Tutorial for the WGCNA package for R
II. Consensus network analysis of liver expression data, female and male mice

4. Relating consensus modules to external microarray sample information and exporting network analysis results

Peter Langfelder and Steve Horvath

February 13, 2016

Contents

0 Preliminaries: setting up the R session

Here we assume that a new R session has just been started. We load the WGCNA package, set up basic parameters and load data saved in the parts 1 and 2 of the tutorial.

```
# Display the current working directory
getwd();
# If necessary, change the path below to the directory where the data files are stored.
# "." means current directory. On Windows use a forward slash / instead of the usual \.
workingDir = ".";
setwd(workingDir);
# Load the WGCNA package
library(WGCNA)
# The following setting is important, do not omit.
options(stringsAsFactors = FALSE);
# Load the data saved in the first part
lnames = load(file = "Consensus-dataInput.RData");
#The variable lnames contains the names of loaded variables.
lnames
# Also load results of network analysis
lnames = load(file = "Consensus-NetworkConstruction-auto.RData");
lnames
exprSize = checkSets(multiExpr);
nSets = exprSize$nSets;
```

We have loaded the variables `multiExpr` and `Traits` containing the expression and trait data, respectively. Further, expression data dimensions are stored in `nGenes` and `nSamples`. The consensus network analysis results are represented by the variables `consMEs`, `moduleLabels`, `moduleColors`, and `consTree`.

4 Relating consensus modules to external microarray sample information

In this section we illustrate the use of module eigengenes to relate consensus modules to external microarray sample information such as classical clinical traits. In this analysis we have available several clinical traits. We relate the traits to consensus module eigengenes in each of the two sets. We remind the reader that while the consensus modules is a single module assignment for all genes, the module eigengenes represent the modules in each of the two sets. In other words, we have a single module assignment for each gene, but we have two sets of consensus module eigengenes, because a given module (set of genes) has a particular expression profile in the female mice, and a different expression profile in the male mice. Similarly, we have the trait data separately for the female and for the male mice.

```
# Set up variables to contain the module-trait correlations
moduleTraitCor = list();
moduleTraitPvalue = list();
# Calculate the correlations
for (set in 1:nSets)
{
  moduleTraitCor[[set]] = cor(consMEs[[set]]$data, Traits[[set]]$data, use = "p");
  moduleTraitPvalue[[set]] = corPvalueFisher(moduleTraitCor[[set]], exprSize$nSamples[set]);
}
```

We now display the module-trait relationships using a color-coded table. We print the correlations and the corresponding p-values, and color-code the entris by the p-value significance.

```
# Convert numerical lables to colors for labeling of modules in the plot
MEColors = labels2colors(as.numeric(substring(names(consMEs[[1]]$data), 3)));
MEColorNames = paste("ME", MEColors, sep="");
# Open a suitably sized window (the user should change the window size if necessary)
sizeGrWindow(10,7)
#pdf(file = "Plots/ModuleTraitRelationships-female.pdf", wi = 10, he = 7);
# Plot the module-trait relationship table for set number 1
set = 1
textMatrix = paste(signif(moduleTraitCor[[set]], 2), "\n(",
                        signif(moduleTraitPvalue[[set]], 1), ")", sep = "");
dim(textMatrix) = dim(moduleTraitCor[[set]])
par(mar = c(6, 8.8, 3, 2.2));
labeledHeatmap(Matrix = moduleTraitCor[[set]],
             xLabels = names(Traits[[set]]$data),
             yLabels = MEColorNames,
             ySymbols = MEColorNames,
             colorLabels = FALSE,
             colors = greenWhiteRed(50),
             textMatrix = textMatrix,
             setStdMargins = FALSE,
             cex.text = 0.5,
             zlim = c(-1,1),
             main = paste("Module--trait relationships in", setLabels[set]))
dev.off();
# Plot the module-trait relationship table for set number 2
set = 2
textMatrix = paste(signif(moduleTraitCor[[set]], 2), "\n(",
                        signif(moduleTraitPvalue[[set]], 1), ")", sep = "");
dim(textMatrix) = dim(moduleTraitCor[[set]])
sizeGrWindow(10,7)
#pdf(file = "Plots/ModuleTraitRelationships-male.pdf", wi = 10, he = 7);
par(mar = c(6, 8.8, 3, 2.2));
```

```
labeledHeatmap(Matrix = moduleTraitCor[[set]],
              xLabels = names(Traits[[set]]$data),
              yLabels = MEColorNames,
              ySymbols = MEColorNames,
              colorLabels = FALSE,
              colors = greenWhiteRed(50),
              textMatrix = textMatrix,
              setStdMargins = FALSE,
              cex.text = 0.5,
              zlim = c(-1,1),
              main = paste("Module--trait relationships in", setLabels[set]))
dev.off();
```

