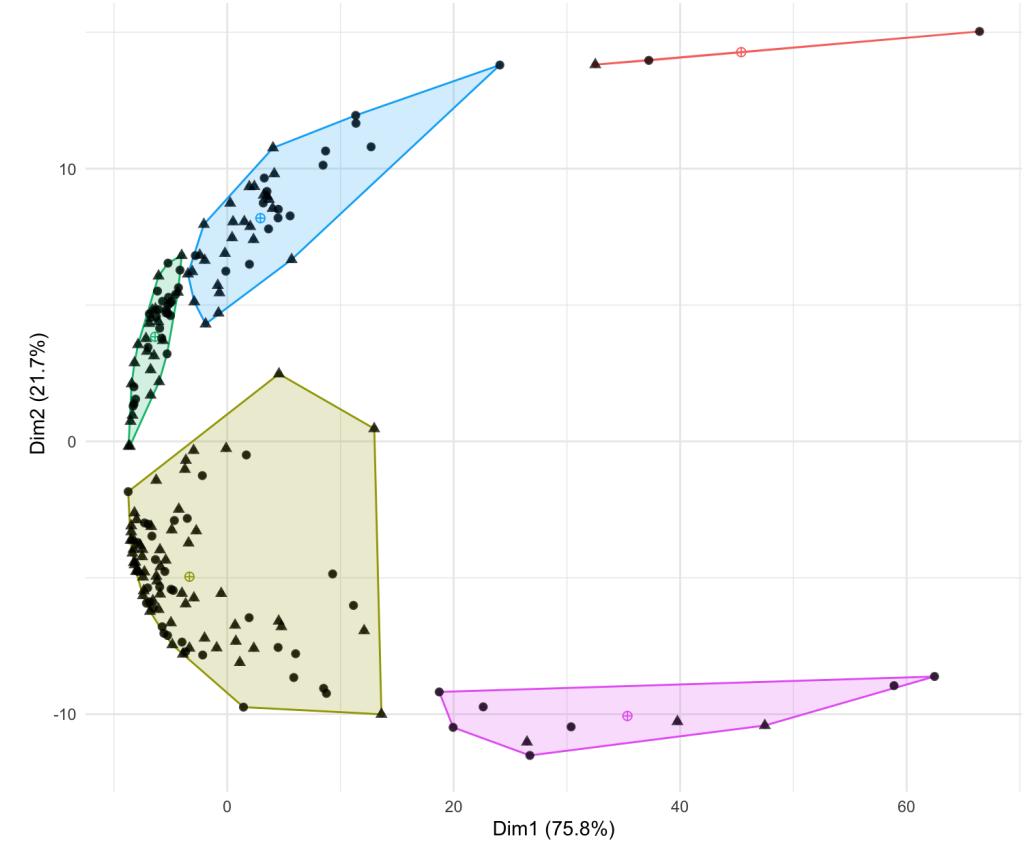
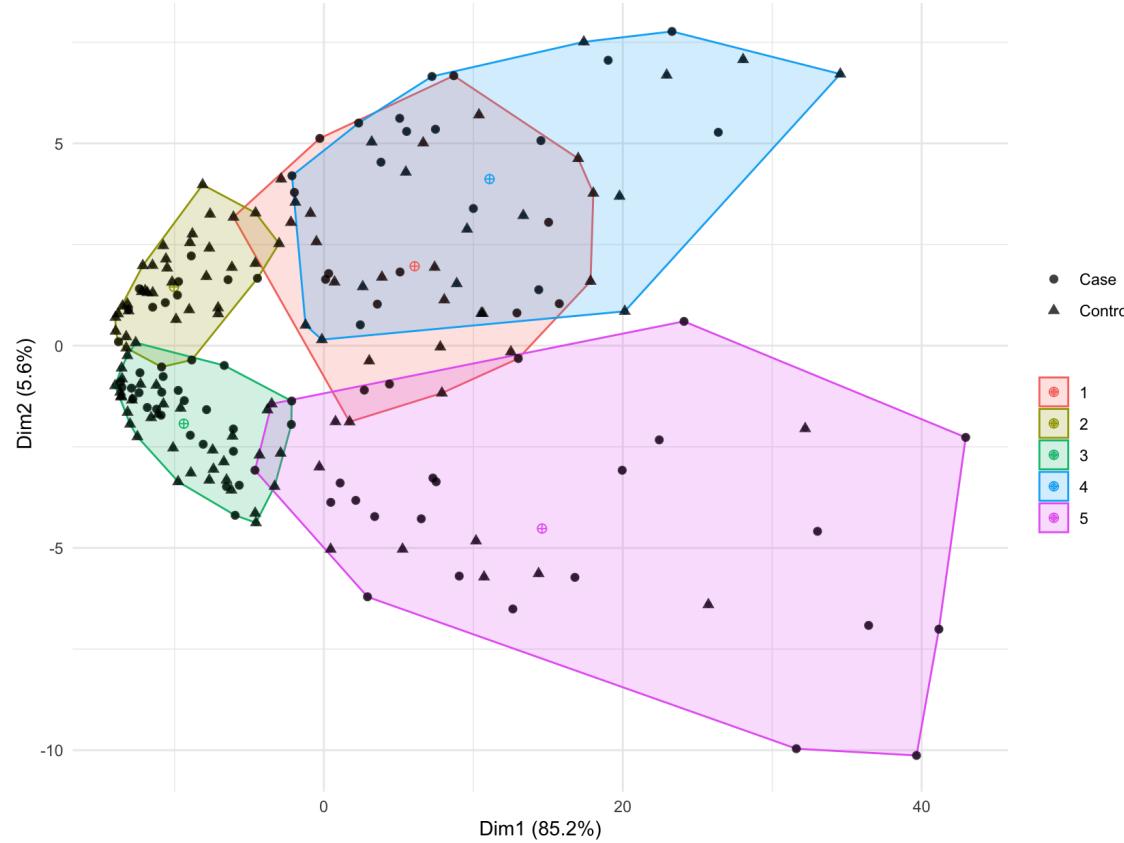
Boxplots of beta methylation values adjusted for covariates which are examples of *trans*-meQTLs. Methylation levels are stratified based on the meQTL SNP genotypes. (a) meQTL with SNP located on chromosome 3 and CpG - on chromosome 1, the effect was observed only at baseline. (b) GR-response meQTL with SNP - on chromosome 11 and CpG - on chromosome 8, the meQTL effect was observed only post-dexamethasone. (c) meQTL with SNP - on chromosome 11 and CpG - on chromosome 7, the meQTL effect was observed in both pre and post-dexamethasone.

# Descriptive analysis of ISNs



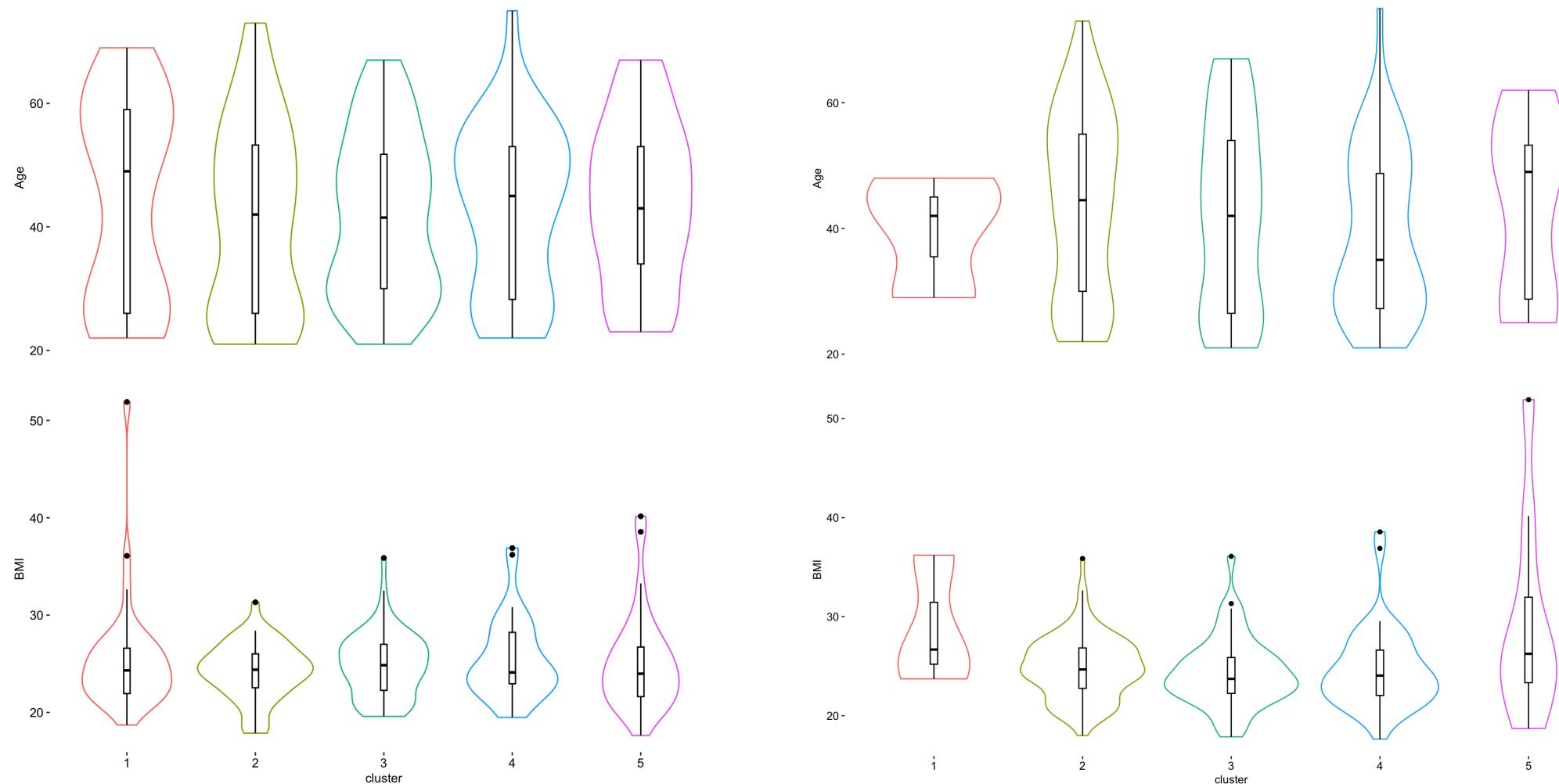
The dendrograms of the results of hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Descriptive analysis of ISNs. Individual cluster plots



