SUPPLEMENTARY MATERIALS

Supplemental Methods

Bioinformatics

All analyses were performed in R (v3.5.1) and RStudio (v1.1.463). The STAGE GRN for CAD and associated modules were extracted from a previous publication[5]. Atherosclerotic arterial wall – specific modules were defined as those modules that had at least 80% contribution of genes from atherosclerotic arterial wall tissue.

Differential gene expression analysis

We used all the female samples of which atherosclerotic tissue RNA-sequencing data was available, leading to 160 females. Differential gene expression analysis between atherosclerotic tissue of 160 females and 160 age-matched males was performed by using DESeq2 (v1.22.1)[35], while correcting for read length of the sequencing batch. A gene was called differentially expressed if FDR < 0.05.

Gene ontology enrichment analysis

Gene ontology of biological processes was performed and visualized by using clusterProfiler (v3.10.0)[36]. Enrichment for ToppCell Atlas was performed in the ToppSuite[37] of which the output can be found in Supplemental Excel File I.

Connectivity analysis

Gene connectivities were calculated by using the WGCNA package (v1.66)[38]. The softConnectivity function was used with a power of 5 to calculate connectivities of 4,889 genes over a spectrum of populations that differed in numbers of females and males. We started by calculating the connectivities in 160 females. Subsequently, we removed one female sample and added one male sample and recalculated the connectivity. This was repeated until the population existed of 160 males. Connectivity patterns were plotted using the pheatmap package.

Construction of co-expression network

RNA-sequencing count data was filtered by keeping only genes that had an average count of at least three. Next, count data was normalized for depth of sequencing by transforming them to RPM-values. The sva package[39] was used to remove batch effects in the data driven by different read lengths. Genes for network construction were selected by a variance cut-off, assuming that higher varying genes contain more information for the network. Genes were selected in a sex-stratified way: genes that had higher variance than 0.1 in a log-transformed sva-corrected RPM matrix of either the female or the male samples. To prevent genes missing that barely made the cut-off, a union of genes was taken with the genes that had a variance higher than 0.09 in the opposite sex, i.e. the female network contained a union of female-varying genes, and male-varying genes that had a variance higher than 0.09 in female samples, and vice versa for males.

Unsigned networks were constructed separately for both sexes using the WGCNA package (v1.66)[38]. A soft thresholding analysis was performed to determine the exponent used for adjacency matrix construction (power = 5). The adjacency matrix was transformed into a topological overlay matrix, which takes the connectivity of neighbouring genes into account. Subsequently, modules were generated by clustering the average distance of the dissimilarity matrix, 1 – topological overlay matrix, and cutting the resulting dendrogram with cutreeDynamic (deepSplit = 2, minClusterSize = 20).

Module differential connectivity analysis

We calculated intramodular connectivities (connectivity of nodes to other nodes within the same module), as implemented in the WGCNA package[38] with the intramodularConnectivity function, on the same adjacency matrices constructed for the network, as well as of adjacency matrices based on (healthy) mammary artery of the same samples, and of adjacency matrices based on female genes, but male samples. Individual intramodular connectivity of genes were summed to obtain the module connectivity. Module differential connectivity (MDC) of diseased and healthy

tissue was then defined as the quotient of module connectivity in the diseased and the healthy state, while the MDC^{sex} was defined as the quotient of the female and the male module connectivities.

Clinical phenotype permutation tests

To determine association to clinical parameters of CAD, module gene expression was correlated to levels of CRP, cholesterol, HDL, LDL, triglycerides, glucose, and Syntax score. The association was summarized for each module by taking the median of the absolute correlation coefficient between the expression of a gene of that module and one of the parameters. Permutations (10000x) were performed by sampling in R to determine whether the observed value was stronger than for random genes. The permutation p-value was obtained by taking the amount of permutations with a higher median absolute correlation coefficient than the observed value and dividing by the number of permutations. P-values of the permutations can be found in the Supplemental Excel File I.

Permutations for CAD gene content

To determine association of modules to previously described CAD genes, we calculated overlap of the modules with different gene lists (see Supplemental Excel File I). We used a list of CAD GWAS genes[40], a union of different lists of genes known to be involved in atherosclerosis (obtained from DisGenet and Ingenuity Pathway Analysis (Qiagen, Germany), see Supplemental Excel File I), and a gene list that contains CAD susceptibility genes[41]. Permutations (10000x) were performed by sampling in R to determine whether the observed value of overlap was stronger than for random genes. P-values of the permutations can be found in the Supplemental Excel File I.

Module prioritization

First, we determined whether or not the module is important for the disease, by calculating module differential connectivities (MDC) of diseased tissue and healthy vascular tissue. For this, atherosclerosis-free mammary artery of the same patients

was used (Fig. 3A). We deemed modules important for disease if the $MDC^{disease}$ was at least 1.5 times as high or as low. For example, the $MDC^{disease}$ of the greenyellow module in females was 4.72, indicating that the module is more highly connected in diseased tissue than in healthy tissue (Fig. 3B). This step removed 5 modules (Fig. 3G). Individual gene connectivities of the different modules corroborated the module connectivities (Suppl. Fig VA). To determine whether or not the module is affected by sex, we calculated the MDC of diseased tissue in female and male samples. A similar cut-off of 1.5 times as high or as low was used for this filter step. Individual gene connectivities of the different modules once again corroborated the module connectivities (Suppl. Fig VB), with less dispersion than the disease-health comparison. Nine of the 27 remaining modules had an $MDC^{sex} > 1.5$ or < 0.66 (Fig. 3C). To further delineate importance for CAD phenotypes, we calculated correlation coefficients between clinical parameters (CRP, cholesterol, HDL, LDL, triglycerides, glucose and Syntax score) and module gene expression (Suppl. Fig. VI-XII). By permuting correlation coefficients for an equal amount of random genes, we determined which modules were more correlated to clinical CAD parameters (Fig. 3CDE). Six out of the nine remaining modules were associated more to clinical CAD parameters than random modules, in which we checked the gene content. Of these six, three modules were enriched for genes known to be involved in atherosclerosis or enriched for CAD GWAS or susceptibility genes (Fig. 3F, Suppl. Fig XIII-XV, and Suppl. Data File).

Hallmark gene set enrichment analysis

Gene-sets from the MSigDB Hallmark were used to calculate gene enrichments[11], with a gene population size of 20,000. The fmsb package was used for generating radar plots.

Motif enrichment

Motif enrichment analysis was performed using RcisTarget (v1.2.1)[42]. In short, RcisTarget identifies transcription factor binding motifs that are enriched in lists of genes. The motif rankings version used was human hg38 refseq-r80 for 500bp up and

100bp down the transcription start site. The motif-to-TF annotation version that was used was human v9.

Bayesian network inference

Bayesian network inference was performed by using the rcausal package (v0.99.0). The Fast Greedy Equivalence Search (max degrees = 100) for continuous data algorithm was used to generate the Bayesian network on the WGCNA-generated modules.

Key driver analysis

Key driver analysis (KDA) was performed as integrated in the Mergeomics package (v1.10.0), number of permutations was set to 20. Key drivers were considered significant if FDR < 0.05.

Human single-cell RNA-sequencing analysis

Plaque samples of 37 patients undergoing carotid endarterectomy (11 females and 26 males) minus the necrotic core were processed immediately after surgery, washed in RPMI and minced prior to digestion. The cell suspension was filtered through a 70 µm cell strainer and washed with RPMI 1640. Cells were kept in this medium with 1% Fetal Calf serum until staining for fluorescence-activated cell sorting. Single cell suspensions were stained with Calcein AM and Hoechst (ThermoFisher Scientific) in PBS supplemented with 5% Fetal Bovine Serum and 0.2% EDTA and subsequently strained through 70 µm cell strainers. Only cells positive for both Calcein AM and Hoechst (= living cells) were sorted for sequencing. Cells were prepped using the SORT-seq protocol[43] and libraries were constructed using the CEL-seq2 protocol[44]. Data was processed in an R 3.5 environment using the Seurat package[45]. Downstream processing was performed using custom R scripts. . In summary, prior to processing, reads were filtered for EERC spike-ins, MALAT1, KCNQ1OT1, UGDH-AS1, and EEF1A. Cells expressing between 400 and 10000 unique genes, between 700 and 30000 RNA molecules and less than 40% mitochondrial

RNA were selected for downstream processing. In total, 6191 cells (4936 male, 1255 female) were divided into 18 distinct clusters (Suppl. Fig. XVIIAB). Cluster identities were determined by marker gene expression (Suppl. Fig. XVIIC) and by integrating identities previously assigned[46]. The module score for female-biased key drivers was calculated using the addModuleScore function in Seurat. Sex differential module scores for key drivers were determined using Welch two sample T-tests in R. Sex differential gene expression in the ACTA2+ SMC and CD14+CD68+ macrophages I module was determined using the FindMarkers function in Seurat.