The two tables are shown in Figs. 1 and 2. The two module-trait relationship tables look similar but they are not the same. For example, they both identify the turquoise, purple and green modules as highly related to weight, although the actual correlations and p-values differ slightly. There are several ways of forming a measure of module-trait relationships that summarizes the two sets into one measure. We will form a very conservative one: for each module-trait pair we take the correlation that has the lower absolute value in the two sets if the two correlations have the same sign, and zero relationship if the two correlations have opposite signs:

```
# Initialize matrices to hold the consensus correlation and p-value
consensusCor = matrix(NA, nrow(moduleTraitCor[[1]]), ncol(moduleTraitCor[[1]]));
consensusPvalue = matrix(NA, nrow(moduleTraitCor[[1]]), ncol(moduleTraitCor[[1]]));
# Find consensus negative correlations
negative = moduleTraitCor[[1]] < 0 & moduleTraitCor[[2]] < 0;
consensusCor[negative] = pmax(moduleTraitCor[[1]][negative], moduleTraitCor[[2]][negative]);
consensusPvalue[negative] = pmax(moduleTraitPvalue[[1]][negative], moduleTraitPvalue[[2]][negative]);
# Find consensus positive correlations
positive = moduleTraitCor[[1]] > 0 & moduleTraitCor[[2]] > 0;
consensusCor[positive] = pmin(moduleTraitCor[[1]][positive], moduleTraitCor[[2]][positive]);
consensusPvalue[positive] = pmax(moduleTraitPvalue[[1]][positive], moduleTraitPvalue[[2]][positive]);
```

We display the consensus module–trait relationships again using a color-coded table:

```
textMatrix = paste(signif(consensusCor, 2), "\n(",
                        signif(consensusPvalue, 1), ")", sep = "");
dim(textMatrix) = dim(moduleTraitCor[[set]])
sizeGrWindow(10,7)
#pdf(file = "Plots/ModuleTraitRelationships-consensus.pdf", wi = 10, he = 7);
par(mar = c(6, 8.8, 3, 2.2));
labeledHeatmap(Matrix = consensusCor,
              xLabels = names(Traits[[set]]$data),
              yLabels = MEColorNames,
              ySymbols = MEColorNames,
              colorLabels = FALSE,
              colors = greenWhiteRed(50),
              textMatrix = textMatrix,
              setStdMargins = FALSE,
              cex.text = 0.5,
              zlim = c(-1,1),
              main = paste("Consensus module--trait relationships across\n",
                          paste(setLabels, collapse = " and ")))
```

The table is shown in Fig. 3. The advantage of the consensus relationship table is that it isolates the module-trait relationships that are present in both sets, and hence may be in a sense considered validated. For example, we confirm that the turquoise, purple, and green modules are highly related to weight in both sets; the brown module is highly related to insulin levels etc. One could now adapt the gene selection technique illustrated in the female expression analysis tutorial to look for particular genes that are highly related to traits as well as being highly connected in a module related to the trait; we leave this analysis as an exercise for the reader.