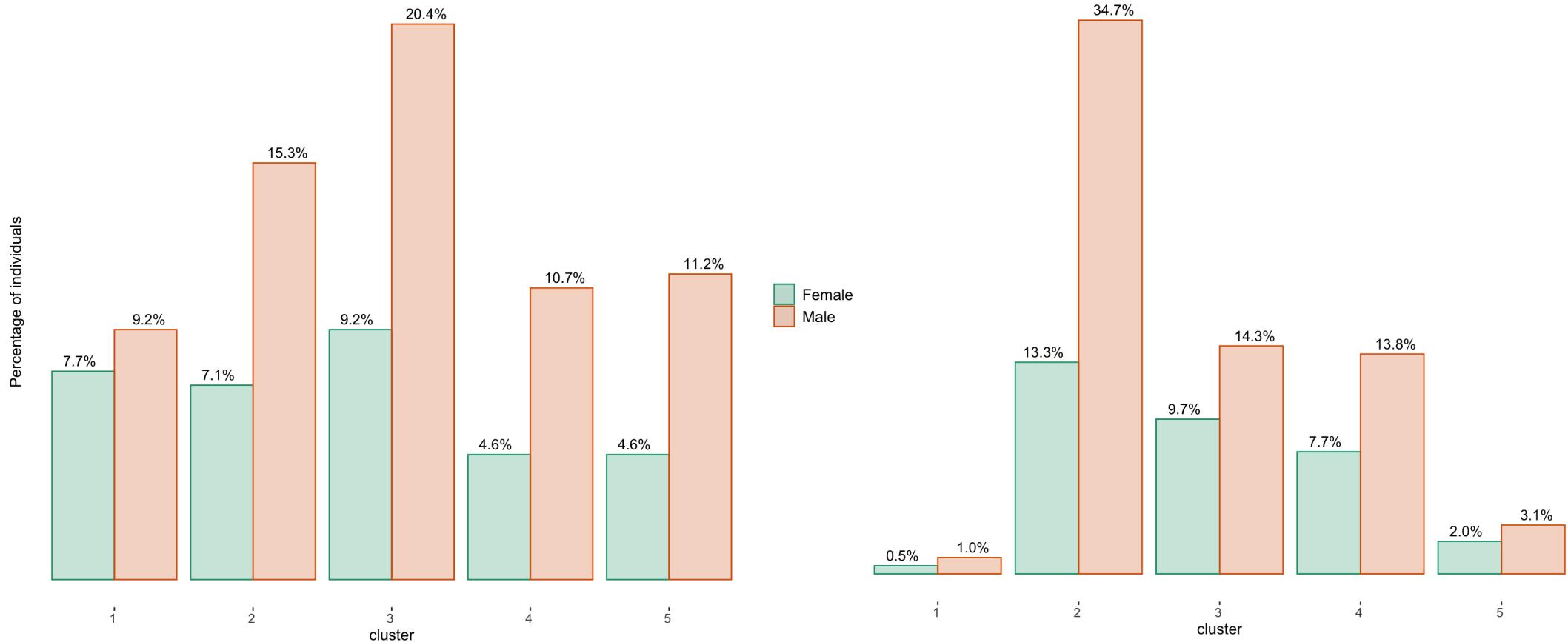
The individual cluster plot of subgroups derived from hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Descriptive analysis of ISNs. Age and BMI distribution



Violon plots of Age (top) and BMI (bottom) in the clusters derived via hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Descriptive analysis of ISNs. Sex distribution



Distribution of males and females in each subgroup derived from hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# MNDA. Global network

## High Variable Variants

### 1. Input for MNDA

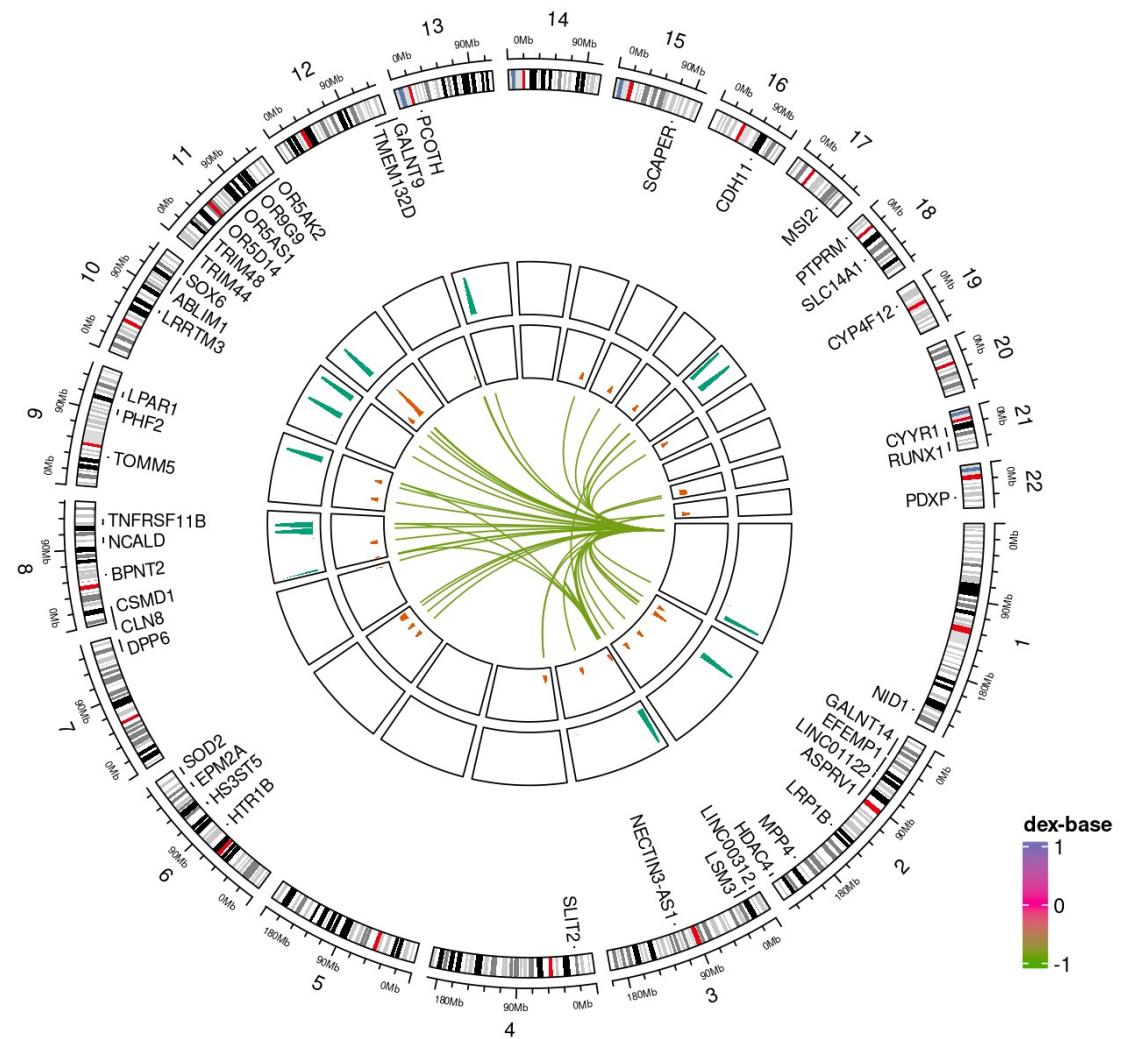
- Disappearing variants, i.e. present only at baseline:  
*baseline weight cut-off is 0.85*
  - Appearing variants, i.e. present only post-dexamethasone:  
*dex weight cut-off is 0.85*
  - Remaining variants, i.e. present in both conditions:  
*baseline and dex weight cut-offs are 0.1*
- ⇒ **2'608 associations**

### 2. Parameters:

- EDNN training repeats: 50
- The dimension of embedding space: 10
- Number of repeats for the random walk: 100
- Number of the random walk steps: 10
- Batch size: 32
- Maximum number of epochs: 50
- BH FDR = 0.05

### 3. MNDA results

- 46 high variable variants (14 CpGs and 32 SNPs), resulting in 53 associations
- Associations with high variable variants are present only at baseline



Circos plot of 46 high variable variants derived from MNDA. Tracks from outside to inside are genes nearby the CpGs and SNPs, the density of high variable CpGs (green) and SNPs (red). Arrows point from the high variable variants to the sites they are associated with and are coloured according to the differences between post-dexamethasone and baseline.

