SUPPLEMENTAL METHODS

Detailed methods for t**ranscription factor binding site consensus (TFBC) analysis**

DNA sequence for the promoters of regulated genes (from -1000 to +200) or for ERBS peak centers (-200 to +200 surrounding peak apices reported by MACS) were downloaded from the UCSC genome browser, using the mouse mm9 build. Background promoter sets were constructed by selecting those genes which were expressed (an average of at least 2 reads from all eight RNA-seq libraries for that tissue) but not regulated by E2 in each tissue (absolute log2 fold change E2/V of less than 0.08), yielding 3089 promoters for aorta and 2119 for liver. The background set for ERBS peaks was an equal number of 400 bp regions whose 3' ends were located 2000 bp to the left (closer to the start of the chromosome) of the 5' edge of the peak. Each foreground and background dataset was scanned for matches to 585 Transfac homology matrices (Transfac Matrix Table 10.3, Biobase GmbH) and 129 matrices from the Jaspar Non-redundant database (1) using the Storm program (part of the CREAD package (2)). Storm was run twice, first with a functional depth threshold of 0.85 and second with a p. value threshold of 5e-4. These two thresholding approaches are complementary, with the first being best at finding near-perfect matches to short sequence matrices (requiring 85% of the maximal fit to the matrix) and the second being best at finding highly significant matches to long sequence matrices (requiring that the probability of as good a fit or better be 0.0005).

Matrix matches from both methods were summed, and the significance of enrichment or anti-enrichment of any TFBC matrix was determined using two-tailed binomial tests with "hits" equal to the foreground matrix matches, "tests" equal to the number of separate regions scanned by STORM (equal to [region_length-TFBC_length]*number of regions, e.g. for aorta downregulated promoters and a 10 bp matrix this would be [1200bp-10bp]*202 genes) and "background frequency" being hits/tests for the

background dataset (e.g. the expressed but non-regulated aorta promoters). These raw p.values were then adjusted for multiple testing (across 714 matrices) using the Benjamini Hochberg method. We then selected all matrices showing significant enrichment (adjusted p. <.05) and an enrichment of 1.1-fold over background (for promoters) or 1.4-fold over background (for ERBSes). We used a lower enrichment threshold for promoters because the longer length of these regions (1200 vs 400 bp) causes the same raw number of sites over background to result in a 3-fold lower enrichment. Note also that, the same fold enrichment gives lower p.values for large input datasets relative to small ones. Accordingly, in order to allow relevant comparisons between enrichment p. values for the 33273 liver ERBS peaks versus the 1754 aorta ERBS peaks, we down-sampled the liver hits and tests numbers used for the binomial tests proportionally (times 1754/33273). This did not change the analysis results, since the final list of significantly enriched TFBC groups (at the p. value and enrichment cut-offs described below) was not changed by use of p. values derived from the original versus sub-sampled numbers.

We used STAMP (3) to examine sequence similarities between enriched TFBC matrices. The results indicated that enriched TFBCs cluster into groups containing matrices that are highly homologous to each other (and to which, by matrix homology, at least, any one of several factors might bind). For instance, the enriched single-matrices V$MAZR_01, V$WT1_Q6 and V$CACBINDINGPROTEIN (and many others) are all highly related to SP1 matrices. This suggests that the enrichment of hits to these matrices might only be a side-effect of the enrichment of sites for a single relevant factor within that group (e.g. a functionally relevant enrichment of SP1 sites would result in an apparent enrichment of MAZR_01 sites, simply because the MAZR_01 matrix is related to SP1 matrices). In cases where homology trees did not positively assign a given matrix to a group (e.g. when one matrix grouped apart from several others with the same base name), we used STAMP to identify the top ten closest matches for that matrix, and used this information to refine group assignments. Next, to avoid biasing our results towards the few matrices for a given factor with the highest enrichment, we added to our list all other matrices with the same base name as a significantly-enriched matrix, if any (e.g. if V$SP1_01 was significantly enriched, we also

included V$SP1_Q6, Jaspar$SP1, etc.). If in doing so we found that more than half of the matrices with the same base name displayed anti-enrichment, this was identified as a case of inconsistent enrichment amongst matrices for that factor, and all matrices for that factor were instead removed from the analysis. Next, for all matrices within each homology group we calculated average enrichment and median adjusted p.value for enrichment (assigning a p. value of 1 to any matrix in the group showing anti-enrichment). Finally, homology groups were identified as enriched in Figs. 3A & 4A if at least one tissue/condition gave a median adjusted enrichment p.value of <0.05 and average fold enrichment of >=1.2 (for ERBSes) or 1.1 (for promoters).

Since the actual binding sites for a factor should generally fit best to matrices for that factor (as opposed to closely related matrices), we expect that the single TFBC matrix showing the highest fold-enrichment and the lowest p.value within any homology group is likely to be for the truly relevant factor. By this logic, for instance, actual ER (represented by the matrices V$ER_Q6 and V$ER_Q6_02) is likely to be the real factor contributing to ER-group enrichment in ERBSes (rather than PAX2, PPARG or T3R) and ATF and CREB are likely to be the real factors contributing to CREB-group enrichment (rather than E4F1, HBP, WHN or XBP1, supplemental tables 3 & 4).