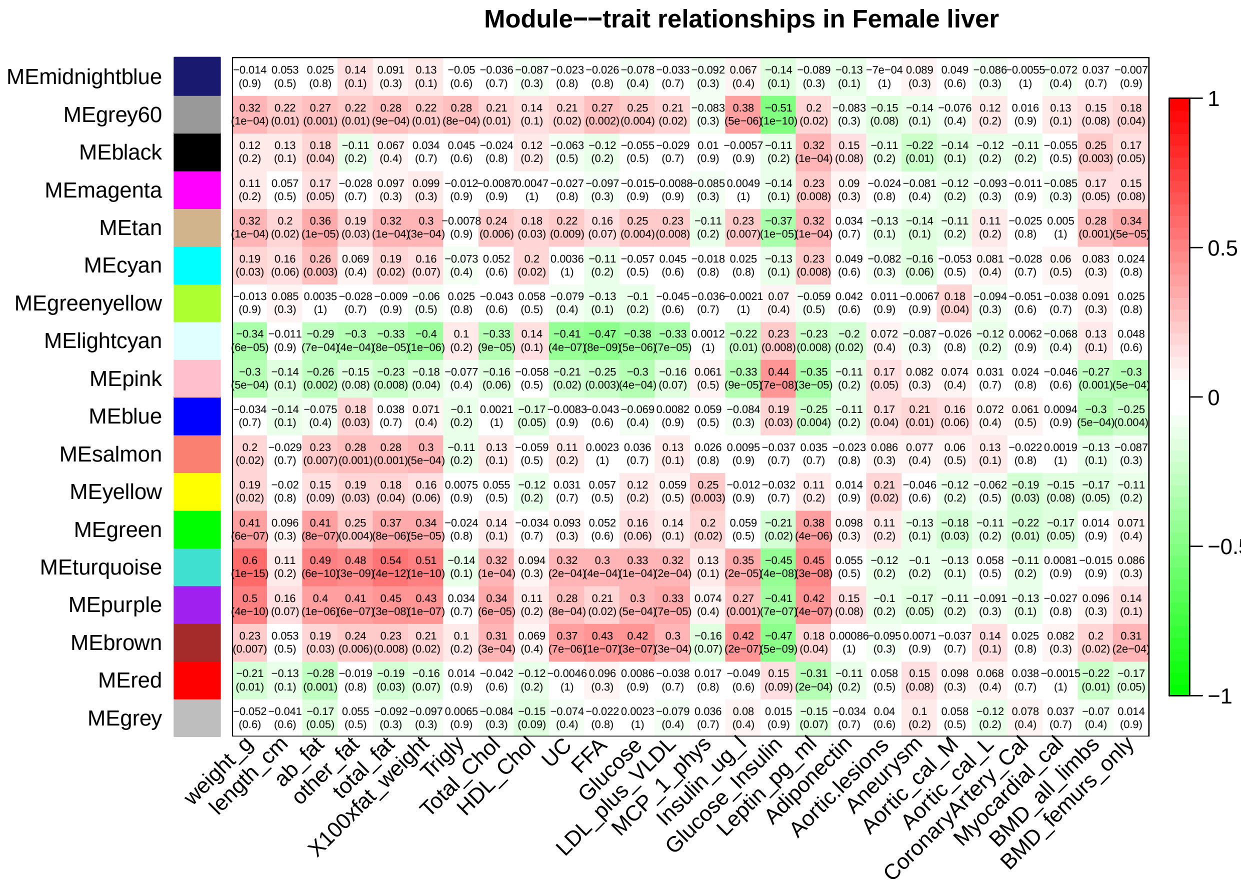


Figure 1: Relationships of consensus module eigengenes and clinical traits in the female data. Each row in the table corresponds to a consensus module, and each column to a trait. Numbers in the table report the correlations of the corresponding module eigengenes and traits, with the p-values printed below the correlations in parentheses. The table is color coded by correlation according to the color legend.