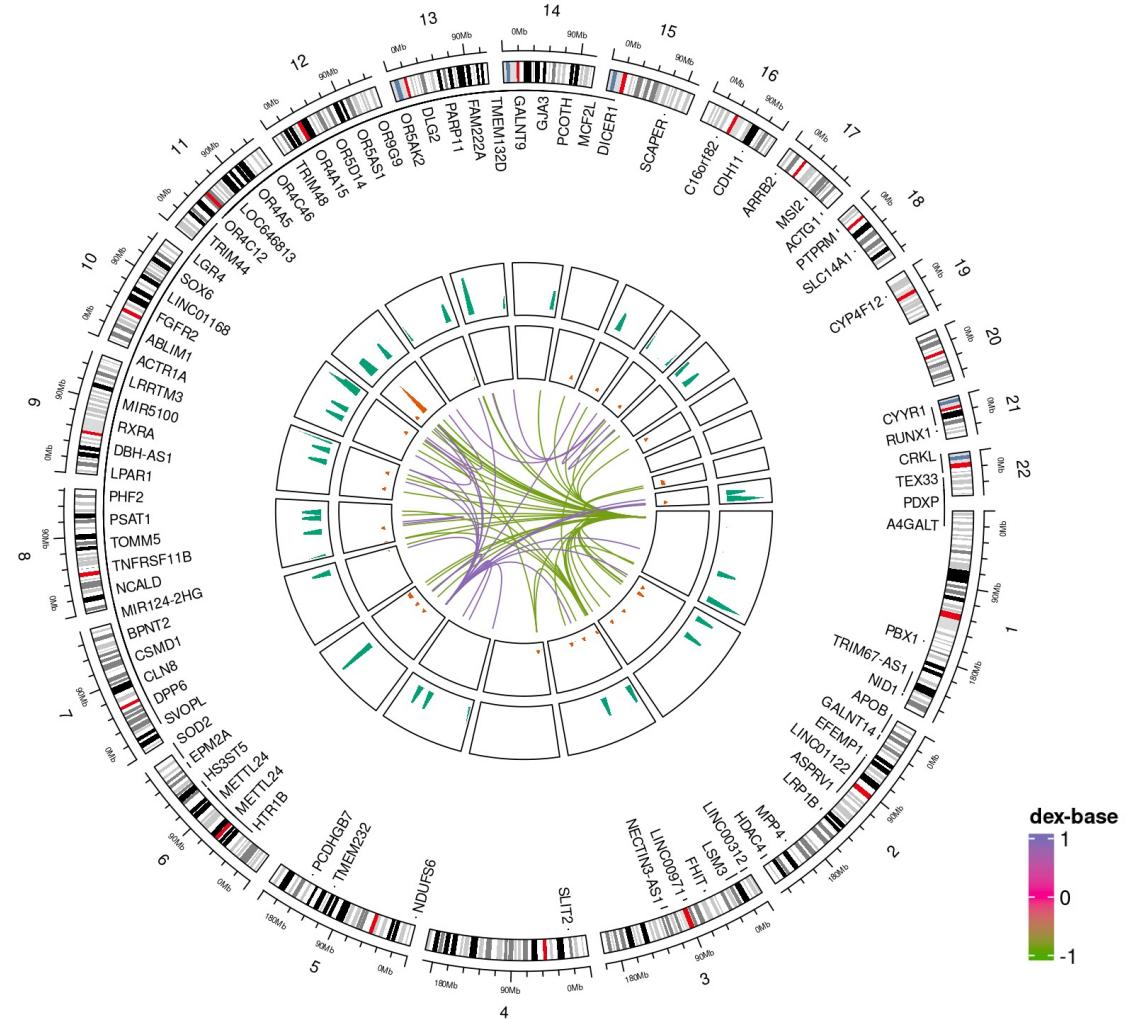
# MNDA. Global network

## Significant Variants

### MNDA results

- 83 significant variants (44 CpGs and 39 SNPs), resulting in 495 associations
- Associations returned at both time points have a weaker connection after stimulation with dexamethasone
- The list of top significant at FDR = 0.05 variants:

ID	CHR	POS	GENE	REGION	FDR
rs188448	chr11	55556411	OR5D14	Distal Intergenic	1.134477e-28
rs2629025	chr15	76943363	SCAPER	Other Intron	1.134477e-28
cg17589341	chr18	43304079	SLC14A1	Promoter (<=1kb)	1.134477e-28
cg09179277	chr10	68835646	LRRTM3	Other Intron	1.134477e-28
rs116777062	chr11	54802952	TRIM48	Distal Intergenic	1.134477e-28
rs7935458	chr11	55804521	OR5AS1	Distal Intergenic	1.134477e-28
rs1033392	chr6	114714143	HS3ST5	Distal Intergenic	1.134477e-28
rs9484950	chr6	145409212	EPM2A	Distal Intergenic	1.134477e-28
rs7817368	chr8	57843906	BPNT2	Distal Intergenic	1.134477e-28
rs10174165	chr2	240441158	HDAC4	Distal Intergenic	1.134477e-28
rs9361214	chr6	77973186	HTR1B	Distal Intergenic	1.134477e-28
rs9933793	chr16	63580572	CDH11	Distal Intergenic	1.134477e-28
rs10083834	chr17	55582918	MSI2	Other Intron	1.134477e-28
rs2834757	chr21	36467070	RUNX1	Other Intron	1.134477e-28
rs6743375	chr2	59519691	LINC01122	Distal Intergenic	1.134477e-28
rs72846578	chr2	142435465	LRP1B	Other Intron	1.134477e-28
rs56116031	chr2	202524879	MPP4	Other Intron	1.134477e-28
rs240873	chr6	159987048	SOD2	Distal Intergenic	1.134477e-28
cg23022828	chr8	4049618	CSMD1	Other Intron	1.134477e-28
cg04940312	chr11	35688285	TRIM44	1st Intron	1.134477e-28
rs55668942	chr2	56162462	EFEMP1	Distal Intergenic	1.134477e-28
rs10758442	chr9	37591236	TOMM5	Promoter (1-2kb)	1.134477e-28
rs7113096	chr11	56474141	OR9G9	1st Intron	1.134477e-28
cg14695497	chr12	132860966	GALNT9	Promoter (2-3kb)	1.134477e-28



Circos plot of 83 significant at FDR = 0.05 variants derived from MNDA. Tracks from outside to inside are genes nearby the CpGs and SNPs, the density of significant CpGs (green) and SNPs (red). Arrows point from the significant variants to the sites they are associated with and are coloured according to the differences between post-dexamethasone and baseline.