To produce the graphs in Figs. 3 and 4, hits to TFBC matrices were binned in 50 bp intervals relative to TSSes or to peak centers, and normalized to sites per kb by dividing the number in each 50 bp (0.05kb) bin by 0.05 times the number of promoters or ERBS peaks in each foreground or background dataset. The plots in Fig. 3 show these results for the indicated single representative TFBC matrices. The plots in Fig. 4 show the average values across all TFBC matrices within the indicated homology groups.

REFERENCES FOR SUPPLEMENTAL METHODS

1. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B 2004 JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32:D91-94

2. Schones DE, Smith AD, Zhang MQ 2007 Statistical significance of cis-regulatory modules. BMC Bioinformatics 8:19

3. Mahony S, Auron PE, Benos PV 2007 DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. PLoS Comput Biol 3:e61

LEGENDS FOR SUPPLEMENTAL FIGURES & TABLES

Supplemental Fig. 1. PhastCons plot (1) showing the increase in mammalian evolutionary conservation at the center of ERBS peaks in aorta (red) and liver (green) relative to 2kb offset background sequences for aorta (dashed pink) and liver (dashed brown). This peak of conservation localized to ERBS peak centers is indicative of conserved DNA sequences that may include binding sites for ERα-interacting or -recruiting TFs.

Supplemental Fig. 2. Newick homology tree (from STAMP (2)) showing the relatedness of many of the Transfac and Jaspar matrices that have some homology to EREs. The matrices fall into three rough groupings, HNF4, ER and intermediate. Note that HNF4 subgroup matrices tend to be highly enriched in liver ERBSes, while ER subgroup matrices tend to be highly enriched in aorta ERBSes.

Supplemental Fig. 3. 2nd and 3rd top IPA networks from E2-regulated genes in aorta. E2-downregulated and upregulated genes are marked in green and pink respectively. For other details see the legend for Fig. 5.

Supplemental Fig. 4. 4th and 5th top IPA networks from E2-regulated genes in aorta, as per supplemental Fig. 3.

Supplemental Fig. 5. 2nd and 3rd top IPA networks from E2-regulated genes in liver, as per supplemental Fig. 3.

Supplemental Fig. 6. 4th and 5th top IPA networks from E2-regulated genes in liver, as per supplemental Fig. 3.

Supplemental Fig. 7. Three way Venn diagram showing the low overlaps between ERBSes in aorta and liver (from our ChIP-seq study of ex vivo tissues), and uterus (from (3)).

Supplemental Fig. 8. Average match frequency per kb for matrices of the E2F (A) and SP1 (B) homology groups are plotted versus peak apexes. Red: aorta, blue: liver, dashed rusty: input chromatin background for aorta, dashed grey-blue: input chromatin background for liver.

Supplemental Fig. 9. (A) The log2 RNA-seq values for genes called as at least 1.35-fold differentially regulated by E2 between aorta and liver ([AoE-AoV]-[LiE-LiV]), derived from only the first two libraries (NuGen method, blue), the second two libraries (PolyA+ mRNA purification, red) or all four libraries (purple), was plotted against the log2 differential E2-regulation values from qRT-PCR experiments on independent biological samples. Best fit line equation and R^2 values are reported. (B) Heat map showing differential gene regulation by E2 between all aorta and liver samples. For all genes called as p. 0.05 significantly regulated by E2 in either tissue in the RNA-seq analysis, we computed the log base 2 value for the ratio of E2 reads per million reads over Vehicle reads per million reads for pairwise samples prepared by the same method and on the same day. The clustering relationships between samples (columns) across regulated genes (rows) was graphed using the heatmap() function in R.

Supplemental Table 1. Gene expression data for all genes that were significantly regulated by E2 in aorta (A), liver (B) or differentially regulated by E2 between aorta and liver (C). “log2FC” is the log, base 2, difference between expression with 4hr E2 treatment versus 4hr vehicle treatment (e.g. +1= 2x increase with E2, -1=2x decrease with E2). Genes with an increase +E2 of >=1. 5 (log2 > .585) are marked in green & those with a decrease of >1. 5 (log 2 < -.585) are marked in pink. “p. value” is raw p. value score assigned by EdgeR. Columns 2 & 3 list log2 fold change and p. values for all genes regulated by E2 in Aorta (A), by E2 in Liver (B) or differentially regulated by E2 between Aorta and Liver (C). Additional

columns show the log2 fold change and p. values for these genes in the other tissue, for comparison (as indicated).

Supplemental Table 2. All of the top 500 Functions Annotations (groups of genes related to the indicated biological functions) from Ingenuity Pathway Analysis which had a p. value of <=.05 for association with E2-regulated genes in aorta (A) or liver (B). E2-regulated genes that are members of each Functions Annotations group are listed. "Predicted activation state" indicates whether the direction of gene regulation by E2 for genes in any enriched functions annotation group correspond to an increase or decrease in that function (noted only if the Z score is greater than 2 or less than -2).