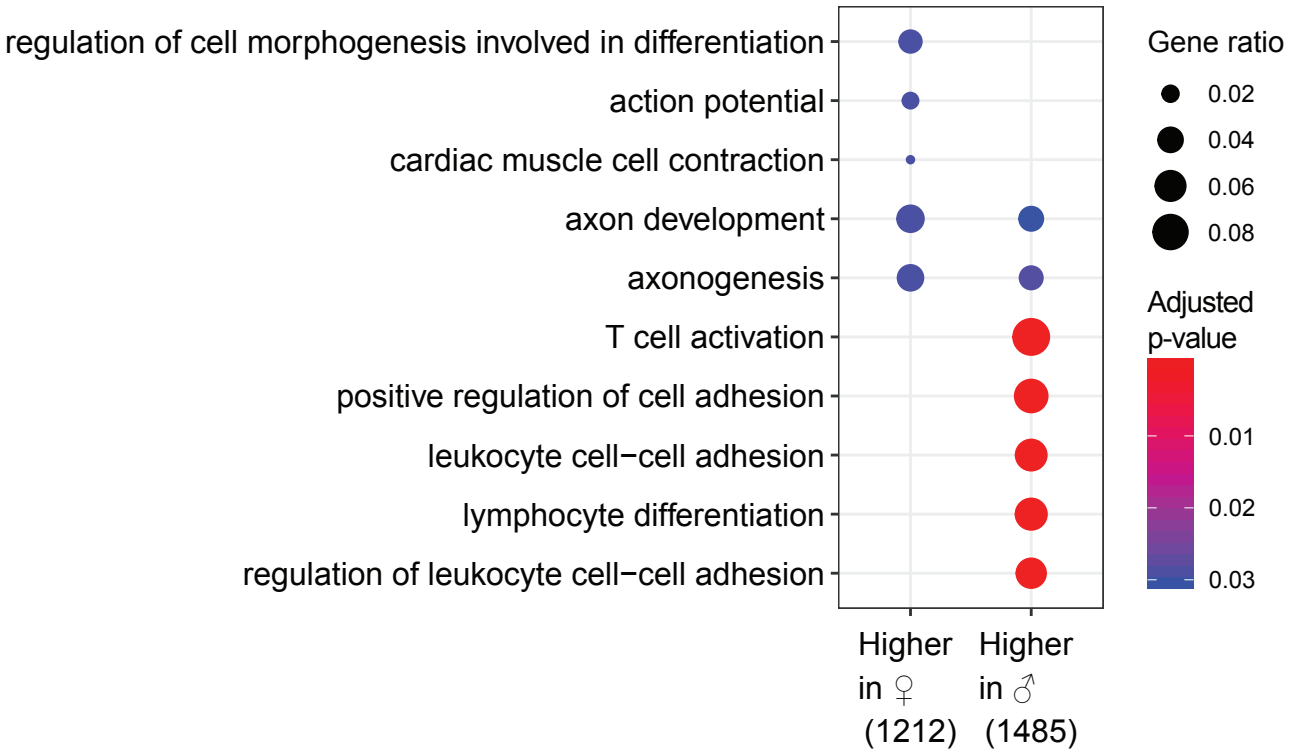
Mouse single-cell RNA-sequencing analysis

scRNA-seq data from microdissected murine late stage BCA plaques was collected as described previously[14]. Raw data is available at GSE150644. In brief, individual atherosclerotic plaques from 18 week Western diet-fed Myh11-Cre^{ERT2} eYFP ApoE$^{-/-}$ Klf4$^{wt/wt}$ or Myh11-Cre^{ERT2} eYFP ApoE$^{-/-}$ Klf4$^{\Delta/\Delta}$ brachiocephalic arteries (BCAs) were removed from underlying media with forceps and deposited into 1% BSA in PBS plus 1 ug/mL Actinomycin D,(Gibco, #11805017). 2,000 cells in each group were targeted in Chromium 10X genomics libraries, which after barcoding, were pooled and sequenced on the Illumina NextSeq™, 150 cycle high-output. Gene-barcode matrices were analyzed in R using Seurat v3. The analysis in this paper includes only the results from Myh11-Cre^{ERT2} eYFP ApoE$^{-/-}$ Klf4$^{wt/wt}$ or Myh11-Cre^{ERT2} eYFP ApoE$^{-/-}$ Klf4$^{\Delta/\Delta}$ mice, and includes 5469 cells from two different experiments (Chromium 10x runs) and 11 animals. Cells were filtered for 200-5000 reads per UMI, 10% or less mitochondrial and less than 5% hemoglobin gene content. Significant principal components of variation (PCs) were calculated using JackStraw test with 10000 repetitions, and clusters were calculated with 19 PCs. The whole dataset and subsets of SMC-lineage traced cells were queried for expression of blue, yellow, and green module genes with a min percent of 10% of cells in a cluster expressing the gene and at least 0.25 log fold-change between SMC^{Klf4} wild type and knockout samples. Differential expression analysis was done using MAST. Code is available upon re-

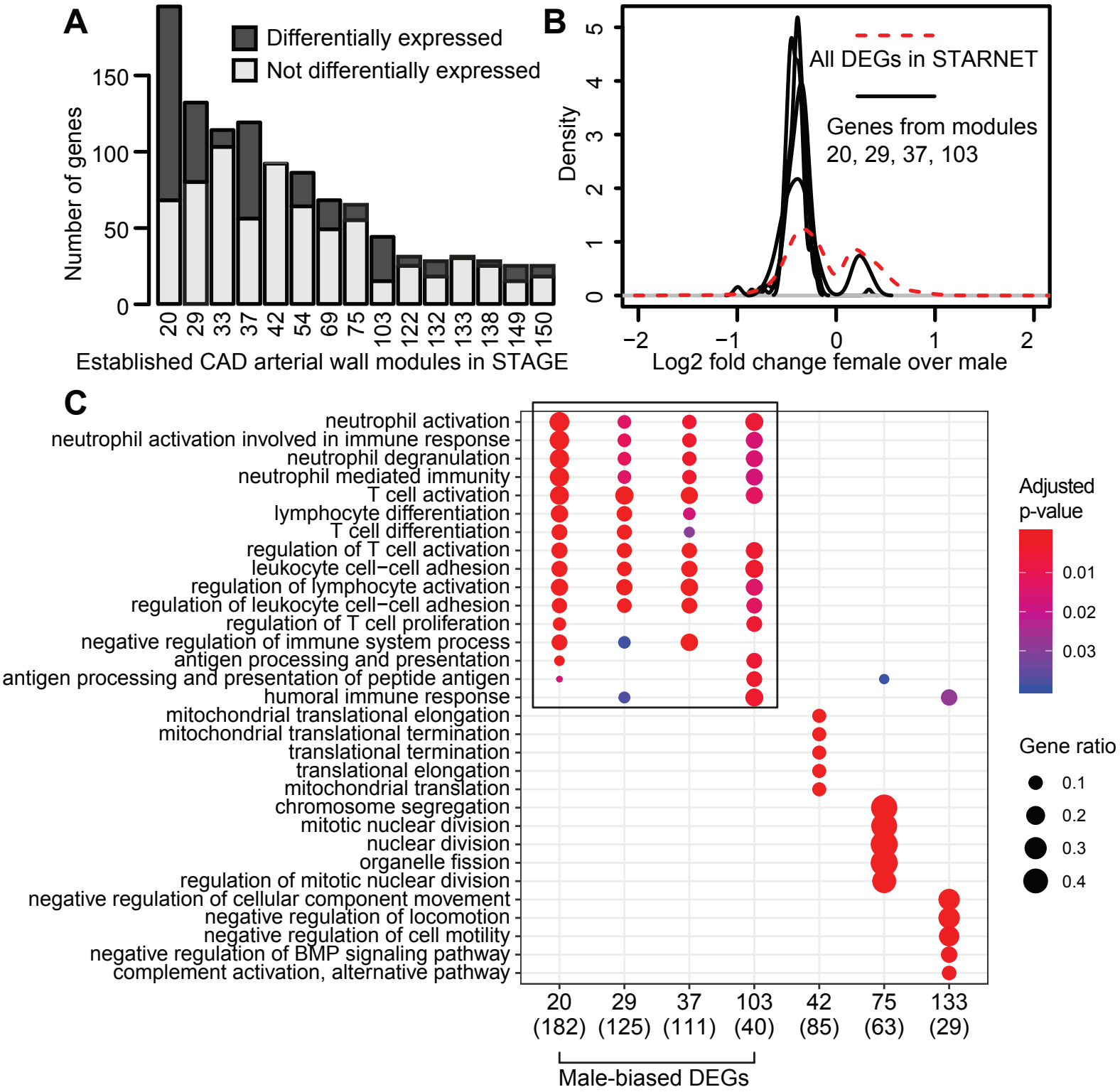
quest.

Supplemental Table I

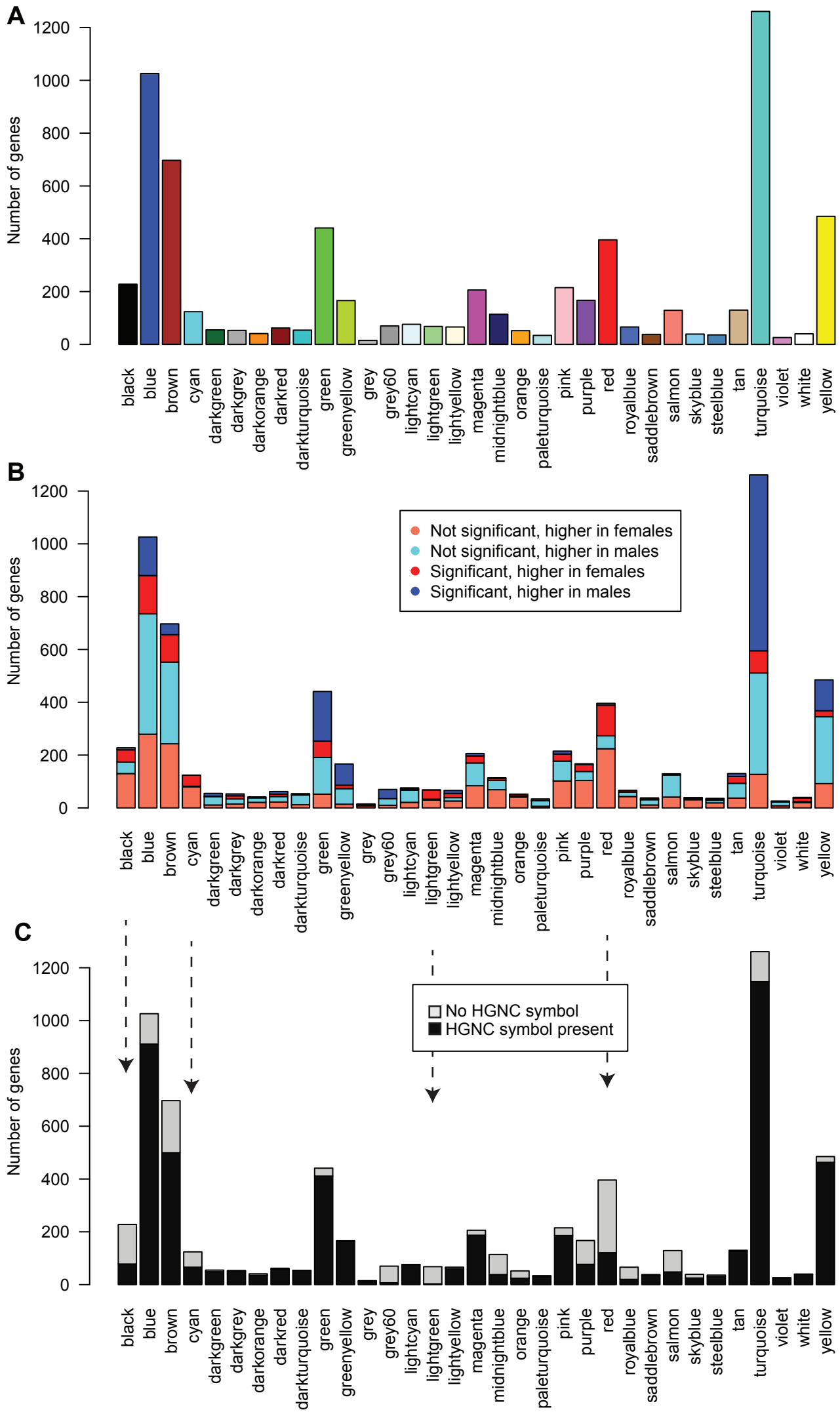
	Female	n = 160	**Male**	n = 160
	Mean	S.D.	Mean	S.D.
Age	69.22	6.91	69.57	7.15
BMI	29.54	4.69	29.06	4.51
CRP	6.93	11.36	3.62	16.88
Chol	5.31	1.38	4.83	1.25
LDL	3.06	1.22	2.55	1.32
HDL	1.81	1.13	2.24	3.22
TG	1.71	1.26	1.56	1.07
Syntax score	42.68	17.06	49.16	15.56
Blood glucose	6.39	2.10	6.57	1.66