# MNDA. ISN example

## High Variable Variants

### 1. Input for MNDA

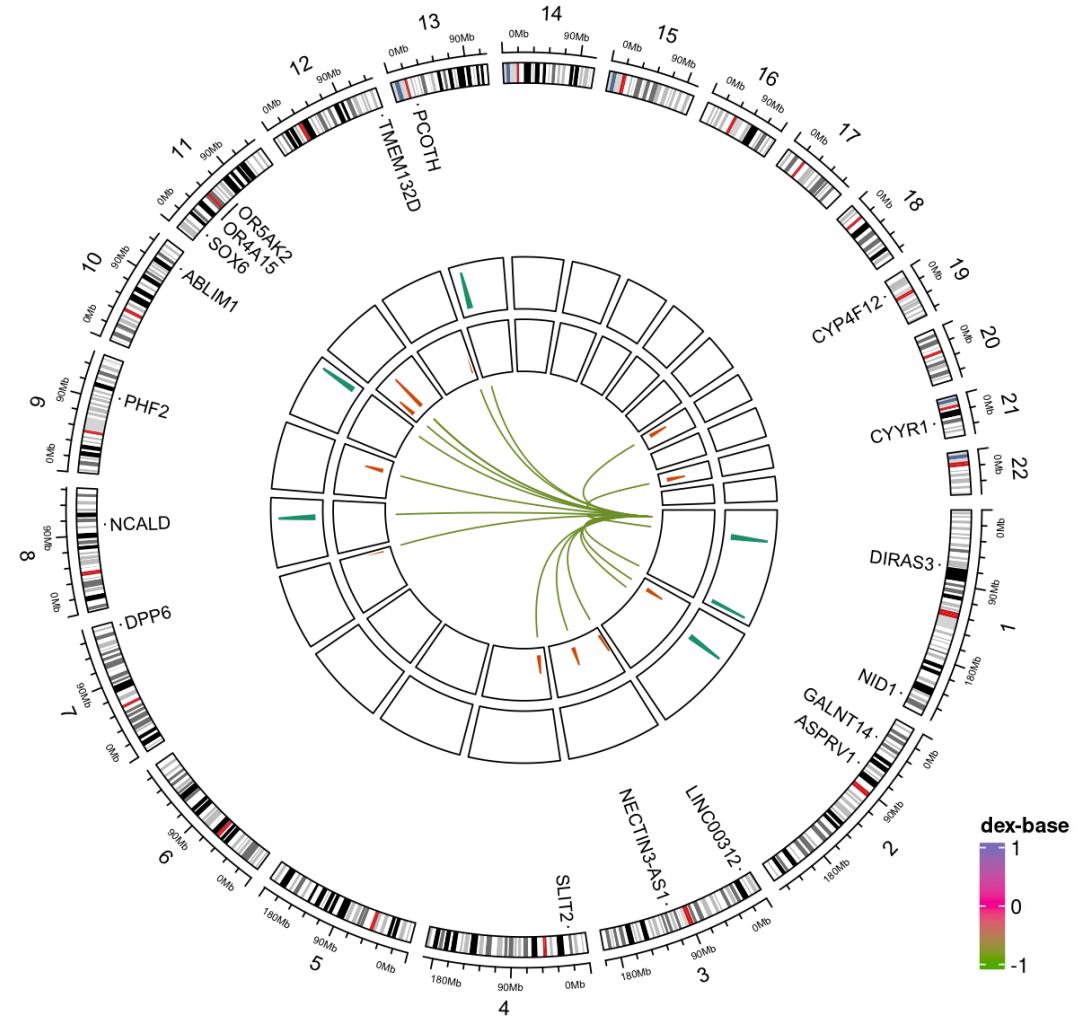
- Disappearing variants, i.e. present only at baseline:  
*baseline weight cut-off is 0.85*
  - Appearing variants, i.e. present only post-dexamethasone:  
*dex weight cut-off is 0.85*
  - Remaining variants, i.e. present in both conditions:  
*baseline and dex weight cut-offs are 0.2*
- ⇒ **1'136 associations**

### 2. Parameters:

- EDNN training repeats: 50
- The dimension of embedding space: 10
- Number of repeats for the random walk: 100
- Number of the random walk steps: 10
- Batch size: 32
- Maximum number of epochs: 50
- BH FDR = 0.05

### 3. MNDA results

- 18 high variable variants (6 CpGs and 12 SNPs), resulting in 18 associations
- 44 significant variants (12 CpGs and 32 SNPs), resulting in 44 associations
- These associations are present only at the baseline
- All genes, except for DIRAS3, associated with these variants for the ISN, are present in global network MNDA analysis



Circos plot of 18 high variable variants derived from MNDA for individual MPIPSYKL\_002794. Tracks from outside to inside are genes nearby the CpGs and SNPs, the density of high variable CpGs (green) and SNPs (red). Arrows point from the high variable variants to the sites they are associated with and are coloured according to the differences between post-dexamethasone and baseline.

# Results. Part II

# Dimensionality reduction

## a) DNAm

The 80% percentile was taken as the cut-off of the Median Absolute Deviation, MAD, score

⇒ CpGs: 196 x 61,440 baseline  
196 x 61,156 post-dex

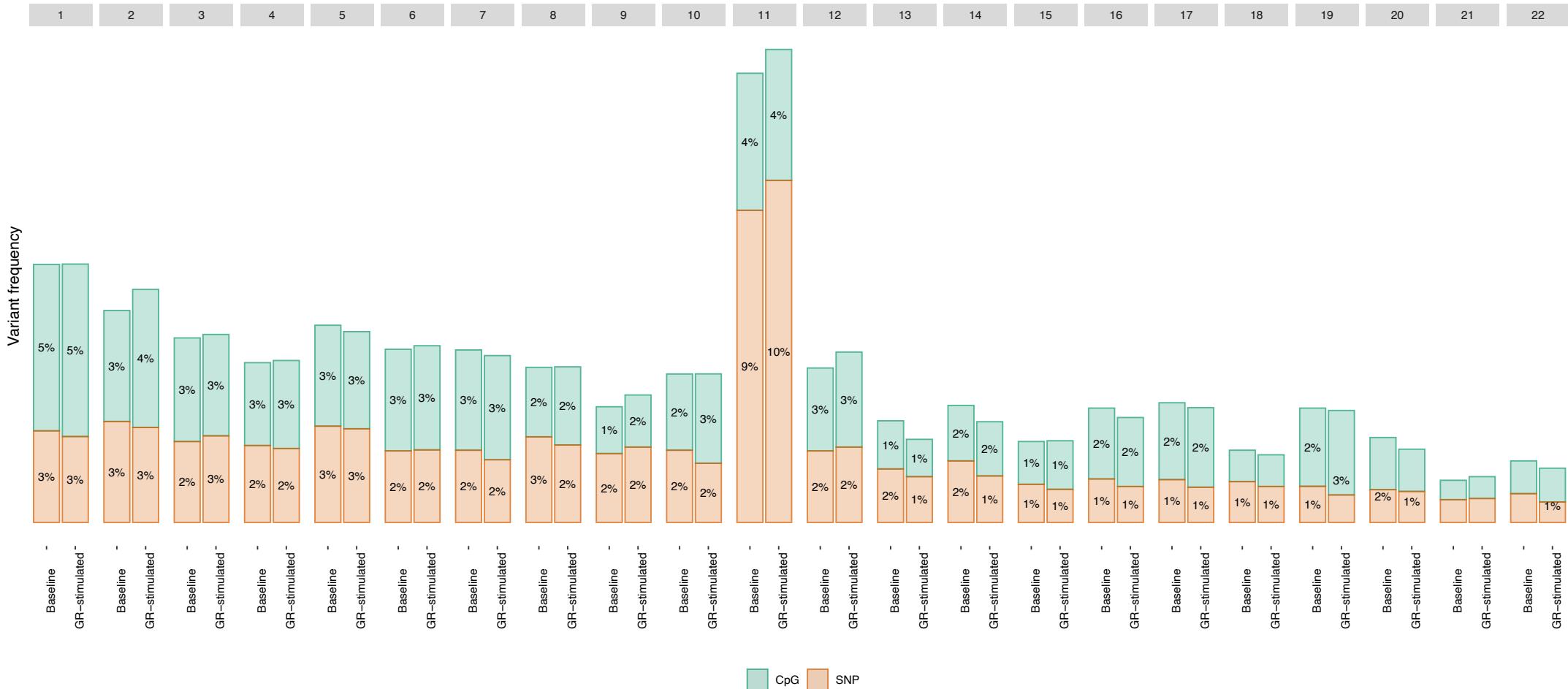
## b) SNPs

LD pruning with the parameters:

- the window size in SNPs = 100
- the number of SNPs to shift the window at each step = 50
- $r^2$  threshold = 0.2