Supplemental Table 3. All TFBC matrices from Transfac or Jaspar non-redundant libraries which were significantly enriched in any E2-regulated gene class (AoU, AoD, LiU or LiD) at p<0.05 and fold-enrichment >=1.1, plus any additional matrices of the same base name (marked with grey text), that were included in calculating the average enrichment and median enrichment p. value for the groups (for Fig. 3A). Columns are generally as described in Fig. 3A, except that the specific p.values are given (blue shading: adjusted p. < 1e-4, light blue: adjusted p. < .05). Instances of significant anti enrichment are given in parentheses after the enrichment p. value of 1 (yellow shading).

Supplemental Table 4. All TFBC matrices from Jaspar or Transfac that were significantly enriched in either aorta or liver ERBS peaks at adjusted p. < .05 and fold-enrichment of >=1.4, as well as any additional matrices of the same base name (marked in grey text), which were included in calculating the average enrichment and median enrichment p. value for the groups (for Fig. 4A). Columns are as described for Fig. 4A and Supplemental Table. 3.

Supplemental Table 5. Normalized expression levels (reads per million reads) and differences in expression for transcription factors and signaling molecules of particular interest. Genes that were not

expressed in either tissue (fewer than 2 reads per transcript in more than half of all samples) are not listed. Because this table was designed to indicate expression differences for individual genes between tissues, expression values were not adjusted for transcript length. Quantitative comparison of transcript abundance between genes of different lengths could be accomplished by first dividing each gene by its mRNA transcript length in kb. The log base 2 ratios of average aorta over average liver expression (average[AoE/AoV]/average[LiE/LiV]) are shown, followed by the fold increase seen in either aorta or liver. Fold increases of greater than 2 are marked in pink, for aorta, or green, for liver. Genes are grouped by their association with specific transcription factor complexes or signal transduction pathways, and groups are sorted alphabetically (with all groups for signal transduction molecules related to rapid signaling listed before transcription factors).

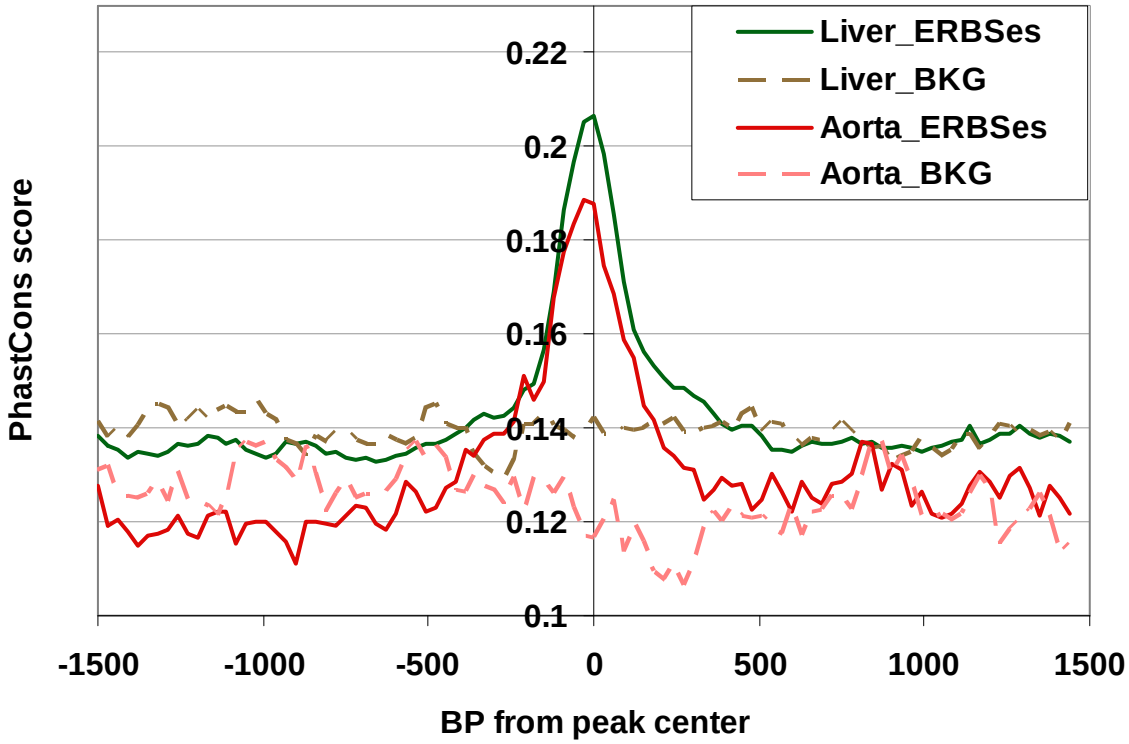
Supplemental Table 6. Primers used for qPCR.

Supplemental Table 7. BED format data for all ERBSes in Aorta (A) or Liver (B). The first line provides track information allowing it to be pasted as a custom track into the UCSC genome browser for visualization.