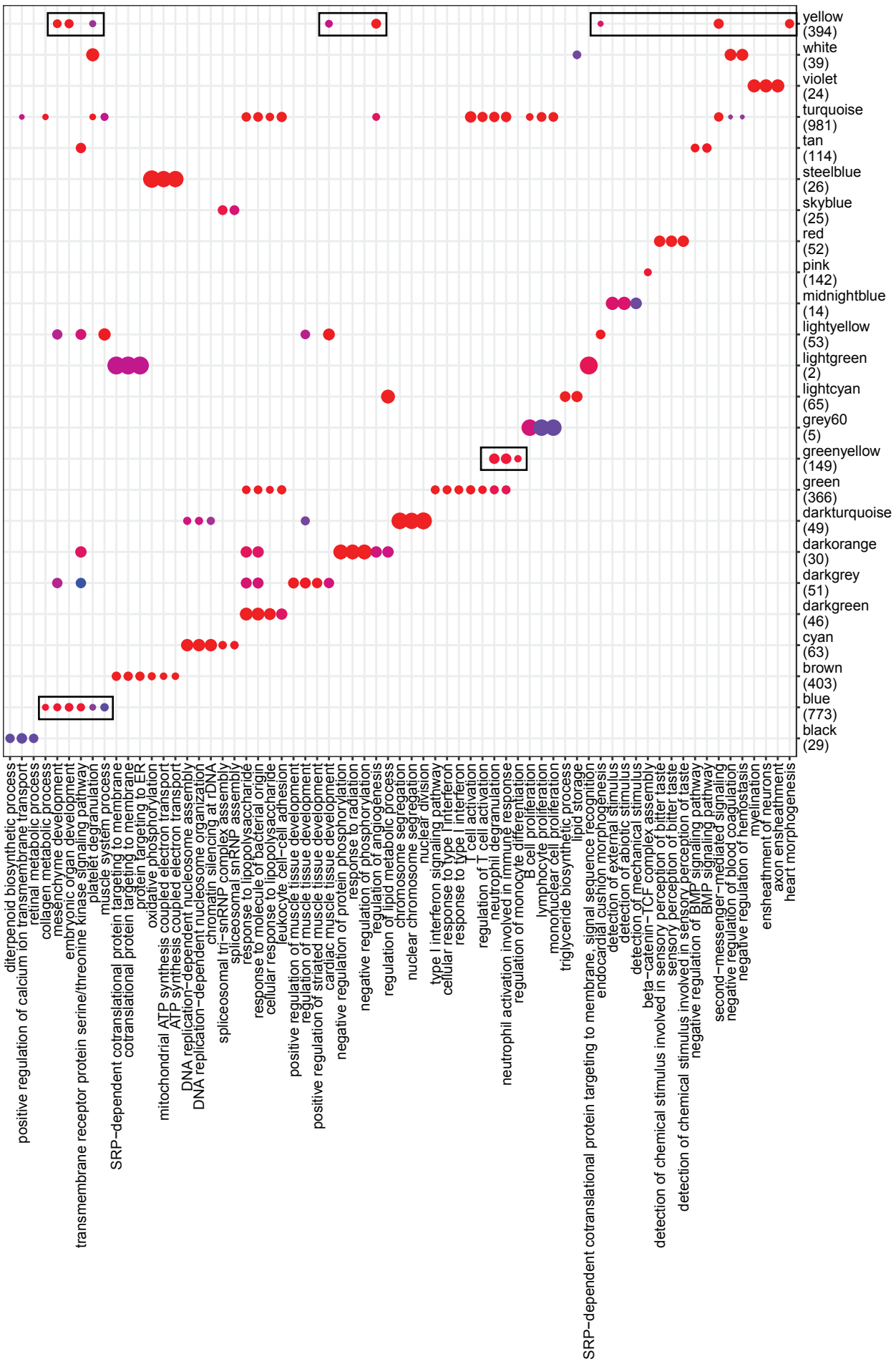
Suppl. Fig. I. Gene enrichment DEGs AOR. A dotplot is shown for the top 5 Gene Ontology enrichments for genes higher expressed in females, and those higher expressed in males. Terms are allocated to the rows, color indicates significance, and size of the dot indicates the ratio of genes from the set present. The number indicates the number of genes that could be found in any of the sets tested.



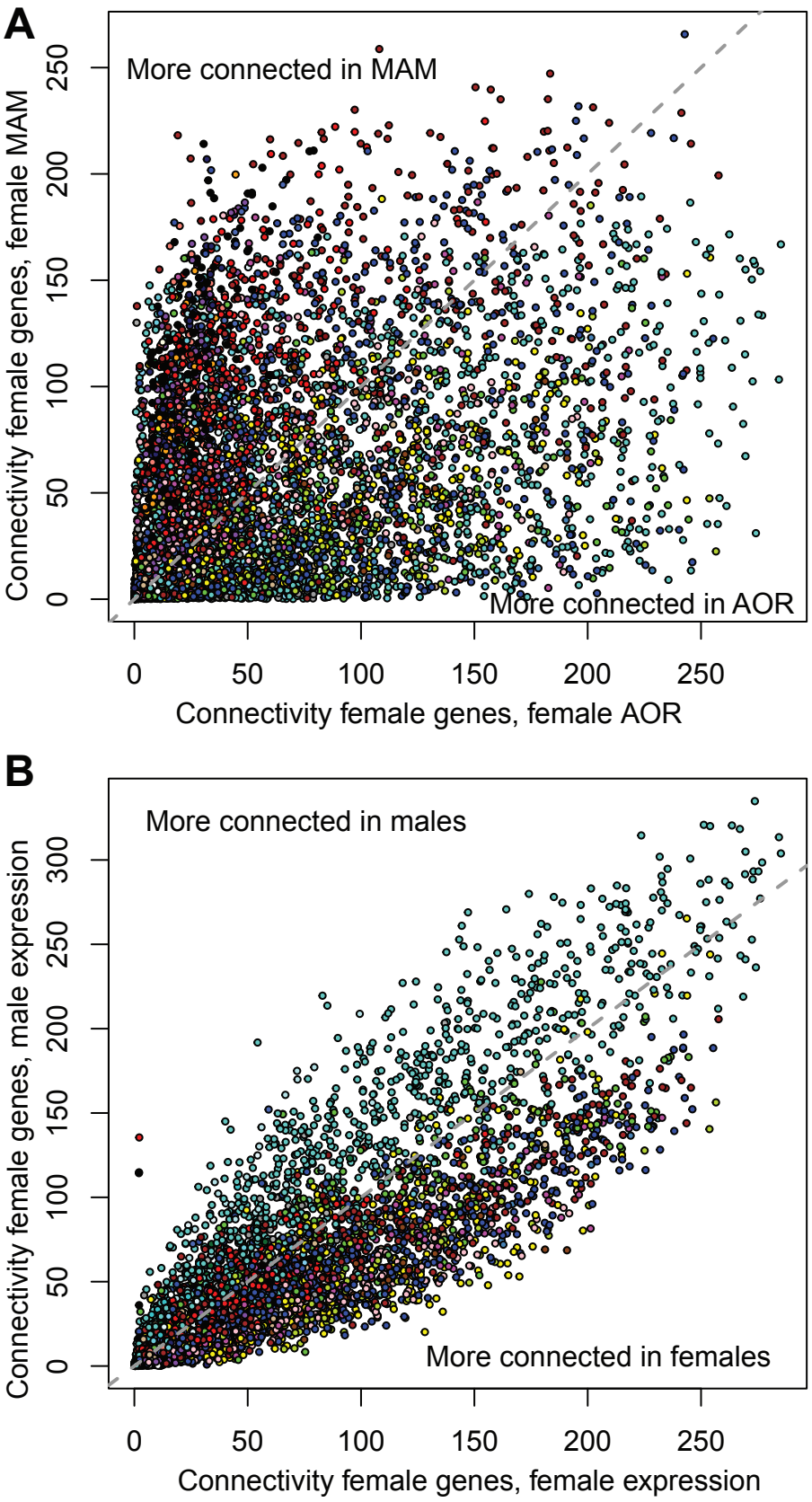
Suppl. Fig. II. Immune-related processes and male-biased STAGE modules. A) A barplot is shown with the number of DEGs (y-axis) between the sexes in AOR of STARNET in AAW modules of STAGE (x-axis). We test differential expression between the sexes using STARNET, since STAGE does not have enough power to detect sex differences. The STAGE modules are numbered with their original module number. Darkgrey indicates the proportion of DEGs between the sexes, while lightgrey points to the number of genes that are not differentially expressed. For example, STAGE module 20 contains a lot of differentially expressed genes between the sexes from STARNET. B) Density plots of log2 fold changes of different gene sets are shown as determined by DEG analysis between the sexes in STARNET, log fold change is female over male. The black lines show log2 fold changes of STAGE modules 20, 29, 37 and 103, which contain the most sex DEGs out of the male-biased AAW STAGE modules. The red line indicates log2 fold changes of all sex DEGs in STARNET AOR. C) A dotplot is shown for the top 5 Gene Ontology enrichments for genes present in multiple STAGE modules (4 AAW-specific modules that are biased towards higher expression in males; 20, 29, 37, 103, and 3 that are not; 42, 75, 133). Terms are allocated to the rows, color indicates significance, and size of the dot indicates the ratio of genes from the set present. The number indicates the number of genes that could be found in any of the sets tested. Immune-related processes are enriched in male-biased AAW-specific modules.