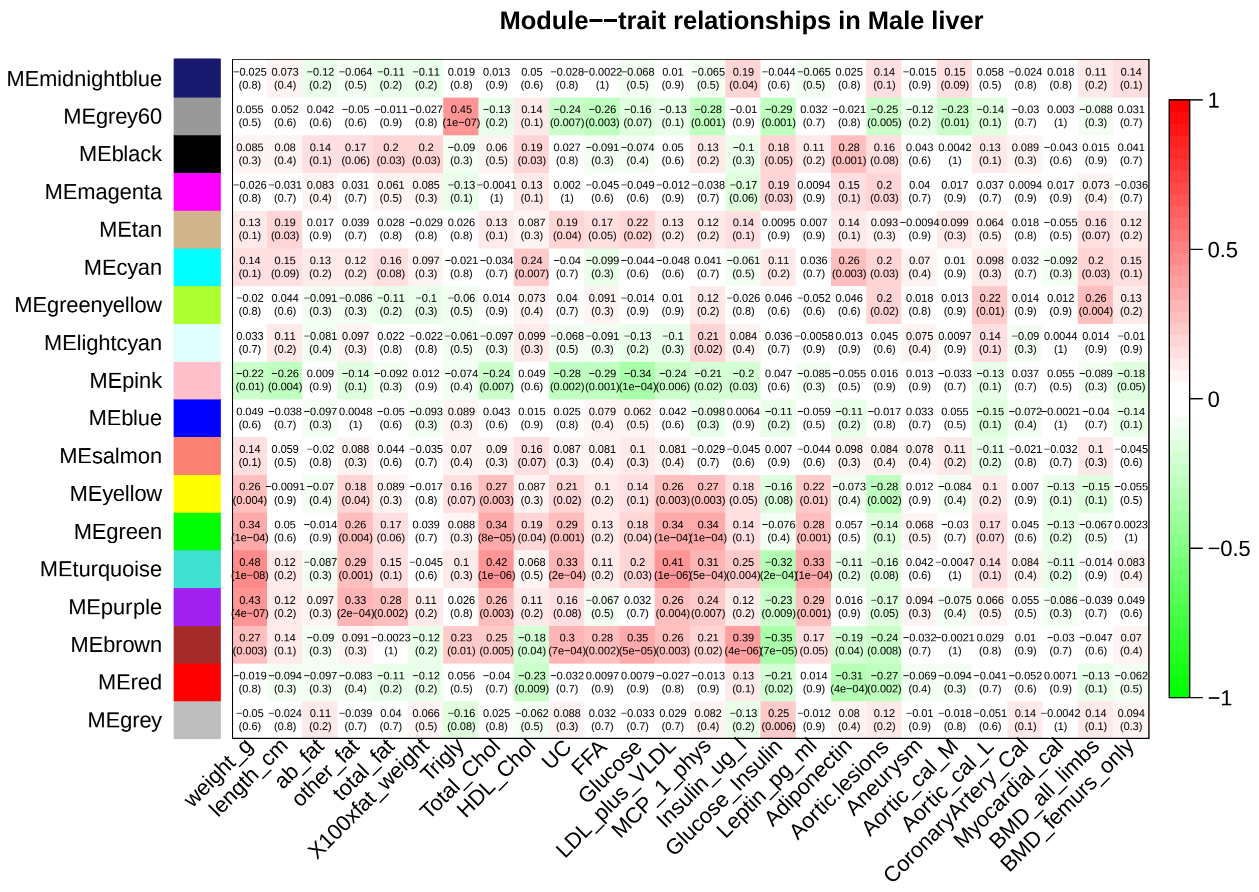