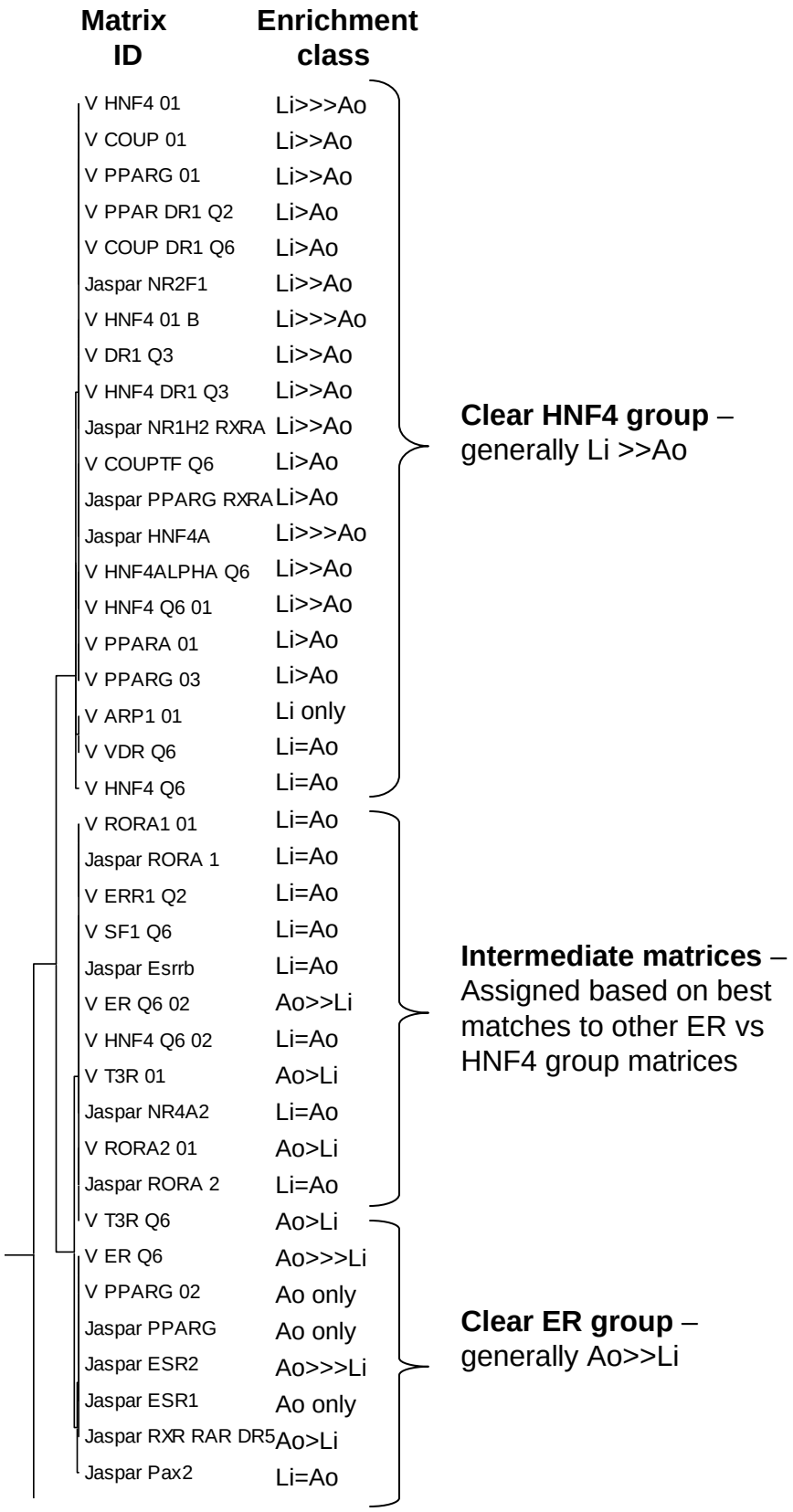
REFERENCES FOR SUPPLEMENTAL FIGURE & TABLE LEGENDS

1. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D 2005 Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034-1050
2. Mahony S, Auron PE, Benos PV 2007 DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. PLoS Comput Biol 3:e61

3. Hewitt SC, Li L, Grimm SA, Chen Y, Liu L, Li Y, Bushel PR, Fargo D, Korach KS 2012 Research resource: whole-genome estrogen receptor alpha binding in mouse uterine tissue revealed by ChIP-seq. Mol Endocrinol 26:887-898