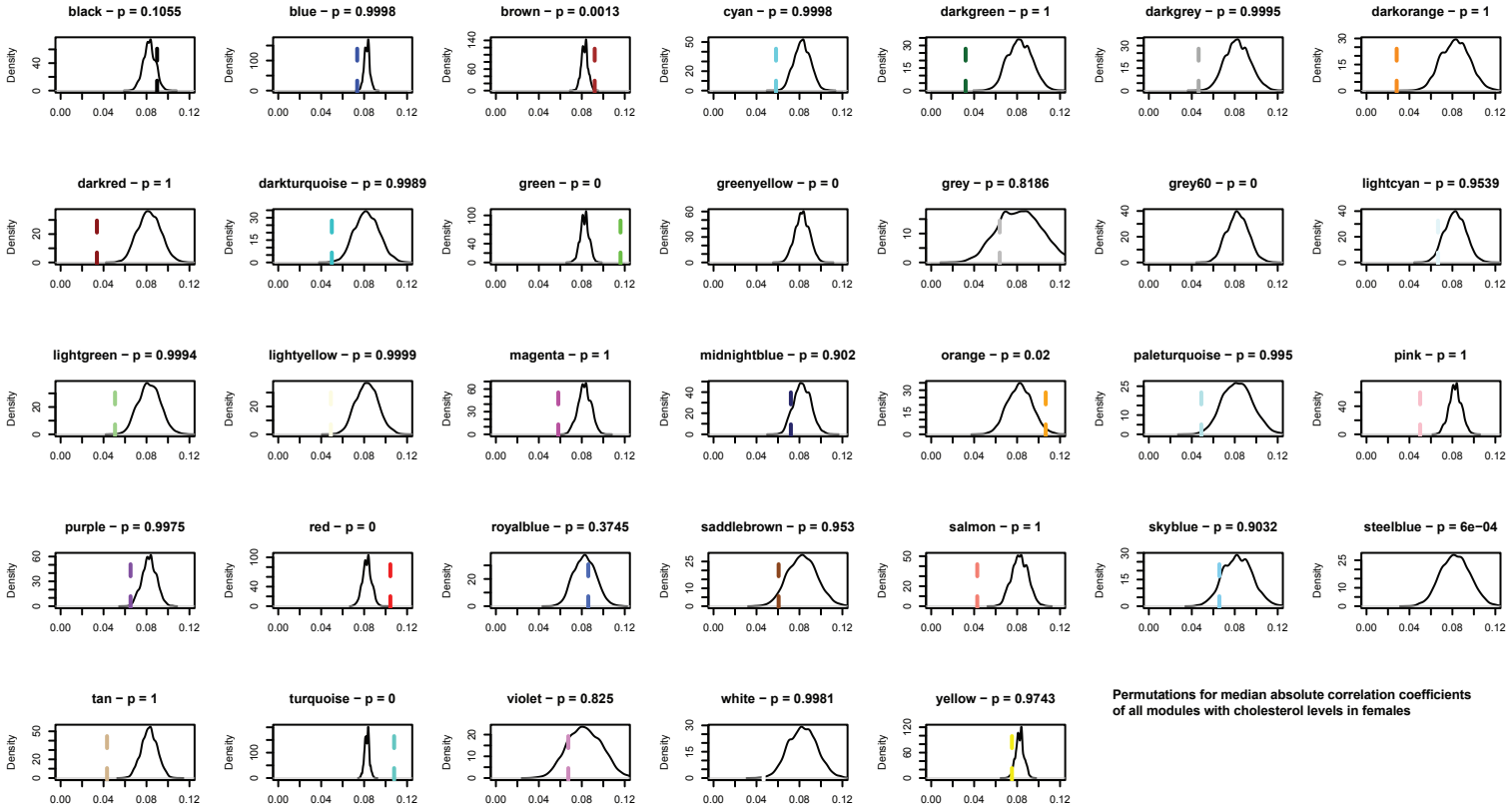
Suppl. Fig. III. Female module sizes, DEGs and annotation. A) A barplot is shown indicating the number of genes present in the module. Color = module. B) A barplot is shown indicating the number of genes present in the module and how many genes in the module are differentially expressed and whether or not they are expressed at a higher level in males or females. Color = direction and significance. C) A barplot is shown indicating the number of genes present in the module and how many genes in the module have an HGNC symbol attached to them. Color = presence of HGNC symbol. Dashed arrows indicate modules with high rate of of female-biased DEGs and lack of annotation.



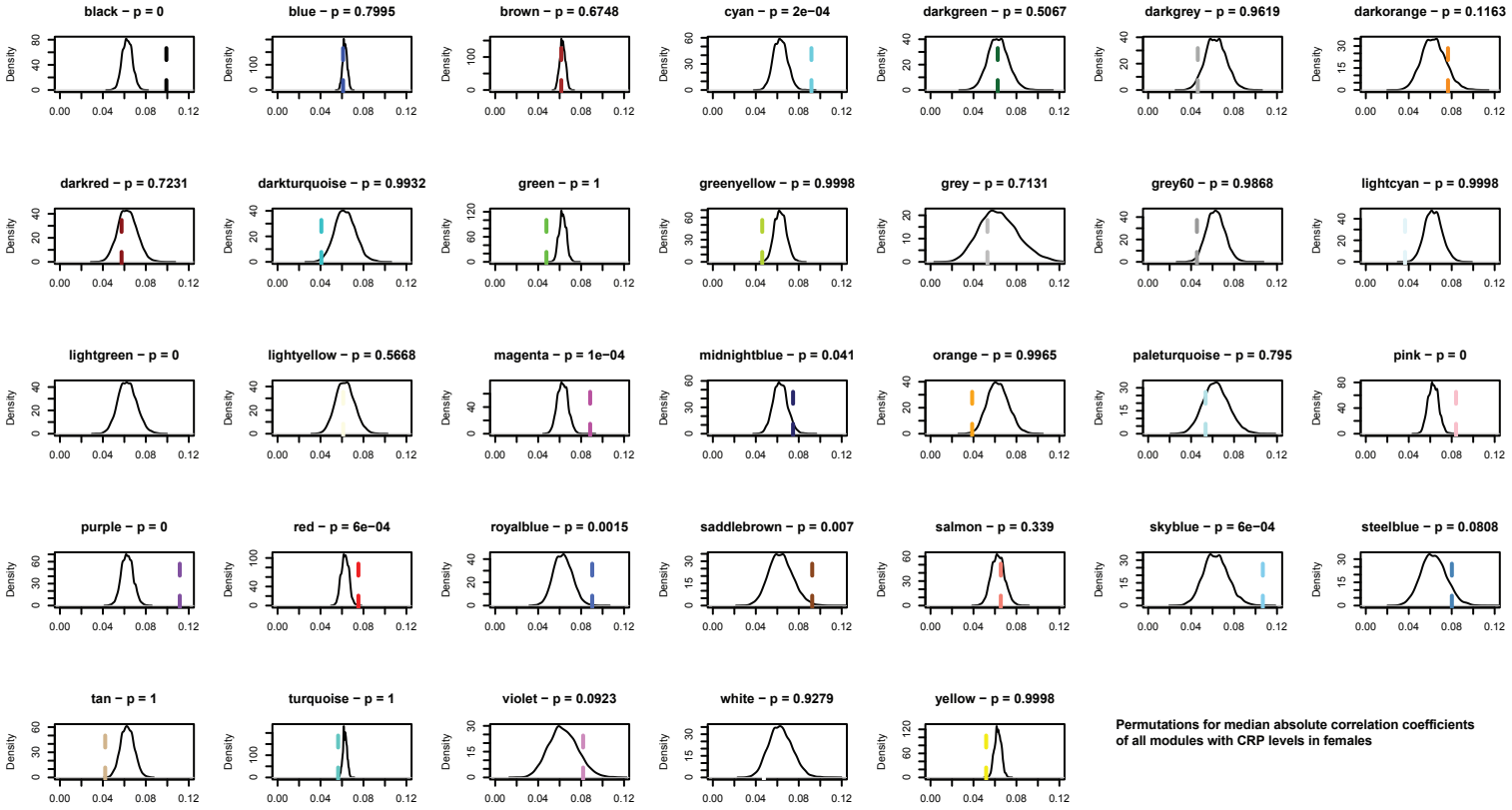
Suppl. Fig. IV. Gene ontology enrichment for generated modules. A dotplot is shown for the top 3 Gene Ontology enrichments for genes present in WGCNA-generated modules from female AOR. Terms are allocated to the columns, color indicates significance, and size of the dot indicates the ratio of genes from the set present. The number indicates the number of genes that could be found in any of the sets tested. Of the 32 tested modules, at least 24 are enriched for biological processes ranging from immunity to metabolism and muscle tissue.