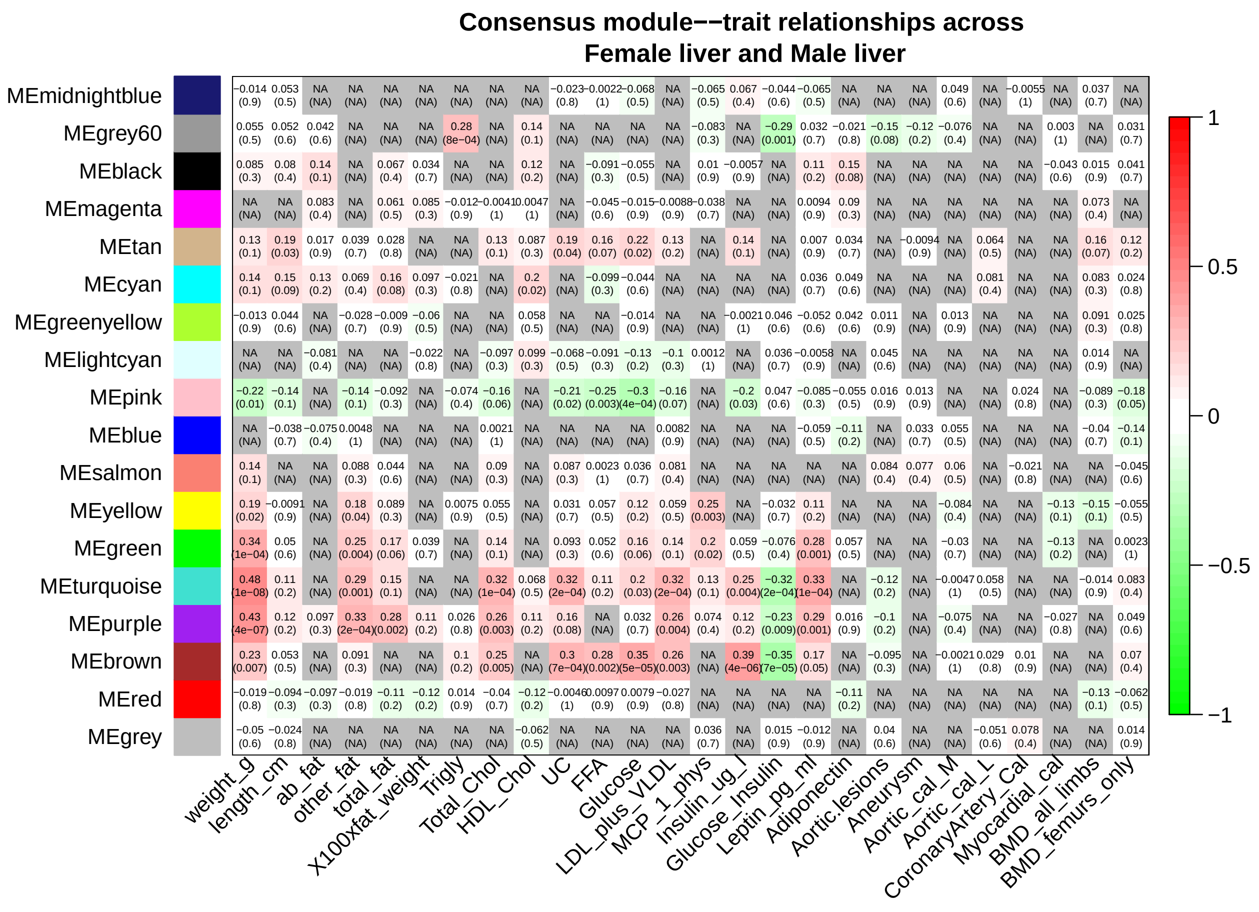