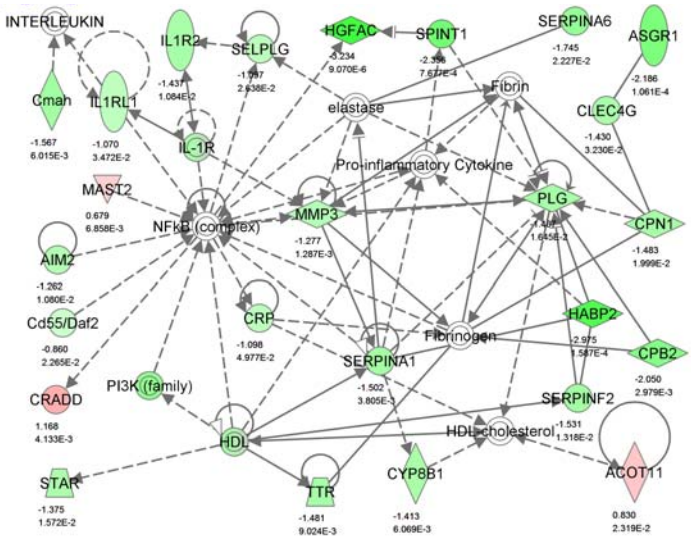


Supplemental Fig. 1

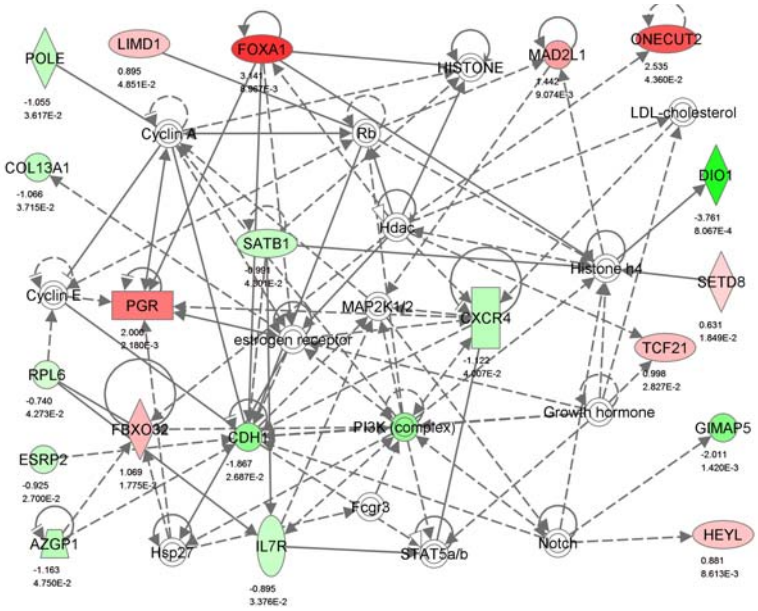


Supplemental Fig. 2

A **Aorta network 2** (23 E2-regl. genes, p. 1e-40)

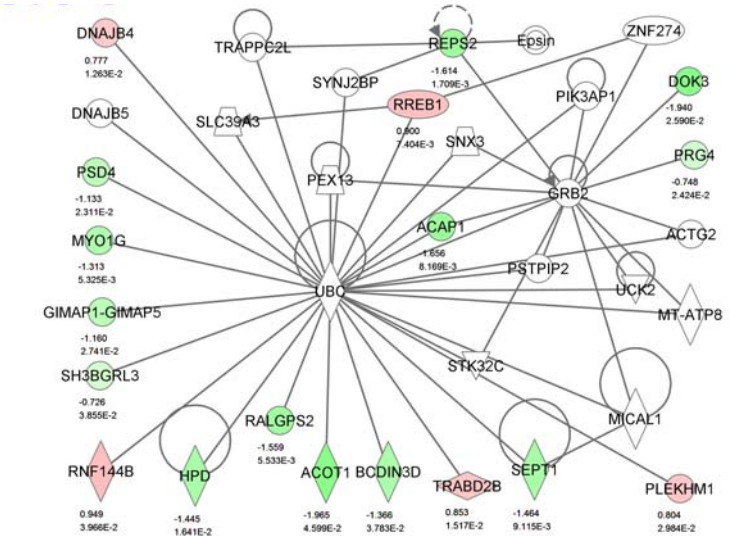


B **Aorta network 3** (20 E2-regl. Genes,p. 1e-31)

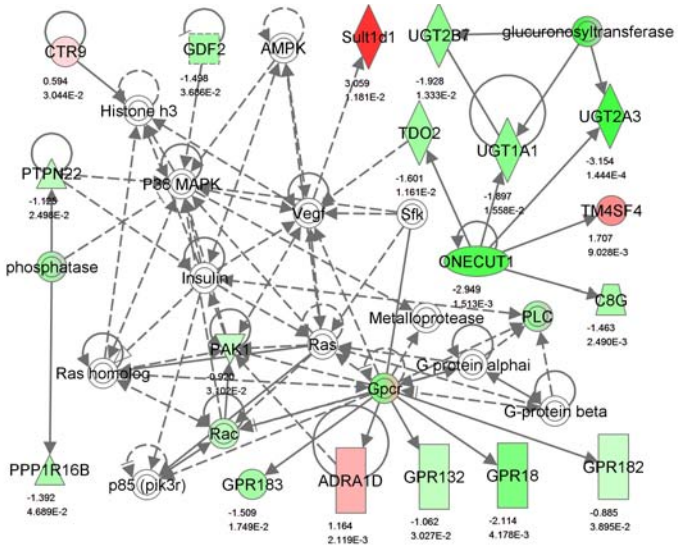


Supplemental Fig. 3

A **Aorta network 4** (18 E2-regl. genes, p. 1e-28)