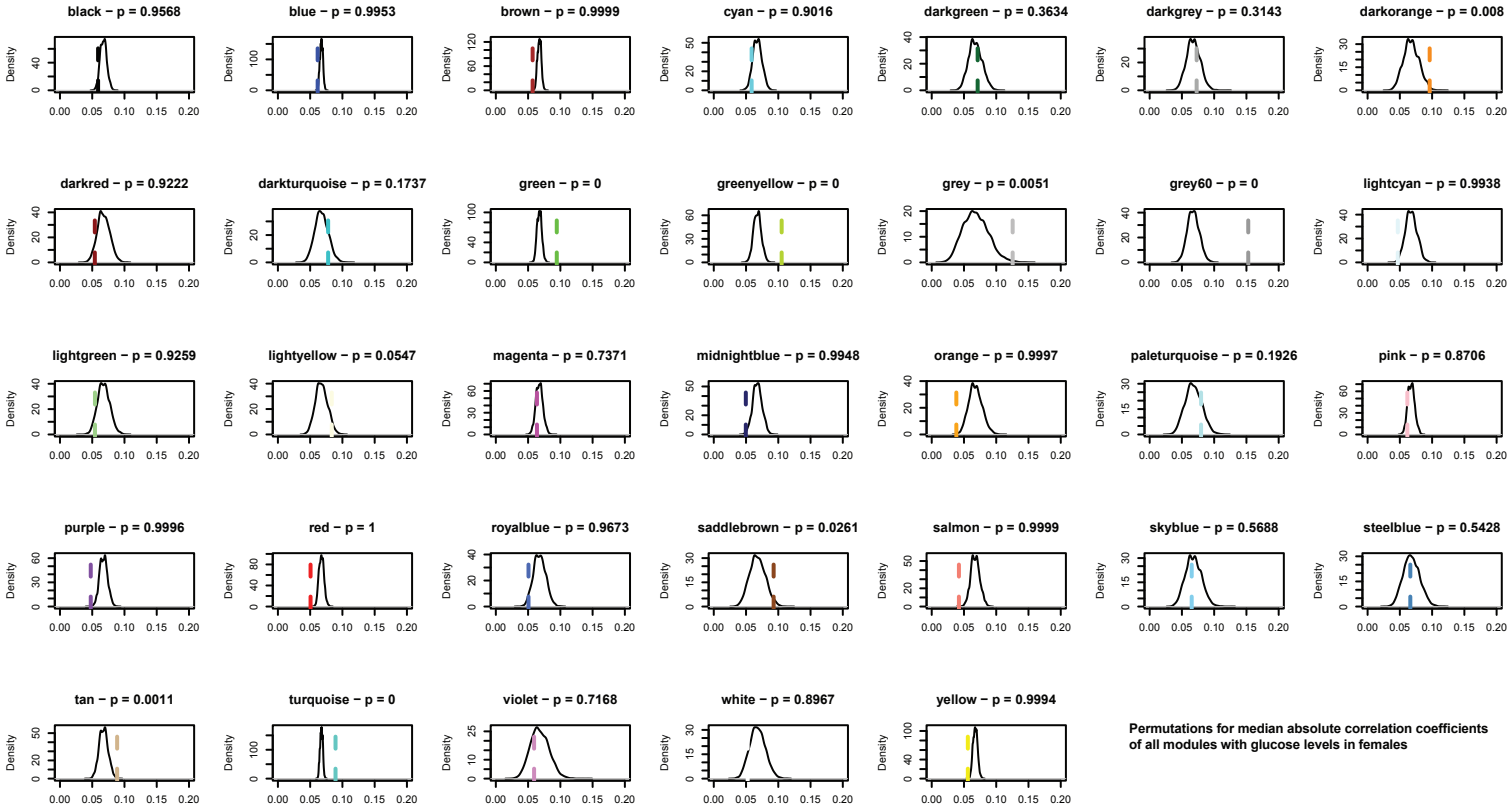
Suppl. Fig. V. Individual gene connectivity. A) A scatterplot is shown for all 6676 individual genes and their connectivity values in female AOR (x-axis) and female MAM (y-axis). Each dot represents a gene, its color represents its module. The dashed line indicates the line of identity. B) A scatterplot is shown for all 6676 individual genes and their connectivity values in female AOR (x-axis) and male AOR (y-axis). Each dot represents a gene, its color represents its module. The dashed line indicates the line of identity.



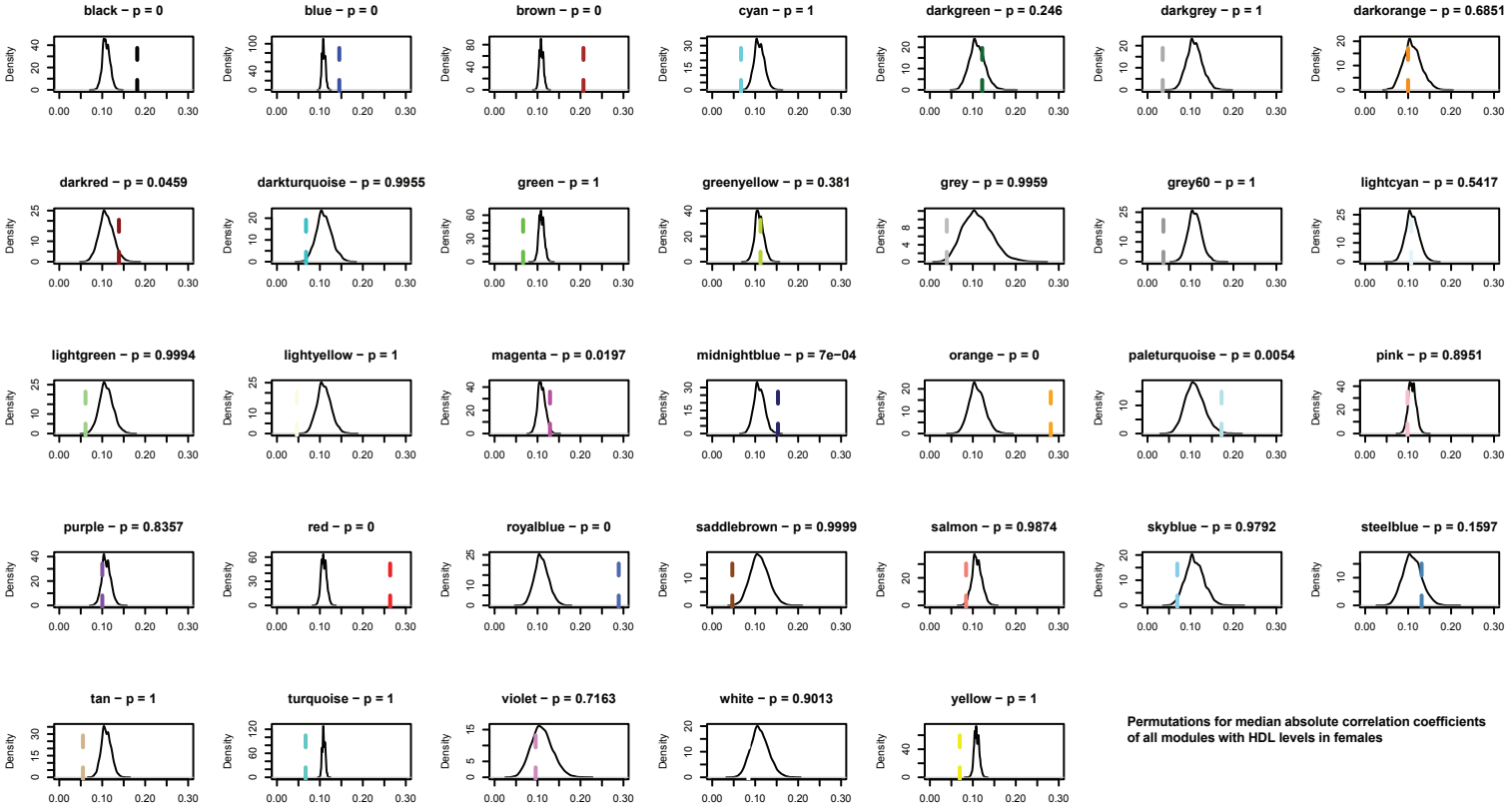
Suppl. Fig. VI. Permutations of cholesterol levels and module gene expression. Density plots are shown for median absolute correlation coefficients of all modules with cholesterol levels in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed median absolute correlation coefficients for that module.



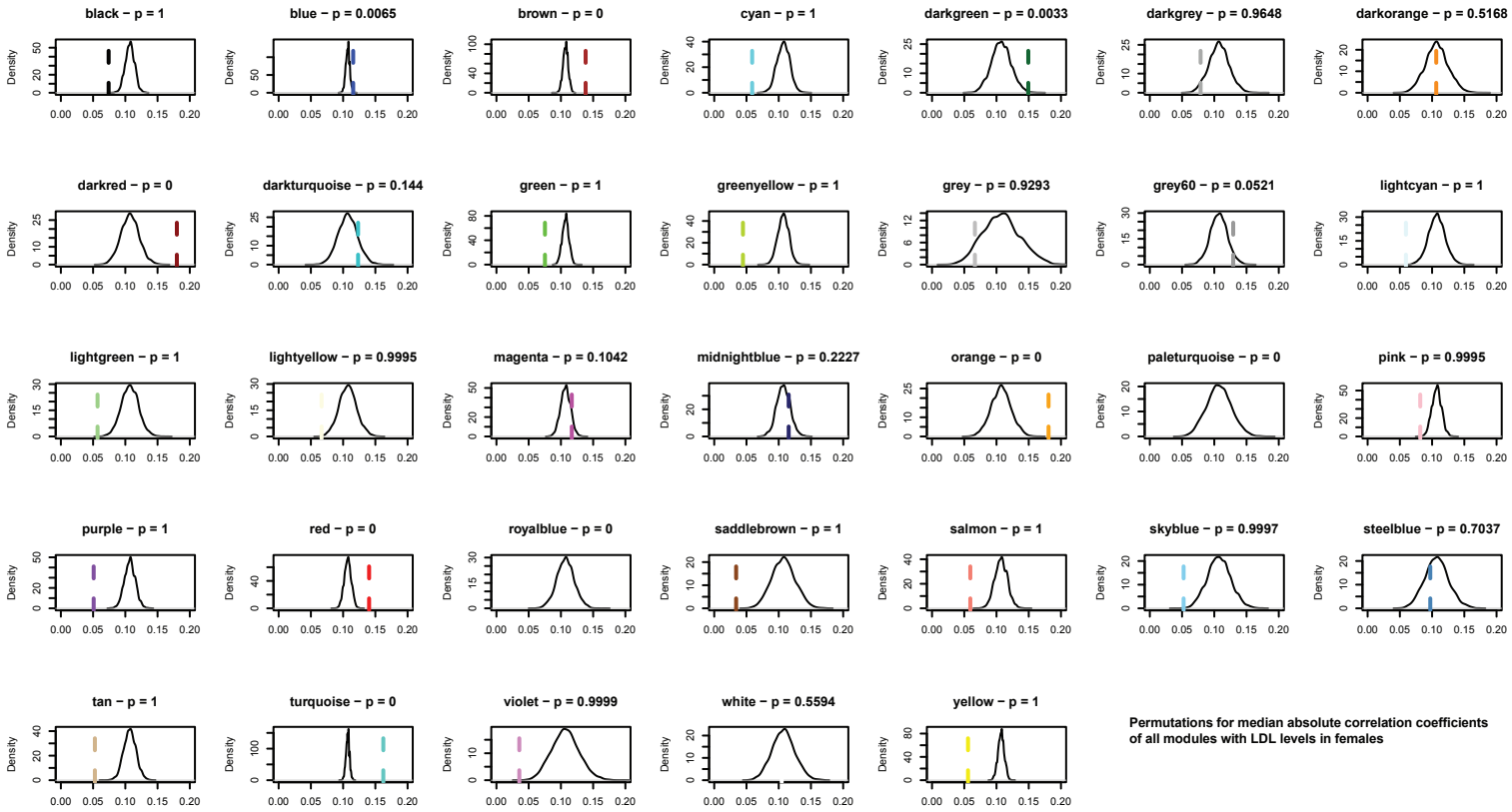
Suppl. Fig. VII. Permutations of CRP levels and module gene expression. Density plots are shown for median absolute correlation coefficients of all modules with CRP levels in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed median absolute correlation coefficients for that module.