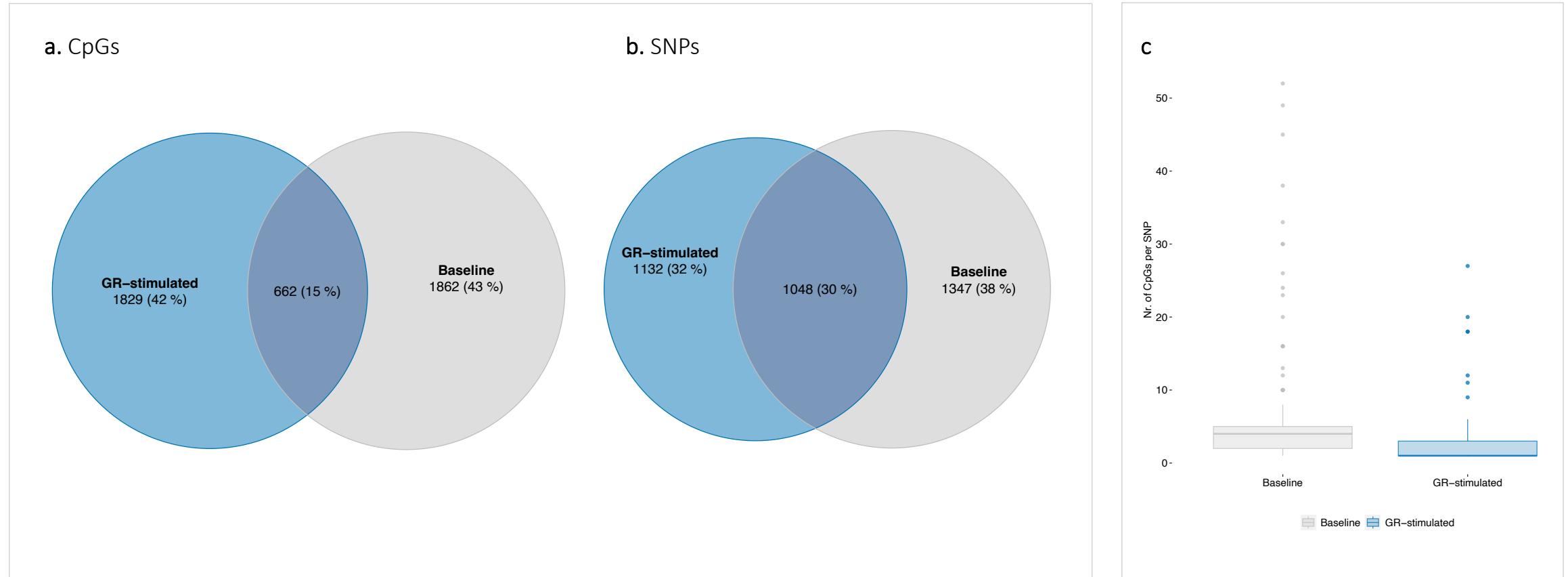
⇒ SNPs: 196 x 124'611

# Global network analysis



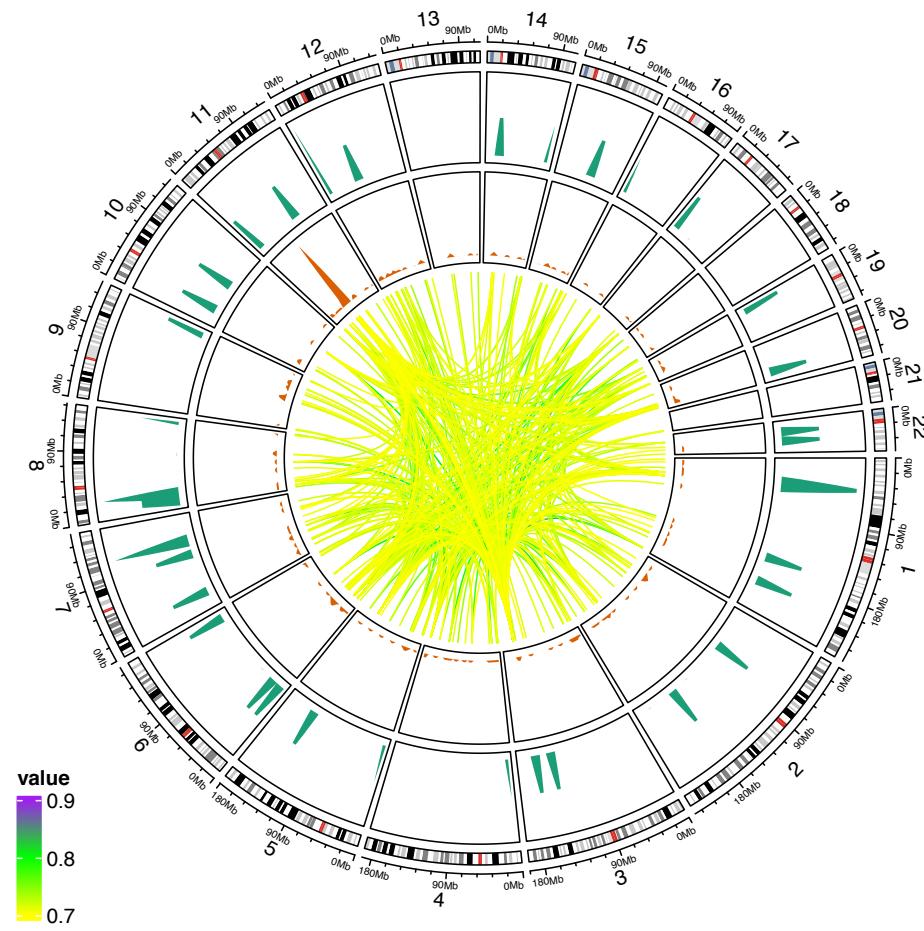
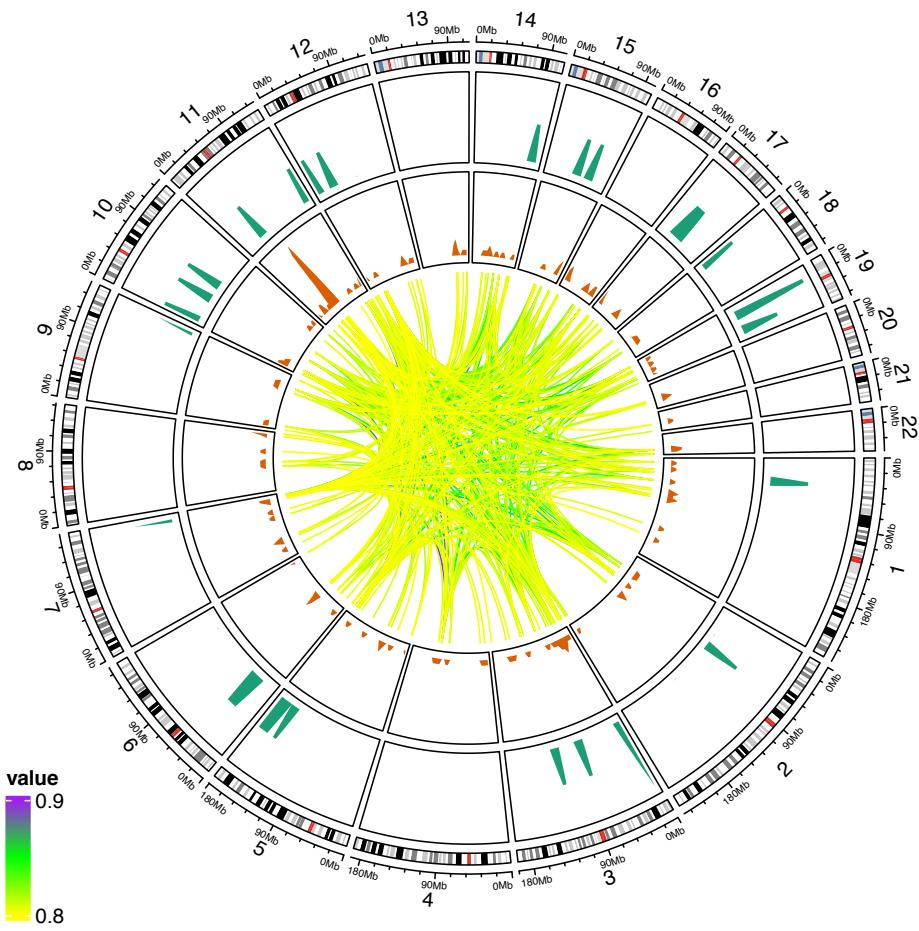
SmCCNet results. Frequency distribution of CpG sites and SNPs selected by SmCCNet across 22 chromosomes. SmCCNet penalty parameters were obtained for baseline using a 5-fold CV ( $\lambda_1 = 0.05$ ,  $\lambda_2 = 0.04$ ) and used for post-dexamethasone network analysis.

# Global network analysis



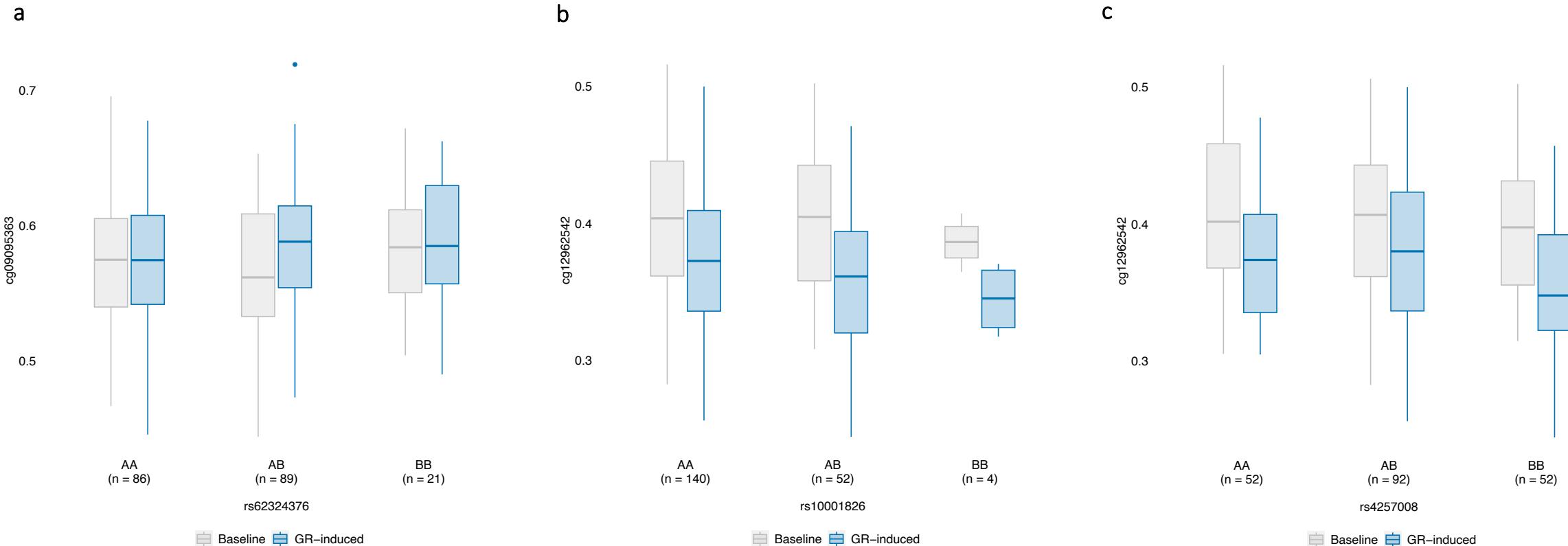
Number of intersection between post-dexamethasone and baseline (a) CpGs and (b) SNPs. (c) Box plot of the number of CpG sites per SNP by treatment. The features selected by SmCCNet and similarity matrix obtained by CCA. The results are shown for the pairs of features with the similarity coefficient  $\geq 0.75$ . On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points.