Figure 2: Relationships of consensus module eigengenes and clinical traits in the male data. Each row in the table corresponds to a consensus module, and each column to a trait. Numbers in the table report the correlations of the corresponding module eigengenes and traits, with the p-values printed below the correlations in parentheses. The table is color coded by correlation according to the color legend.



	weight_g	length_cm	ab_fat	other_fat	total_fat	X100xfat_weight	Trigly	Total_Chol	HDL_Chol	UC	FFA	Glucose	LDL_plus_VLDL	MCP_1_phys	Insulin_ug_l	Glucose_Insulin	Leptin_pg_ml	Adiponectin	Aortic.lesions	Aneurysm	Aortic_cal_M	Aortic_cal_L	CoronaryArtery_Cal	Myocardial_cal	BMD_all_limbs	BMD_femurs_only
MEmidnightblue	−0.014 (0.9)	0.053 (0.5)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	−0.023 (0.8)	−0.0022 (1)	−0.068 (0.5)	NA (NA)	−0.065 (0.5)	0.067 (0.4)	−0.044 (0.6)	−0.065 (0.5)	NA (NA)	NA (NA)	NA (NA)	0.049 (0.6)	NA (NA)	−0.0055 (1)	NA (NA)	0.037 (0.7)	NA (NA)
MEgrey60	0.055 (0.5)	0.052 (0.6)	0.042 (0.6)	NA (NA)	NA (NA)	NA (NA)	0.28 (8e−04)	NA (NA)	0.14 (0.1)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	−0.083 (0.3)	NA (NA)	−0.29 (0.001)	0.032 (0.7)	−0.021 (0.8)	−0.15 (0.08)	−0.12 (0.2)	−0.076 (0.4)	NA (NA)	NA (NA)	0.003 (1)	NA (NA)	0.031 (0.7)
MEblack	0.085 (0.3)	0.08 (0.4)	0.14 (0.1)	NA (NA)	0.067 (0.4)	0.034 (0.7)	NA (NA)	NA (NA)	0.12 (0.2)	NA (NA)	−0.091 (0.3)	−0.055 (0.5)	NA (NA)	0.01 (0.9)	−0.0057 (0.9)	NA (NA)	0.11 (0.2)	0.15 (0.08)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	−0.043 (0.6)	0.015 (0.9)	0.041 (0.7)
MEmagenta	NA (NA)	NA (NA)	0.083 (0.4)	NA (NA)	0.061 (0.5)	0.085 (0.3)	−0.012 (0.9)	−0.0041 (1)	0.0047 (1)	NA (NA)	−0.045 (0.6)	−0.015 (0.9)	−0.0088 (0.9)	−0.038 (0.7)	NA (NA)	NA (NA)	0.0094 (0.9)	0.09 (0.3)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	0.073 (0.4)	NA (NA)
MEtan	0.13 (0.1)	0.19 (0.03)	0.017 (0.9)	0.039 (0.7)	0.028 (0.8)	NA (NA)	NA (NA)	0.13 (0.1)	0.087 (0.3)	0.19 (0.04)	0.16 (0.07)	0.22 (0.02)	0.13 (0.2)	NA (NA)	0.14 (0.1)	NA (NA)	0.007 (0.9)	0.034 (0.7)	NA (NA)	−0.0094 (0.9)	NA (NA)	0.064 (0.5)	NA (NA)	NA (NA)	0.16 (0.07)	0.12 (0.2)
MEcyan	0.14 (0.1)	0.15 (0.09)	0.13 (0.2)	0.069 (0.4)	0.16 (0.08)	0.097 (0.3)	−0.021 (0.8)	NA (NA)	0.2 (0.02)	NA (NA)	−0.099 (0.3)	−0.044 (0.6)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	0.036 (0.7)	0.049 (0.6)	NA (NA)	NA (NA)	NA (NA)	0.081 (0.4)	NA (NA)	NA (NA)	0.083 (0.3)	0.024 (0.8)
MEgreenyellow	−0.013 (0.9)	0.044 (0.6)	NA (NA)	−0.028 (0.7)	−0.009 (0.9)	−0.06 (0.5)	NA (NA)	NA (NA)	0.058 (0.5)	NA (NA)	NA (NA)	−0.014 (0.9)	NA (NA)	NA (NA)	−0.0021 (1)	0.046 (0.6)	−0.052 (0.6)	0.042 (0.6)	0.011 (0.9)	NA (NA)	0.013 (0.9)	NA (NA)	NA (NA)	NA (NA)	0.091 (0.3)	0.025 (0.8)
MElightcyan	NA (NA)	NA (NA)	−0.081 (0.4)	NA (NA)	NA (NA)	−0.022 (0.8)	NA (NA)	−0.097 (0.3)	0.099 (0.3)	−0.068 (0.5)	−0.091 (0.3)	−0.13 (0.2)	−0.1 (0.3)	0.0012 (1)	NA (NA)	0.036 (0.7)	−0.0058 (0.9)	NA (NA)	0.045 (0.6)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	0.014 (0.9)	NA (NA)