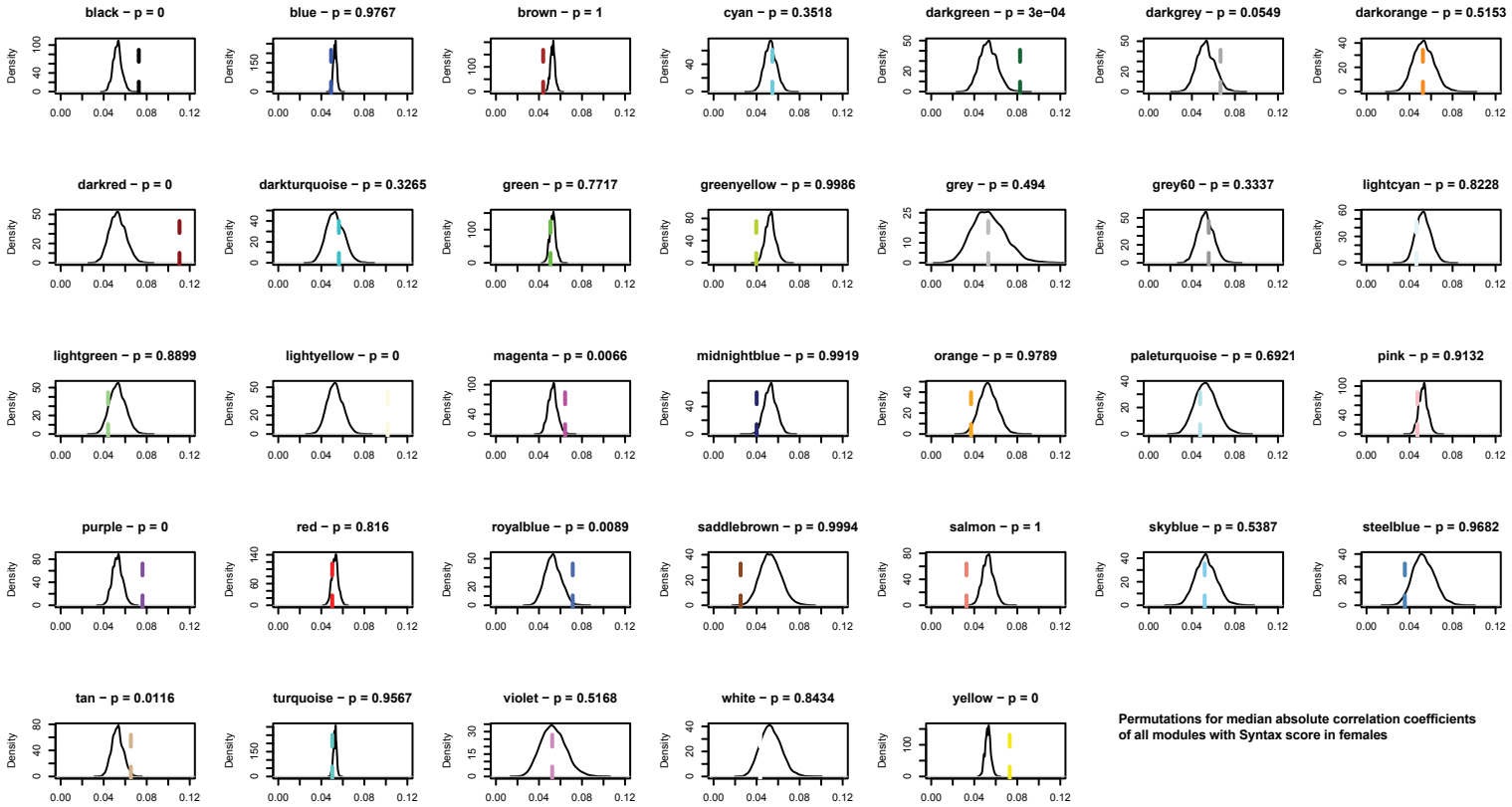
Suppl. Fig. VIII. Permutations of glucose levels and module gene expression. Density plots are shown for median absolute correlation coefficients of all modules with glucose levels in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed median absolute correlation coefficients for that module.



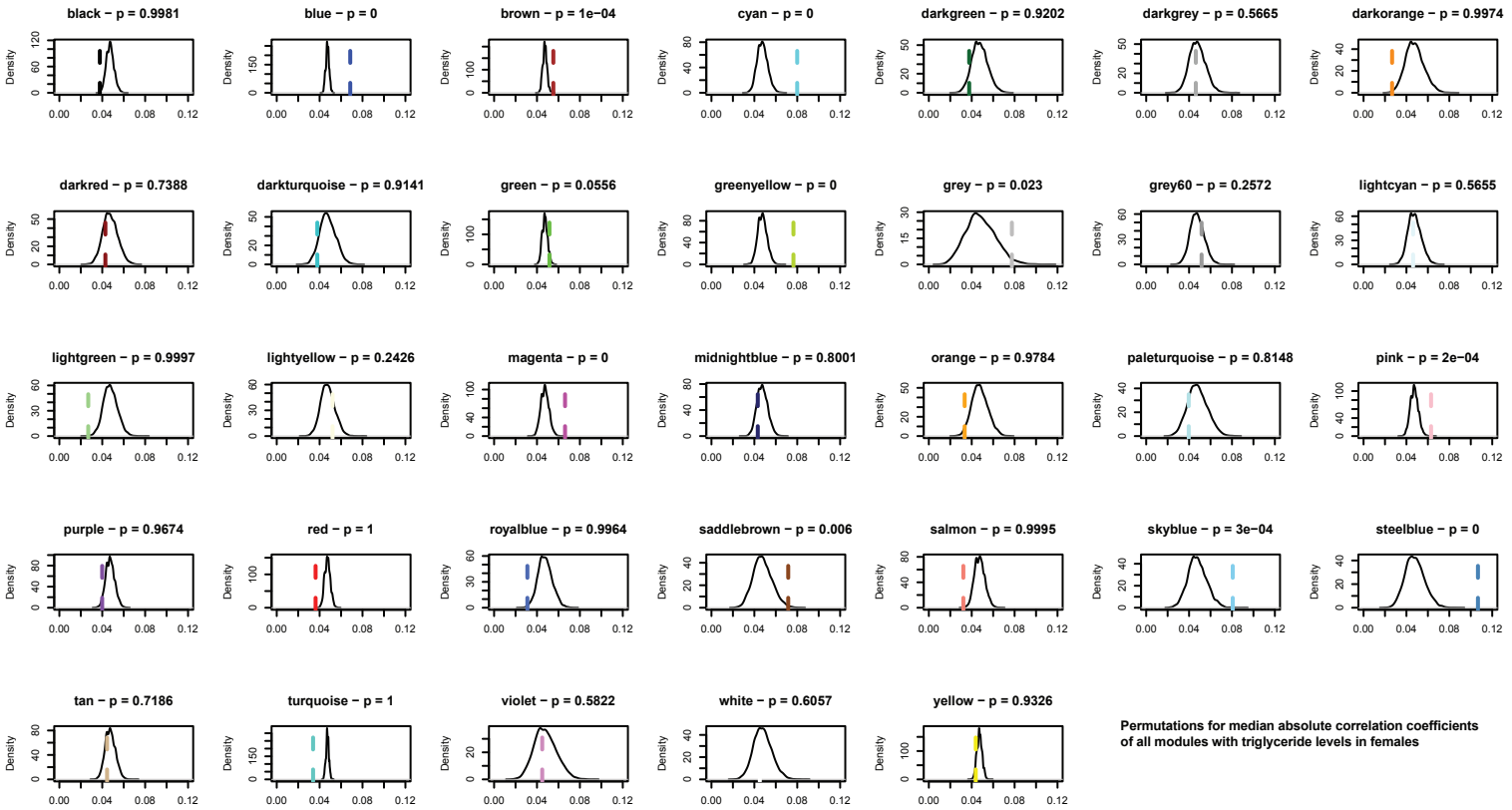
Suppl. Fig. IX. Permutations of HDL levels and module gene expression. Density plots are shown for median absolute correlation coefficients of all modules with HDL levels in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed median absolute correlation coefficients for that module.