# Global network analysis



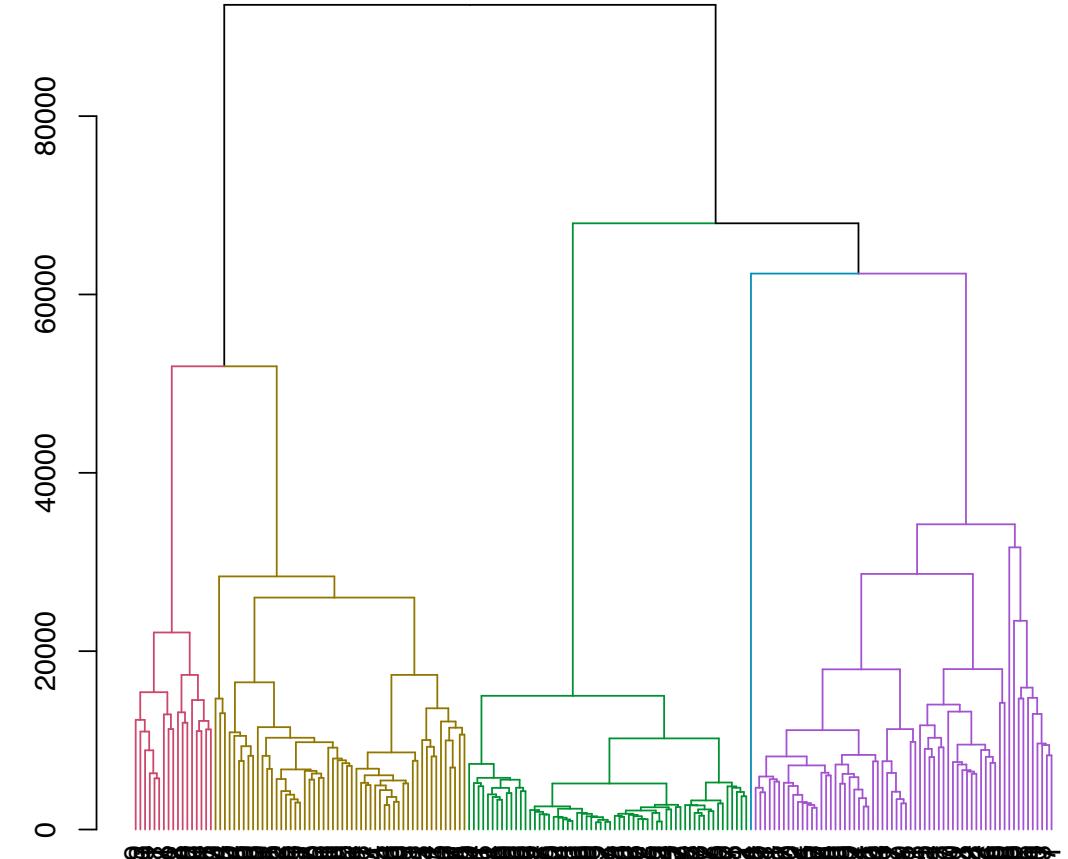
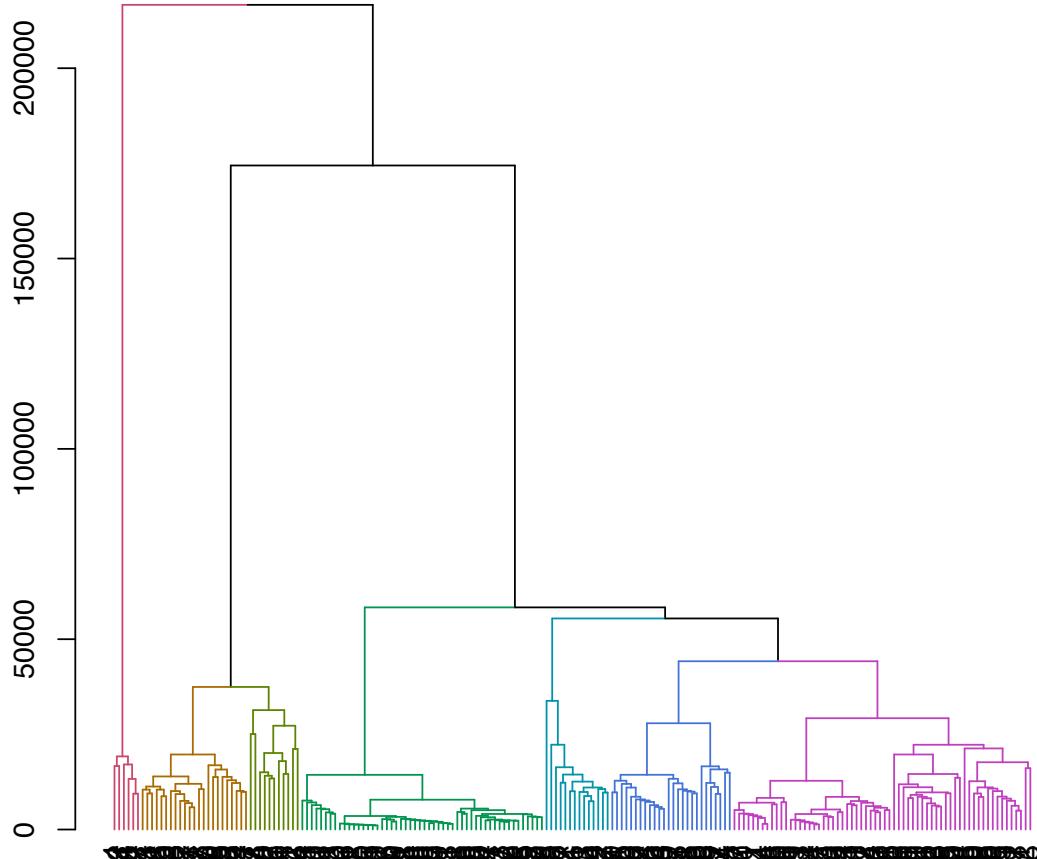
Circos plot of the top 500 meQTLs with the highest similarity coefficient for baseline (left) and post-dexamethasone (right). Tracks from outside to inside are: density of CpGs (green) and SNPs (red). Arrows are pointing from SNPs to the CpG sites they are associated with, and are coloured according to the similarity coefficient level.