MEpink	−0.22 (0.01)	−0.14 (0.1)	NA (NA)	−0.14 (0.1)	−0.092 (0.3)	NA (NA)	−0.074 (0.4)	−0.16 (0.06)	NA (NA)	−0.21 (0.02)	−0.25 (0.003)	−0.3 (4e−04)	−0.16 (0.07)	NA (NA)	−0.2 (0.03)	0.047 (0.6)	−0.085 (0.3)	−0.055 (0.5)	0.016 (0.9)	0.013 (0.9)	NA (NA)	NA (NA)	0.024 (0.8)	NA (NA)	−0.089 (0.3)	−0.18 (0.05)
MEblue	NA (NA)	−0.038 (0.7)	−0.075 (0.4)	0.0048 (1)	NA (NA)	NA (NA)	NA (NA)	0.0021 (1)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	0.0082 (0.9)	NA (NA)	NA (NA)	NA (NA)	−0.059 (0.5)	−0.11 (0.2)	NA (NA)	0.033 (0.7)	0.055 (0.5)	NA (NA)	NA (NA)	NA (NA)	−0.04 (0.7)	−0.14 (0.1)
MEsalmon	0.14 (0.1)	NA (NA)	NA (NA)	0.088 (0.3)	0.044 (0.6)	NA (NA)	NA (NA)	0.09 (0.3)	NA (NA)	0.087 (0.3)	0.0023 (1)	0.036 (0.7)	0.081 (0.4)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	0.084 (0.4)	0.077 (0.4)	0.06 (0.5)	NA (NA)	−0.021 (0.8)	NA (NA)	NA (NA)	−0.045 (0.6)
MEyellow	0.19 (0.02)	−0.0091 (0.9)	NA (NA)	0.18 (0.04)	0.089 (0.3)	NA (NA)	0.0075 (0.9)	0.055 (0.5)	NA (NA)	0.031 (0.7)	0.057 (0.5)	0.12 (0.2)	0.059 (0.5)	0.25 (0.003)	NA (NA)	−0.032 (0.7)	0.11 (0.2)	NA (NA)	NA (NA)	NA (NA)	−0.084 (0.4)	NA (NA)	NA (NA)	−0.13 (0.1)	−0.15 (0.1)	−0.055 (0.5)
MEgreen	0.34 (1e−04)	0.05 (0.6)	NA (NA)	0.25 (0.004)	0.17 (0.06)	0.039 (0.7)	NA (NA)	0.14 (0.1)	NA (NA)	0.093 (0.3)	0.052 (0.6)	0.16 (0.06)	0.14 (0.1)	0.2 (0.02)	0.059 (0.5)	−0.076 (0.4)	0.28 (0.001)	0.057 (0.5)	NA (NA)	NA (NA)	−0.03 (0.7)	NA (NA)	NA (NA)	−0.13 (0.2)	NA (NA)	0.0023 (1)
MEturquoise	0.48 (1e−08)	0.11 (0.2)	NA (NA)	0.29 (0.001)	0.15 (0.1)	NA (NA)	NA (NA)	0.32 (1e−04)	0.068 (0.5)	0.32 (2e−04)	0.11 (0.2)	0.2 (0.03)	0.32 (2e−04)	0.13 (0.1)	0.25 (0.004)	−0.32 (2e−04)	0.33 (1e−04)	NA (NA)	−0.12 (0.2)	NA (NA)	−0.0047 (1)	0.058 (0.5)	NA (NA)	NA (NA)	−0.014 (0.9)	0.083 (0.4)
MEpurple	0.43 (4e−07)	0.12 (0.2)	0.097 (0.3)	0.33 (2e−04)	0.28 (0.002)	0.11 (0.2)	0.026 (0.8)	0.26 (0.003)	0.11 (0.2)	0.16 (0.08)	NA (NA)	0.032 (0.7)	0.26 (0.004)	0.074 (0.4)	0.12 (0.2)	−0.23 (0.009)	0.29 (0.001)	0.016 (0.9)	−0.1 (0.2)	NA (NA)	−0.075 (0.4)	NA (NA)	NA (NA)	−0.027 (0.8)	NA (NA)	0.049 (0.6)
MEbrown	0.23 (0.007)	0.053 (0.5)	NA (NA)	0.091 (0.3)	NA (NA)	NA (NA)	0.1 (0.2)	0.25 (0.005)	NA (NA)	0.3 (7e−04)	0.28 (0.002)	0.35 (5e−05)	0.26 (0.003)	NA (NA)	0.39 (4e−06)	−0.35 (7e−05)	0.17 (0.05)	NA (NA)	−0.095 (0.3)	NA (NA)	−0.0021 (1)	0.029 (0.8)	0.01 (0.9)	NA (NA)	NA (NA)	0.07 (0.4)
MEred	−0.019 (0.8)	−0.094 (0.3)	−0.097 (0.3)	−0.019 (0.8)	−0.11 (0.2)	−0.12 (0.2)	0.014 (0.9)	−0.04 (0.7)	−0.12 (0.2)	−0.0046 (1)	0.0097 (0.9)	0.0079 (0.9)	−0.027 (0.8)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	−0.11 (0.2)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	−0.13 (0.1)	−0.062 (0.5)
MEgrey	−0.05 (0.6)	−0.024 (0.8)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	−0.062 (0.5)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	0.036 (0.7)	NA (NA)	0.015 (0.9)	−0.012 (0.9)	NA (NA)	0.04 (0.6)	NA (NA)	NA (NA)	−0.051 (0.6)	0.078 (0.4)	NA (NA)	NA (NA)	0.014 (0.9)