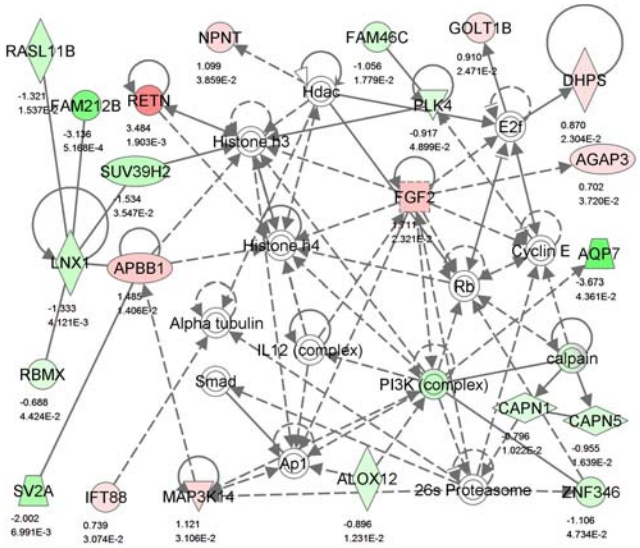


B **Aorta network 5** (18 E2-regl. Genes,p. 1e-28)

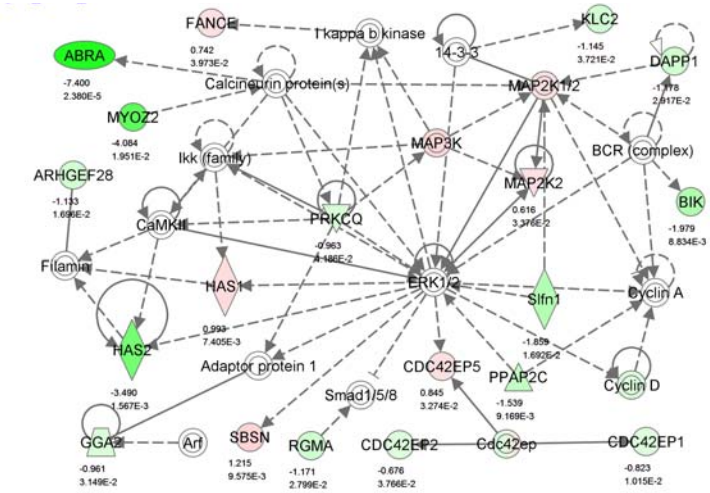


Supplemental Fig. 4

A **Liver network 2** (22 E2-regl. genes, p. 1e-35)

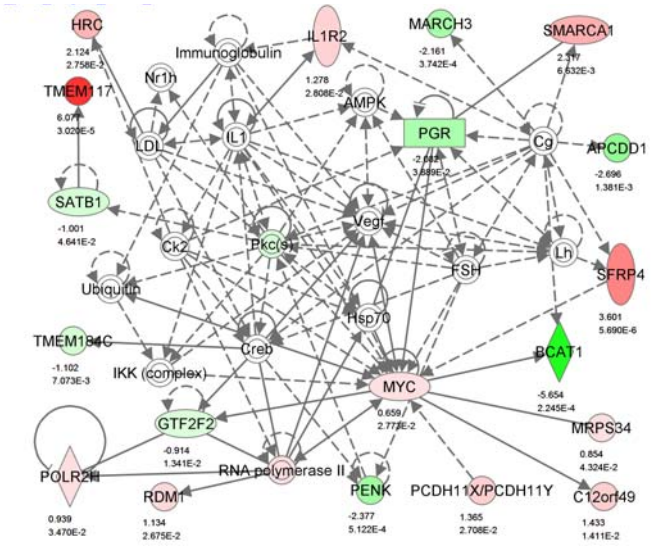


B **Liver network 3** (19 E2-regl. Genes,p. 1e-29)