# Global network analysis



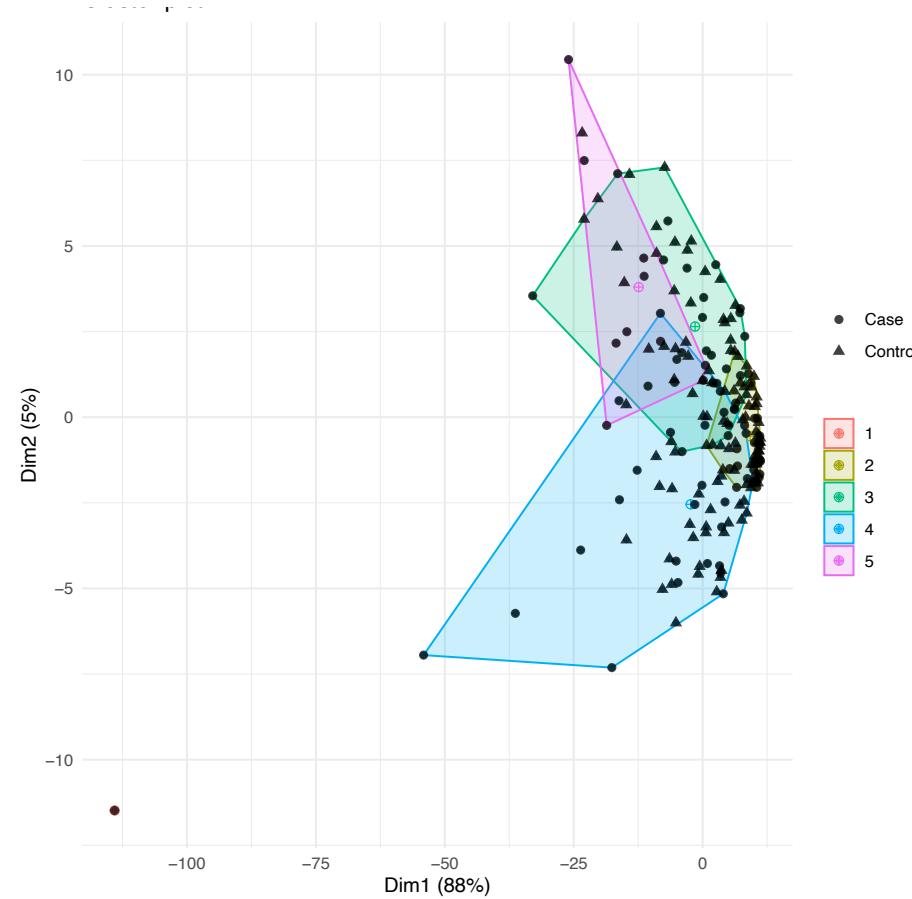
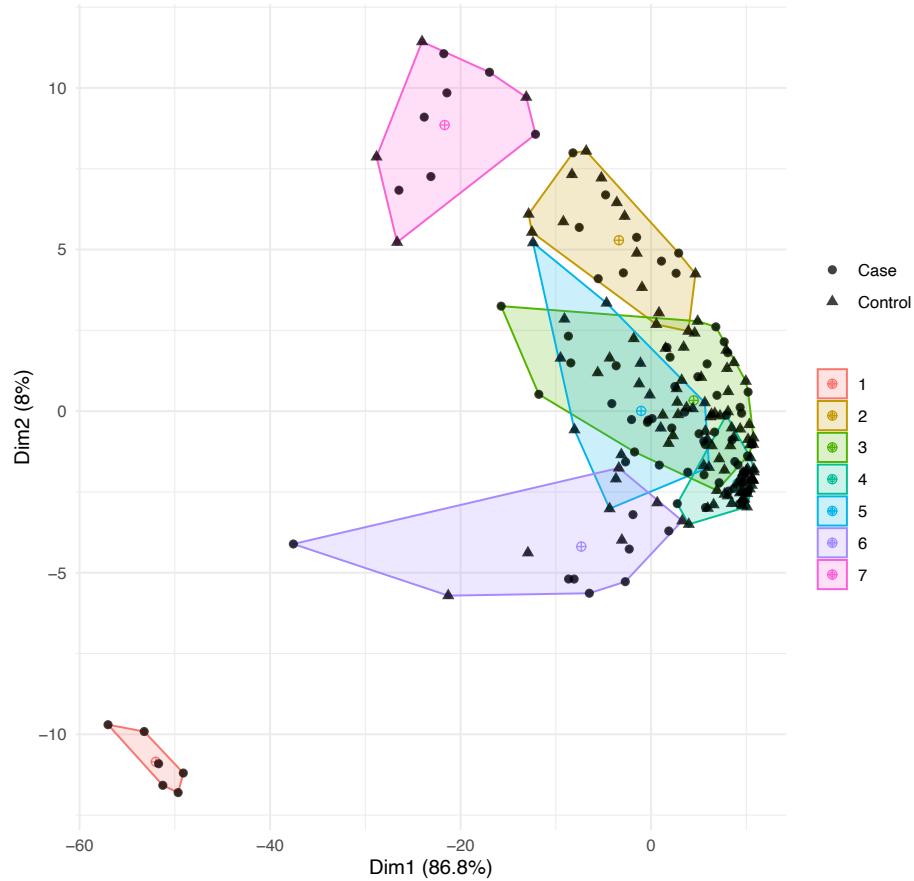
Boxplots of beta methylation values adjusted for covariates which are examples of *trans*-meQTLs. Methylation levels are stratified based on the meQTL SNP genotypes. (a) meQTL with SNP located on chromosome 4 and CpG - on chromosome 7, the effect was observed only at baseline. (b) GR-response meQTL with SNP - on chromosome 12 and CpG - on chromosome 3, the meQTL effect was observed only post-dexamethasone. (c) meQTL with SNP - on chromosome 4 and CpG - on chromosome 3, the meQTL effect was observed in both pre and post-dexamethasone.

# Descriptive analysis of ISNs



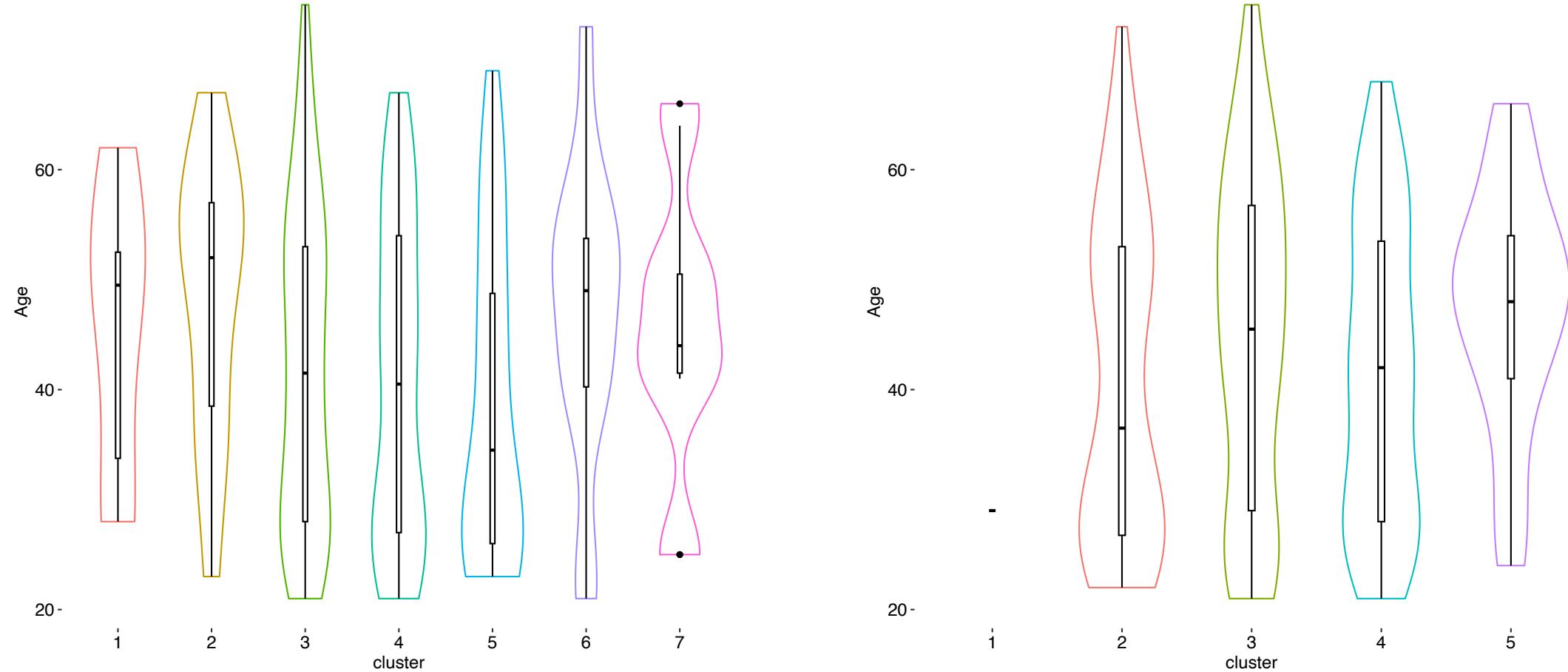
The dendograms of the results of hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Descriptive analysis of ISNs. Individual cluster plots