Figure 3: Consensus relationships of consensus module eigengenes and clinical traits across the female and male data. Each row in the table corresponds to a consensus module, and each column to a trait. Numbers in the table report the consensus correlations of the corresponding module eigengenes and traits, with the p-values printed below the correlations in parentheses. The table is color coded by correlation according to the color legend. Missing (NA) entries indicate that the correlations in the male and female data sets have opposite signs and no consensus can be formed.

4.a Exporting results of the network analysis

We now put together a data frame that summarizes the results of network analysis, namely the gene significances (GS) and module memberships (also known as kME) of all probes. We start by loading the gene annotation table.

```
file = gzfile(description = "GeneAnnotation.csv.gz");
annot = read.csv(file = file);
# Match probes in the data set to the probe IDs in the annotation file
probes = names(multiExpr[[1]]$data)
probes2annot = match(probes, annot$substanceBXH)
```

We next (re-)calculate the module eigengenes in the "alphabetic" order and calculate the gene significances and module memberships in each data set.

```
consMEs.unord = multiSetMEs(multiExpr, universalColors = moduleLabels, excludeGrey = TRUE)
GS = list();
kME = list();
for (set in 1:nSets)
{
  GS[[set]] = corAndPvalue(multiExpr[[set]]$data, Traits[[set]]$data);
  kME[[set]] = corAndPvalue(multiExpr[[set]]$data, consMEs.unord[[set]]$data);
}
```

We perform a very simple "meta-analysis" by combining the Z scores of correlations from each set to form a meta-Z score and the corresponding p-value.

```
GS.metaZ = (GS[[1]]$Z + GS[[2]]$Z)/sqrt(2);
kME.metaZ = (kME[[1]]$Z + kME[[2]]$Z)/sqrt(2);
GS.metaP = 2*pnorm(abs(GS.metaZ), lower.tail = FALSE);
kME.metaP = 2*pnorm(abs(kME.metaZ), lower.tail = FALSE);
```

Next we form matrices holding the GS and kME. We use a simple re-shaping trick to put the values and the associated p-values and meta-analysis results next to one another.

```
GSmat = rbind(GS[[1]]$cor, GS[[2]]$cor, GS[[1]]$p, GS[[2]]$p, GS.metaZ, GS.metaP);
nTraits = checkSets(Traits)$nGenes
traitNames = colnames(Traits[[1]]$data)
dim(GSmat) = c(nGenes, 6*nTraits)
rownames(GSmat) = probes;
colnames(GSmat) = spaste(
    c("GS.set1.", "GS.set2.", "p.GS.set1.", "p.GS.set2.", "Z.GS.meta.", "p.GS.meta"),
    rep(traitNames, rep(6, nTraits)))
# Same code for kME:
kMEmat = rbind(kME[[1]]$cor, kME[[2]]$cor, kME[[1]]$p, kME[[2]]$p, kME.metaZ, kME.metaP);
MEnames = colnames(consMEs.unord[[1]]$data);
nMEs = checkSets(consMEs.unord)$nGenes
dim(kMEmat) = c(nGenes, 6*nMEs)
rownames(kMEmat) = probes;
colnames(kMEmat) = spaste(
    c("kME.set1.", "kME.set2.", "p.kME.set1.", "p.kME.set2.", "Z.kME.meta.", "p.kME.meta"),
    rep(MEnames, rep(6, nMEs)))
```

Finally we put together the full information data frame and write it into a plain text CSV file that can be read by standard spreadsheet programs. Note that the probes are not sorted in any particular way; many sort orders are possible and we leave it to the reader to either modify the code or to perform the sort in a spreadsheet software.

```
info = data.frame(Probe = probes, GeneSymbol = annot$gene_symbol[probes2annot],
            EntrezID = annot$LocusLinkID[probes2annot],
            ModuleLabel = moduleLabels,
            ModuleColor = labels2colors(moduleLabels),
            GSmat,
            kMEmat);
 write.csv(info, file = "consensusAnalysis-CombinedNetworkResults.csv",
          row.names = FALSE, quote = FALSE);
```