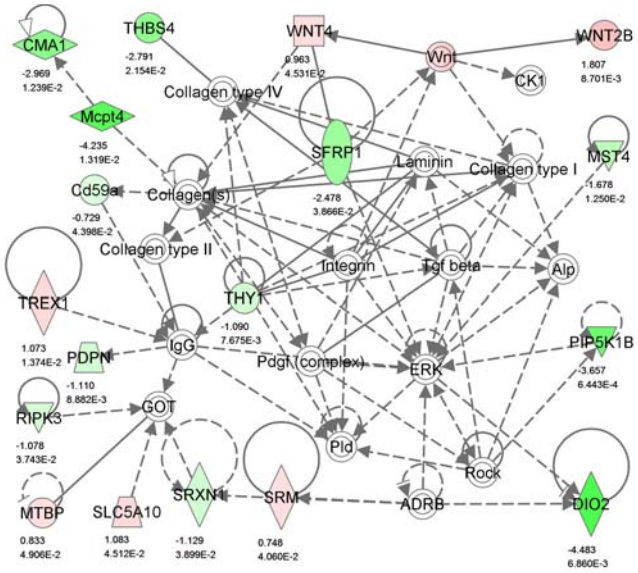


Supplemental Fig. 5

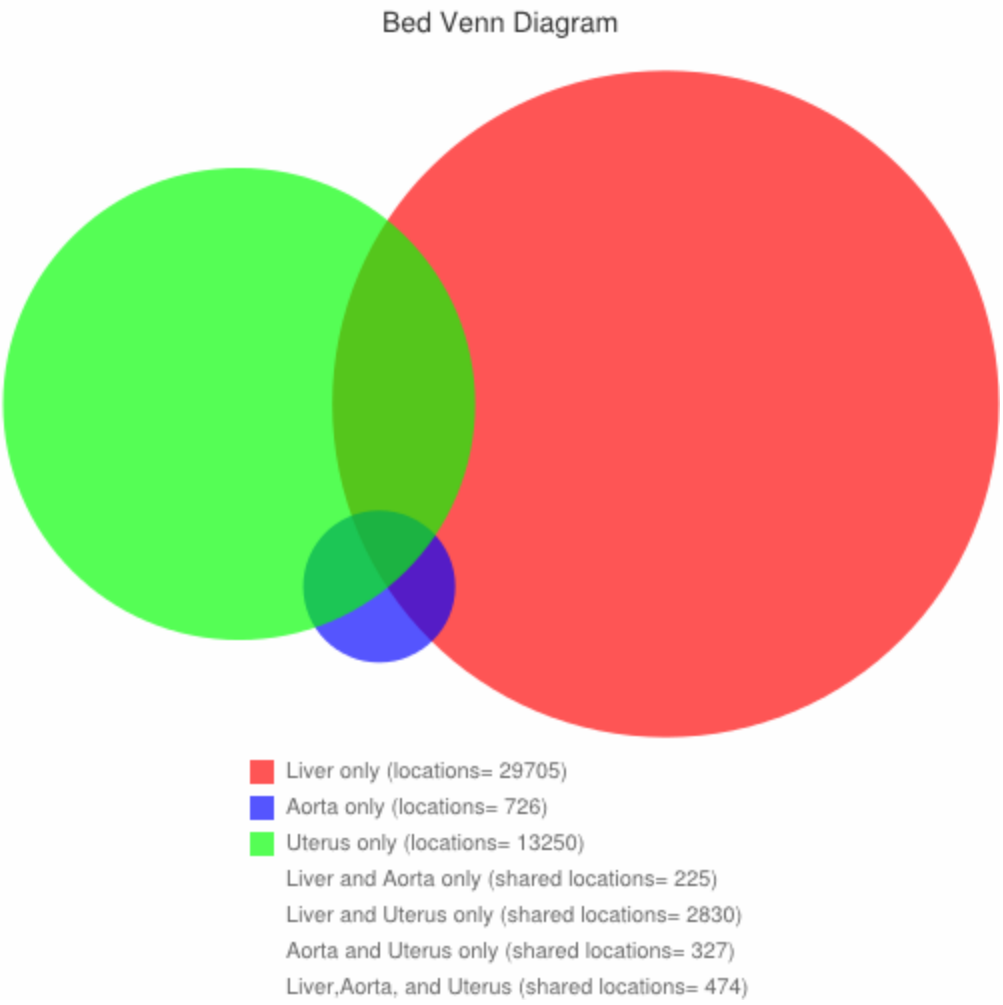
A **Liver network 4** (19 E2-regl. genes, p. 1e-28)



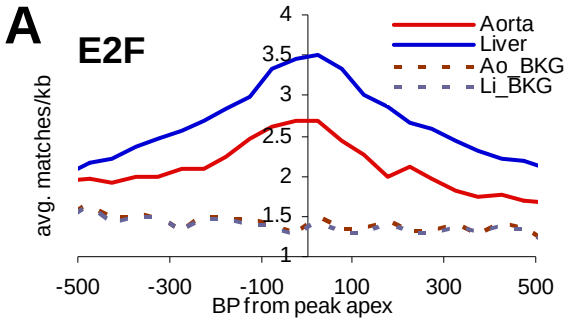
B **Liver network 5** (18 E2-regl. Genes,p. 1e-27)