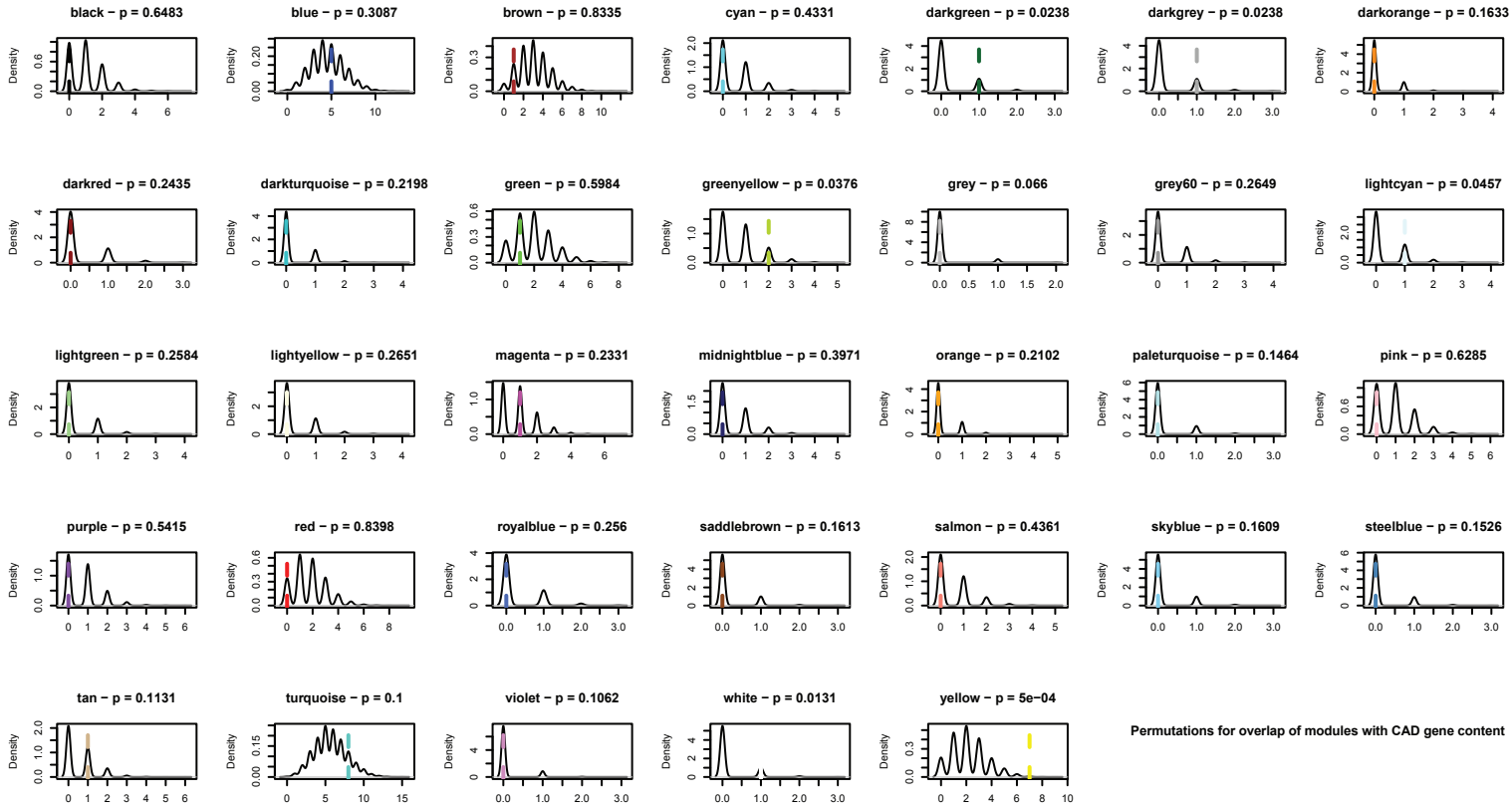
Suppl. Fig. X. Permutations of LDL levels and module gene expression. Density plots are shown for median absolute correlation coefficients of all modules with LDL levels in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed median absolute correlation coefficients for that module.



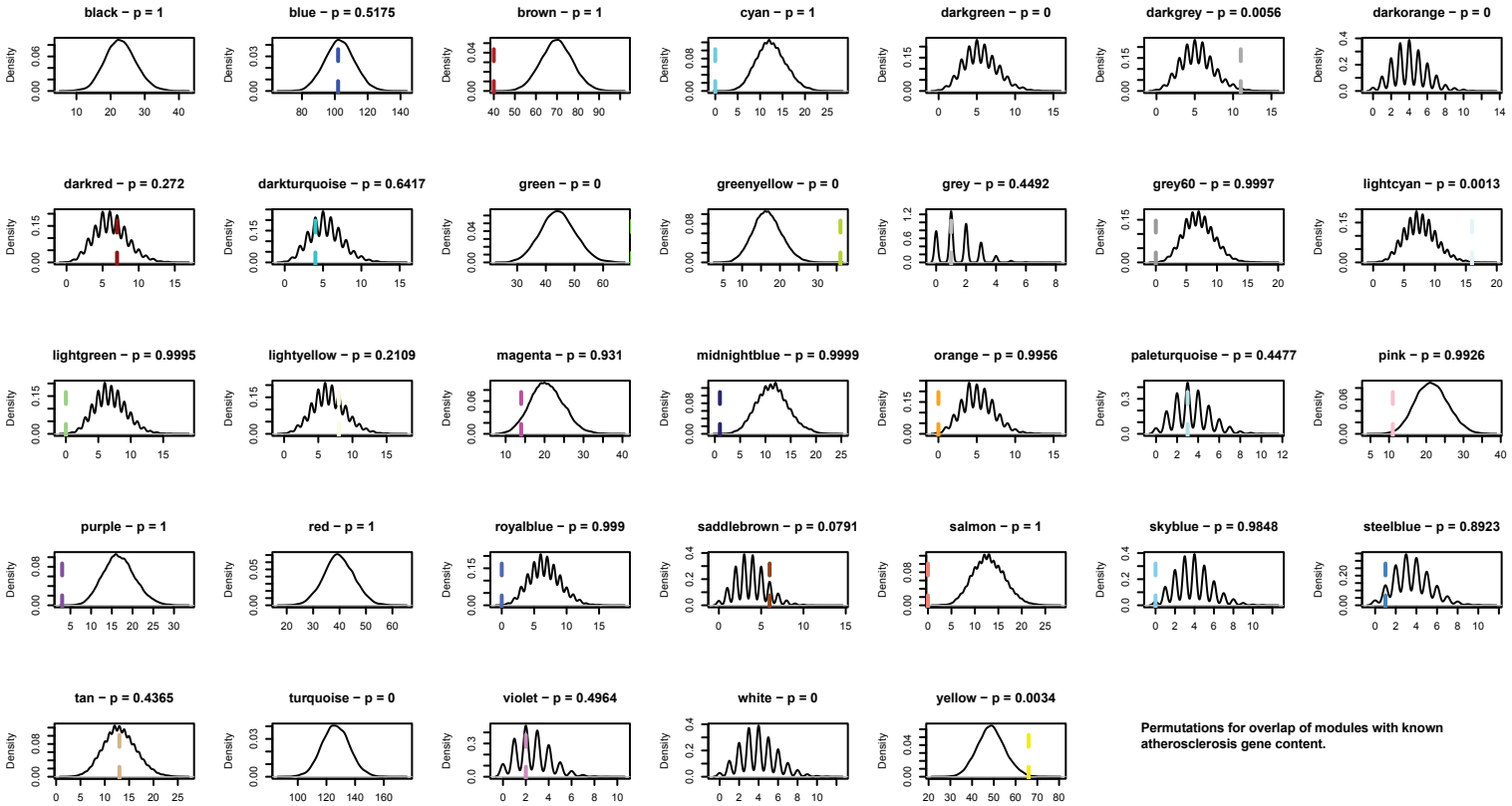
Suppl. Fig. XI. Permutations of syntax score and module gene expression. Density plots are shown for median absolute correlation coefficients of all modules with syntax score in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed median absolute correlation coefficients for that module.



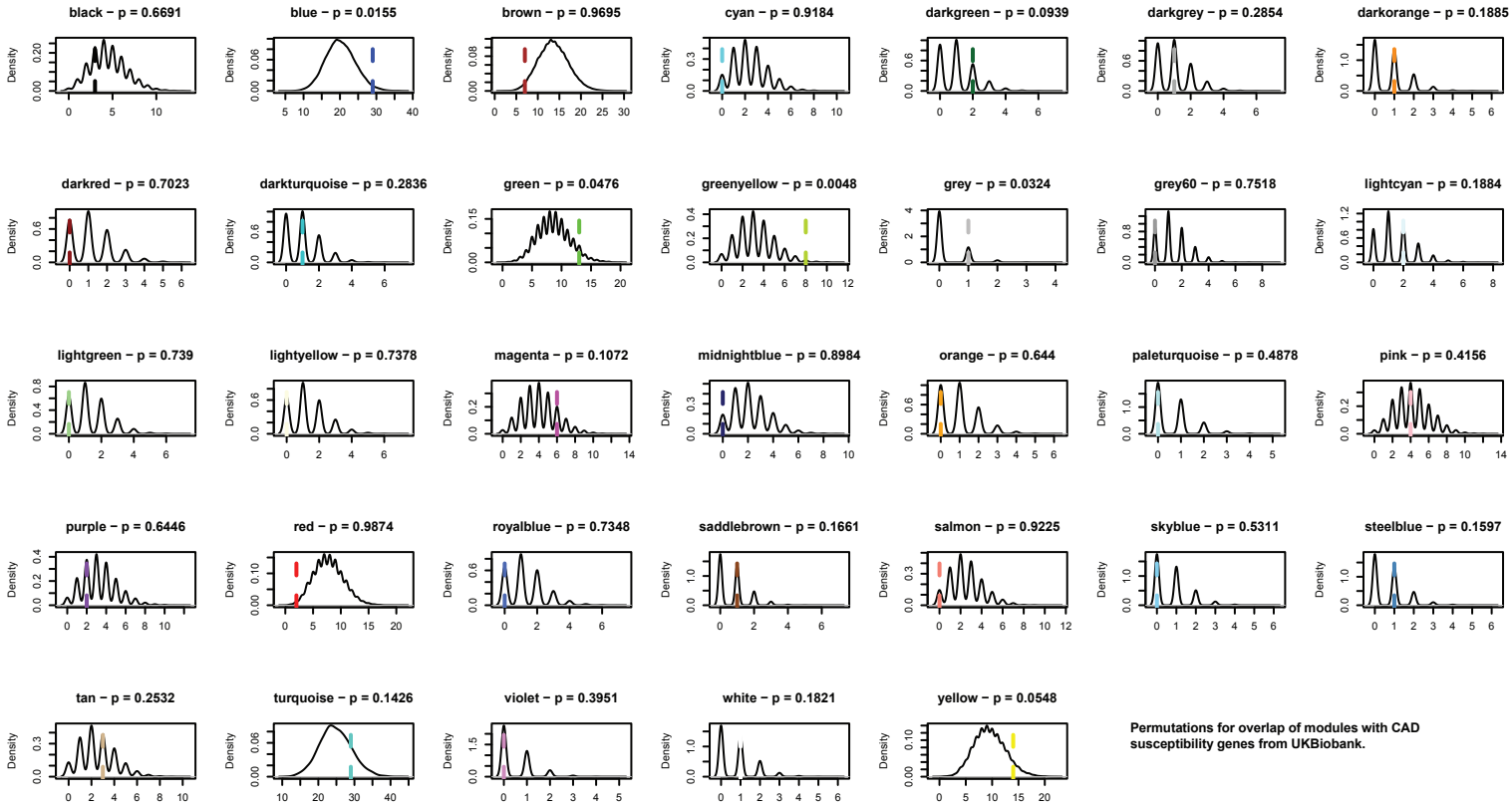
Suppl. Fig. XII. Permutations of triglyceride levels and module gene expression. Density plots are shown for median absolute correlation coefficients of all modules with triglyceride levels in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed median absolute correlation coefficients for that module.