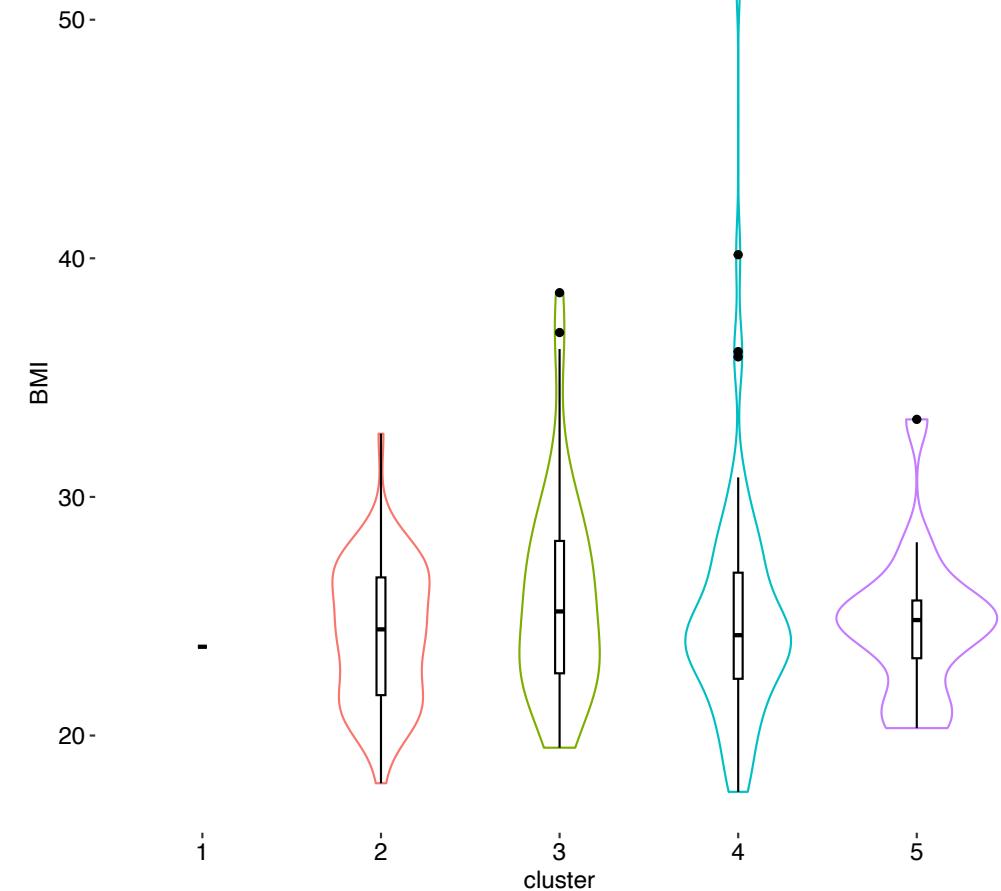
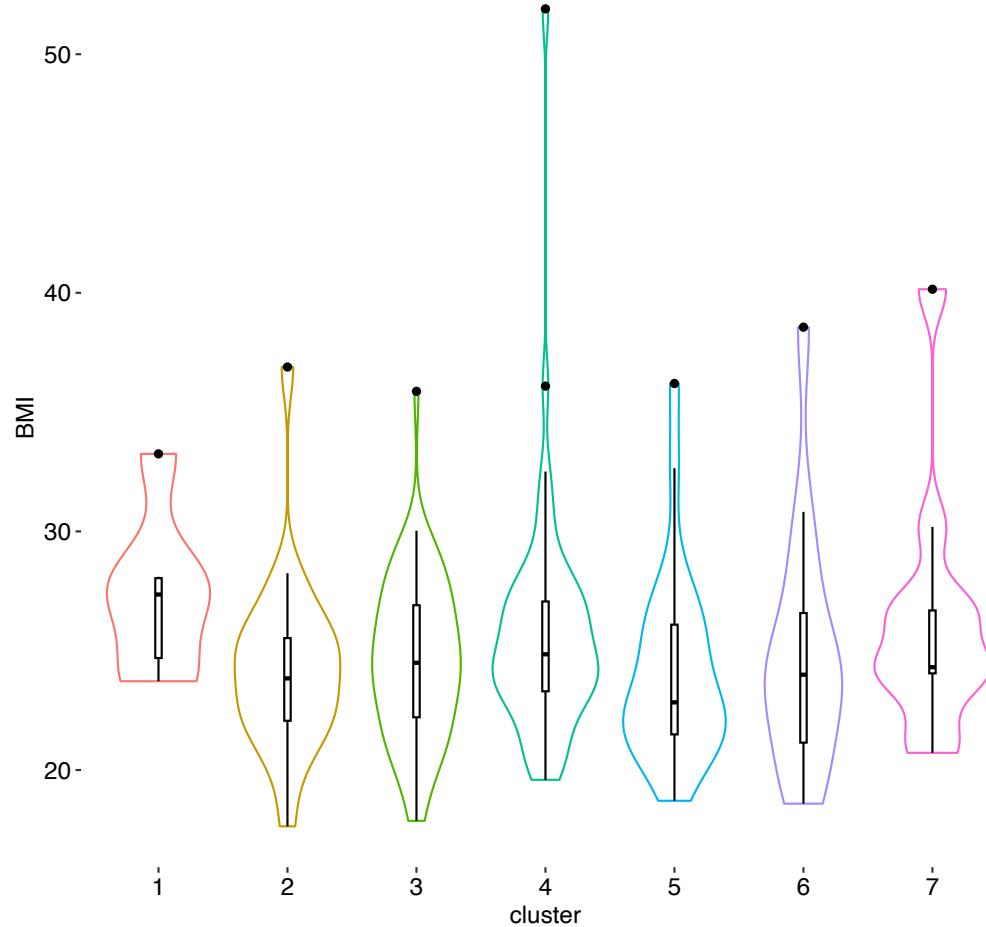
The individual cluster plot of subgroups derived from hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Descriptive analysis of ISNs. Age distribution



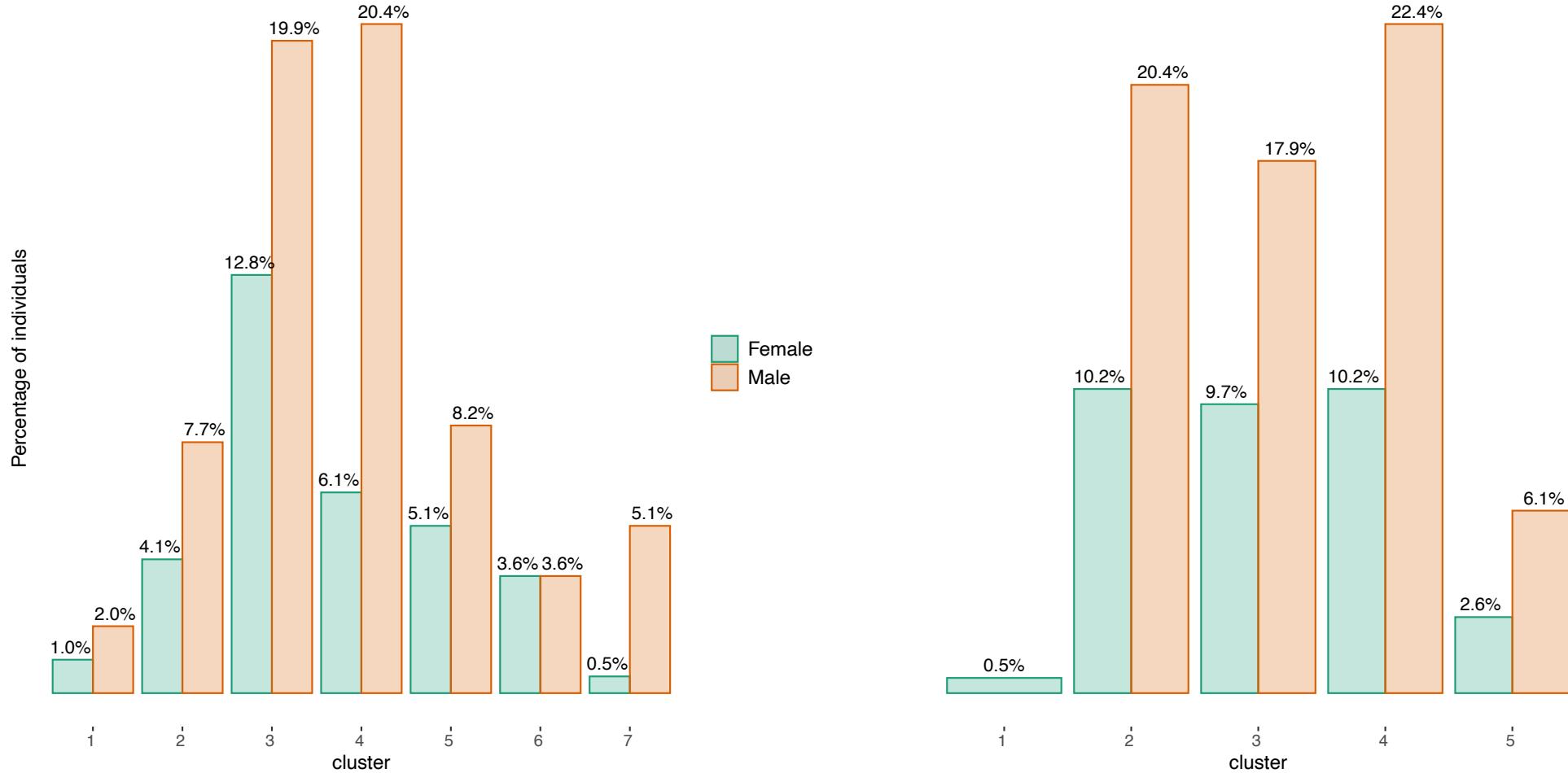
Violon plots of Age in the clusters derived via hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Descriptive analysis of ISNs. BMI distribution



Violon plots of BMI in the clusters derived via hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Descriptive analysis of ISNs. Sex distribution



Distribution of males and females in each subgroup derived from hierarchical clustering on the Euclidean distances between ISNs for baseline (left) and post-dexamethasone (right).

# Summary

1. The analysis should be re-run taking the more extensive set of variants or using another approach for feature selection. The set of SNPs derived from SmCCNet doesn't look relevant to extract plausible information regarding gene-environment interactions in relation to psychiatric traits.
2. Not only *cis*- but also *trans*-interactions should be considered.
3. There are external post-dexamethasone drivers which affect the profiles of individuals and their population structure.
4. Network-based approaches are a powerful tool to extract valuable biological information (for population stratification and biological drivers identification)
5. ISNs are a promising approach for population stratification, which may further contribute to precision medicine optimisation.

## Outlook

1. Repeat analysis for a larger set of features (is a must): less stringent thresholds for MAD filtration and LD pruning
2. Analysis by chromosome (time-consuming, trans effect)
3. Another approach for feature selection and comparison with SmCCNet results
4. Integrate gene expression as an additional layer

To discuss...

1. SNPs on chr 11
2. Analysis by chromosome
3. GPU Computing
4. Selection of penalties parameters in SmCCNet for feature selection
5. Unsupervised vs supervised approaches for feature selection

# Acknowledgment



MAX PLANCK INSTITUTE  
OF PSYCHIATRY



Prof. Dr. Dr. Kristel Van Steen  
Federico Melograna  
Zuqi Li  
Behnam Yousefi  
@ TranSYS Consortium

Dr. Janine Knauer-Arloth  
Prof. Dr. Dr. Elisabeth Binder



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 860895.