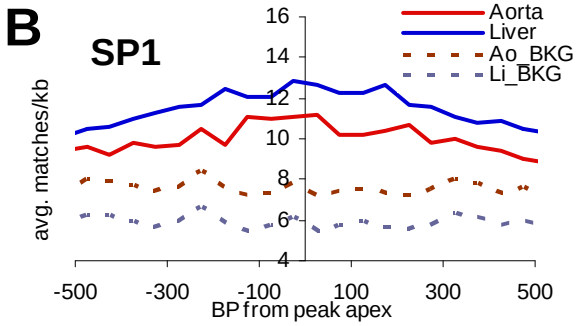


Supplemental Fig. 6

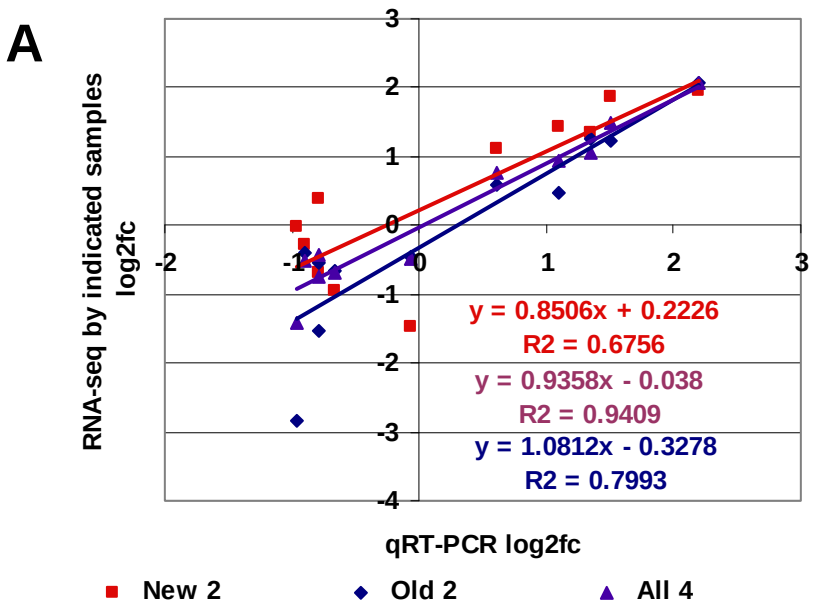


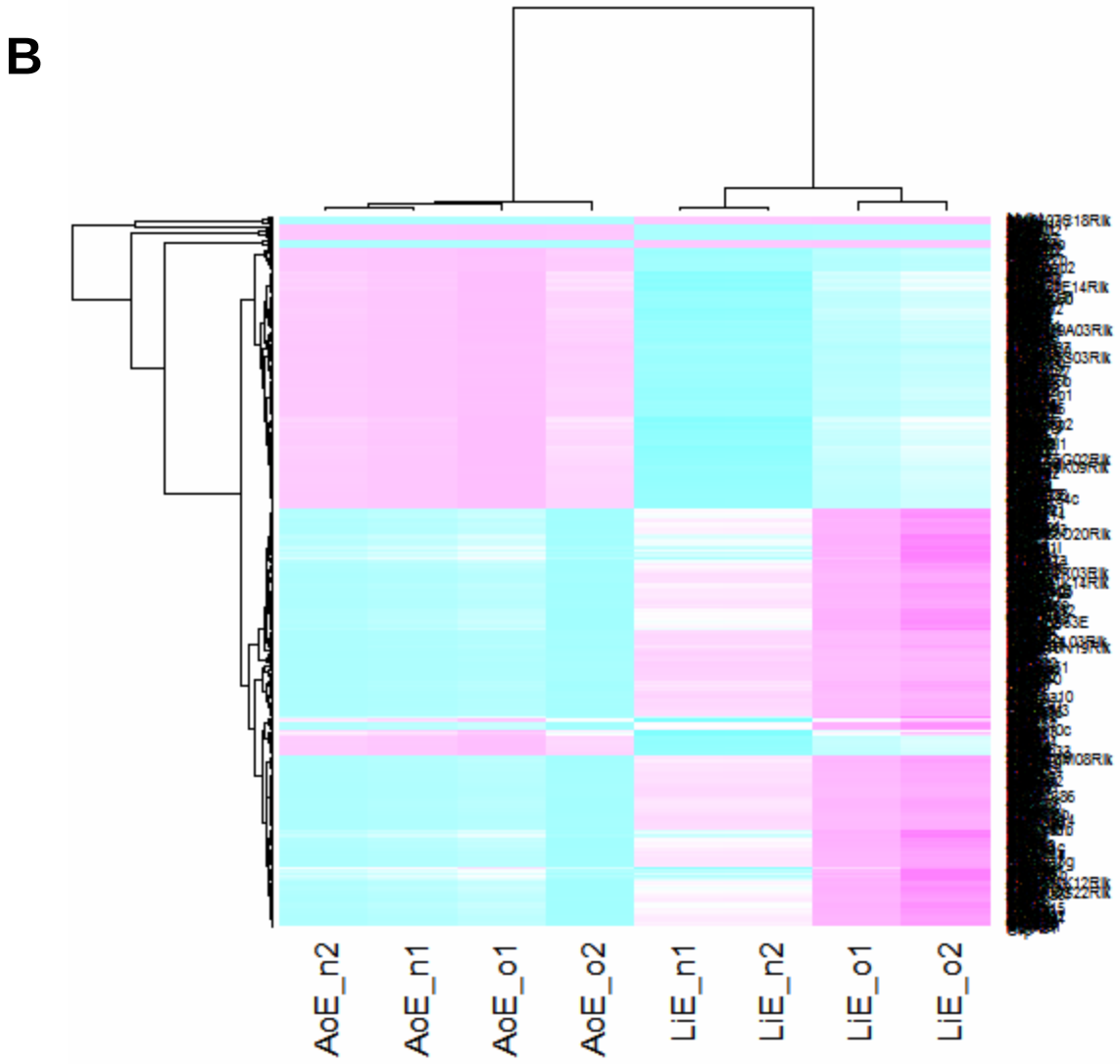
Supplemental Fig. 7





Supplemental Fig. 8





Supplemental Fig. 9