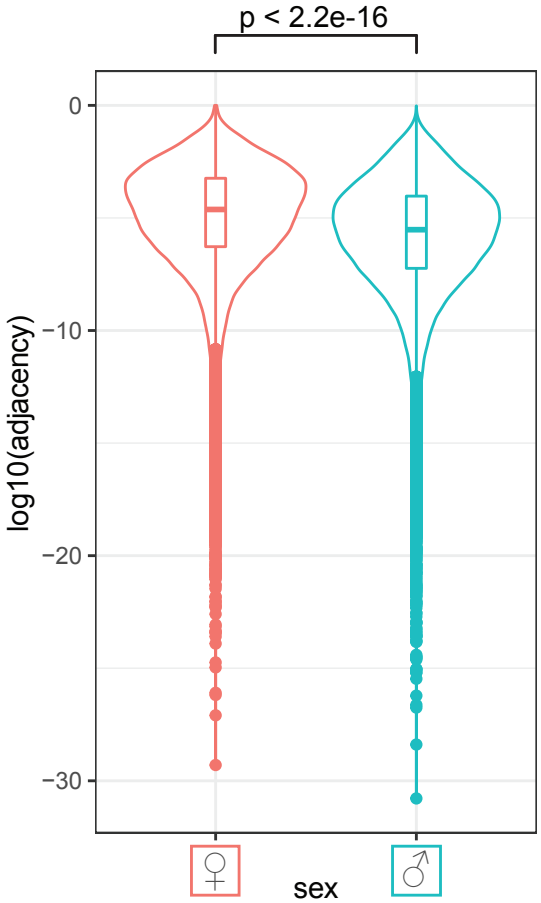
Suppl. Fig. XIII. Permutations of CAD GWAS gene content.
Density plots are shown for overlap permutations of random genes with CAD GWAS gene content in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed CAD GWAS gene content for that module.



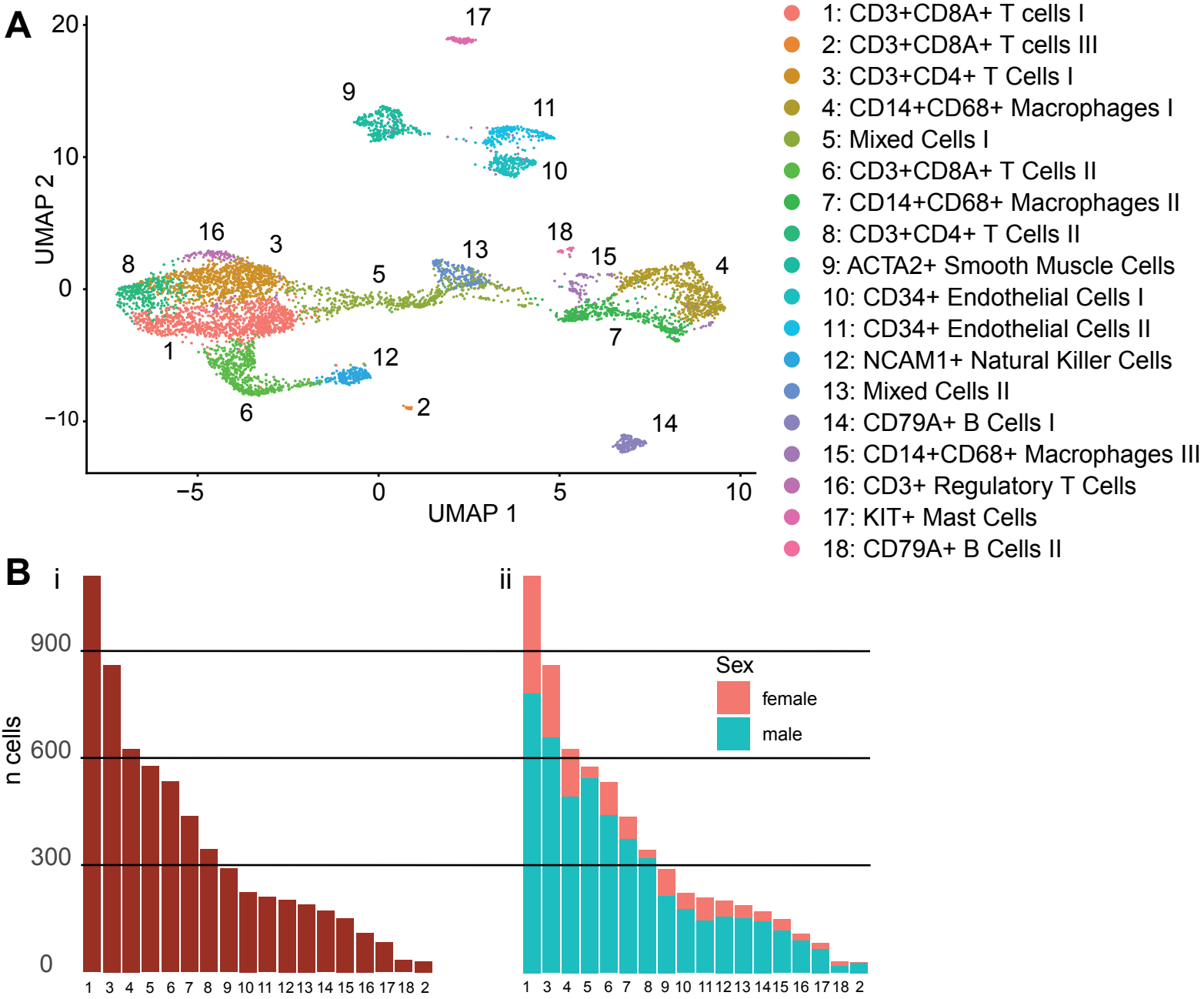
Suppl. Fig. XIV. Permutations of atherosclerosis gene content.
Density plots are shown for overlap permutations of random genes with atherosclerosis gene content in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed atherosclerosis gene content for that module.

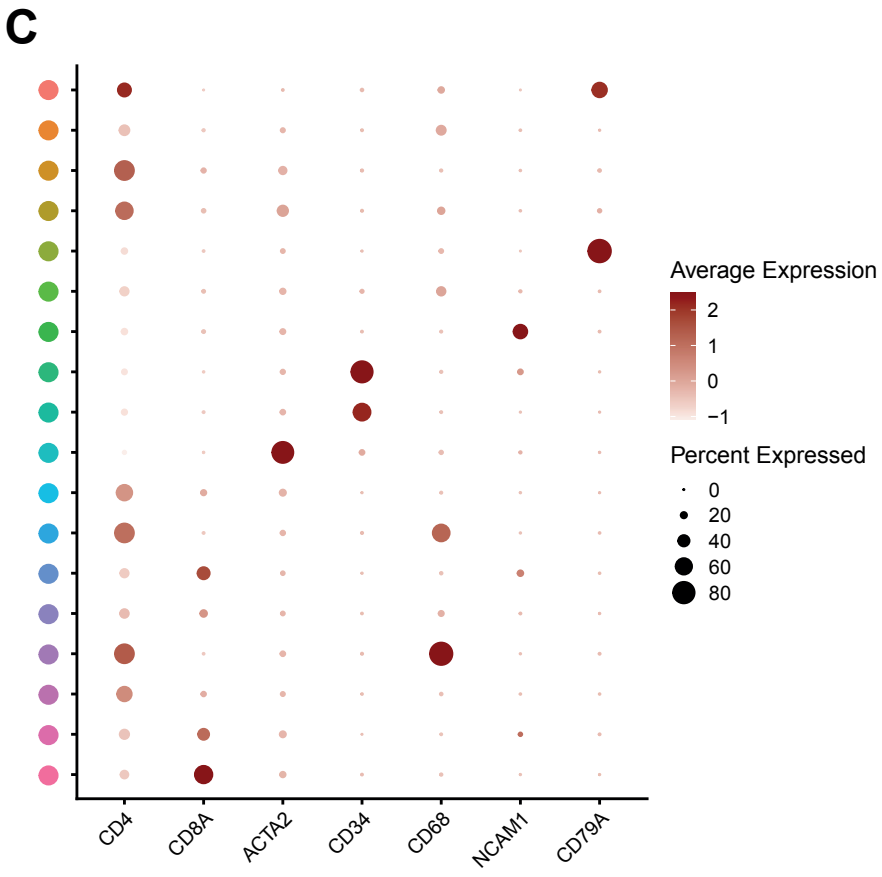


Suppl. Fig. XV. Permutations of CAD susceptibility gene content.
Density plots are shown for overlap permutations of random genes with CAD susceptibility gene content in females. Permutations were performed 10000x. P-value is shown in the title of the separate plots, 1 for each module. Dotted line indicates the observed CAD susceptibility gene content for that module.



Suppl. Fig. XVI. Validation activity yellow module in cultured endothelial cells. Violin boxplots are shown for log10-adjacency values of the yellow module in cultured human aortic endothelial cells. Adjacency values were calculated from microarray data on 43 females and 126 males (Welch Two Sample t-test; p<2.2e-16). Expression data are available in Gene Expression Omnibus accession GSE30169.





Suppl. Fig. XVII. scRNA sequencing of human atherosclerotic plaque. A) UMAP plot showing 18 distinct cell populations derived from 6191 human plaque cells. B) Distribution of number of cells per population (i) and their distribution by sex (ii). C) Average expression of marker genes across 18 cell populations.