Supplemental Materials

Data sets used for the meta-analysis

In this section we describe in some detail the designs of the data sets we used in our analysis. A summary of the microarray platforms used is in Table S1. Details of our preprocessing methods (quality control and batch correction) are described in the next subsections. Details on the originally reported differential expression results for each study are also given in a later section.

Downing – The goal of the study described in (Downing et al. 2012) was to elucidate genetic pathways involved in susceptibility to alcohol teratogenesis by examining global changes in gene expression in embryos and placentae from two mouse strains, C57BL/6J, DBA/2J, and their first generation reciprocal crosses. C57BL/6J embryos are known to be susceptible to morphological malformations following prenatal alcohol exposure, while DBA/2J are known to be relatively resistant. Pregnant dams were intragastrically intubated on GD9 with either 5.8 g/kg ethanol (20% w/v) or an isocaloric amount of maltose dextrin; a non-intubated control group was also included. There were five litters per treatment and genotype (B6B6, D2D2, B6D2, D2B6). Dams were sacrificed four hours after intubation and RNA was extracted from whole embryos and placentae. Within a litter, all embryos were pooled for RNA extraction. In total, there were 120 samples (3 treatments x 4 strains x 5 litter x 2 tissues), of which we considered only 60 embryonic samples. Two groups of control samples, 20 intubated with maltose dextrin and 20 non-intubated ones, were not found to have significant differences in terms of gene expression and therefore were considered as one control group in our study. This data set was retrieved with permission from PhenoGen (http://phenogen.ucdenver.edu).

GSE1074 – This study, described in (Green et al. 2007), examined the effect of ethanol exposure in two related mouse strains, C57BL/6J and C57BL/6N, which are known to have different susceptibility to teratogenic effects of alcohol. Pregnant dams received intraperitoneal injection on the 8th day of gestation (GD8); alcohol-exposed animals received 22% absolute alcohol in isotonic saline (2.9 g ethanol per kg body weight), while control animals received saline injection. Subset of animals also received an injection of PK11195, a ligand previously shown to protect mouse embryos from teratogen-induced eye and brain malformations. The four samples treated with PK11195 were excluded from our study.

Treated dams were euthanized three hours after the treatment and the RNA was extracted from microdisected embryonic cranial neural folds (headfolds). The same 12 samples were run on two different microarray platforms, Perkin Elmer and Affymetrix; we used only microarrays generated on the latter. In total, we considered four alcohol-treated and four control samples from this study. Each group had two B6J and two B6N samples.

GSE9545 – In this study (Zhou et al. 2011), teratogenic effects of alcohol was assessed using whole embryonic culture of C57BL/6 mice. Pregnant dams were sacrificed on GD8.25 and the gravid uteri were removed and transformed into whole embryonic cultures. Treatment consisted of adding ethanol (88 mM) to alcohol-exposed embryo cultures for 46 hours, after which RNA was extracted from whole embryos.

Two independent experiments were performed using the same experimental procedure, but different microarray platforms. Experiment 1 (GSE9542.2) was done on 8 samples, four alcohol-treated and four controls, and Experiment 2 (GSE9542.1) used 7 alcohol-treated and 4 control samples. The data set IDs are taken from Gemma (Zoubarev et al. 2012), where the original GEO data set GSE9545 was split in two based on the microarray platform used. These two data sets were analyzed separately in our study.

GSE34305 – This study (Kleiber et al. 2012) examines long-term changes in gene expression patterns in brains of adult mice prenatally exposed to alcohol via maternal preference drinking. Experimental C57BL/6J female mice were given access to 10% ethanol solution and pure water sipping tubes starting from 14 days prior to fertilization until 10 days postpartum. Control dams had access to water only. The ethanol and water consumption was measured daily and only experimental dams that consumed ethanol above certain threshold were considered. RNA was extracted from whole brains of adult male offspring at PD70. Two independent microarray experiments were conducted using offspring from different ethanol-consuming and control dams. In the first experiment, RNA samples were pooled from three non-littermate mice to reduce litter effect and two biological replicates were performed per treatment group (2:2 samples). In the second experiment, RNA samples were pooled from two non-littermate mice and three biological replicates were performed per treatment group (3:3). We analyzed the microarrays from two experiments together, after correcting for batch effects.

GSE34469 – This experiment, described in (Laufer et al. 2013), was conducted by the same group as GSE34305, but this time mice were exposed to two acute doses of alcohol at neurodevelopmental times representing the human first and second gestation trimester equivalent. The group used C57BL/6J for two independent experiments: in the first one, relating to the human first gestational trimester, the dams were given two doses of either ethanol (2.5g/kg of ethanol in saline) or saline injections spaced two hours apart at each GD8 and GD11. RNA was extracted from whole brains of adult male mice at P70. RNA samples from three mice were pooled to reduce litter effects and two biological replicates per treatment group were performed (2:2). In the second experiment (modeling trimester 2), the dams were given two 2-hour spaced injections of either ethanol or saline injections at GD14 and GD16. RNA was extracted in the same way as in experiment 1. We analyzed the microarrays from two experiments together, after correcting for batch effects.

GSE34549 – This data set was generated as a part of the same study as GSE34469 (Laufer et al. 2013) with the focus on the human third gestation trimester equivalent. Since in mice this developmental period occurs postnatally, the mice pups were treated directly with two 2-hour spaced injections of either ethanol (2.5 g/kg) or saline on PD4 and PD7. The adult male mice were sacrificed at postnatal day 70 (PD60 in the GEO record) and RNA was extracted from whole brain tissue. Samples from three male mice from three different litters were pooled as one biological replicate; there are two biological replicates per treatment group (2:2).

GSE23105 – This study (Kaminen-Ahola et al. 2010a) focuses on growth restriction phenotype in mice subjected to prenatal alcohol exposure, but also examines gene expression changes in kidneys of alcohol-exposed animals after birth. The study used a mouse model of PAE based on

maternal ad libitum ingestion of 10% ethanol. Experimental pregnant C57BL/6J dams were given 10% ethanol solution in water from fertilization to GD8.5. RNA was extracted from kidneys of six control males and from six ethanol-exposed males at PD28.

GSE23106 – This study (Kaminen-Ahola et al. 2010b) was conducted by the same group and in a similar way as the GSE23105 study described above. While not specified by the authors, it is likely that the samples used in these two studies were extracted from the same animals. As described previously, experimental pregnant C57BL/6J dams were subjected to ad libitum ingestion of 10% ethanol from fertilization to GD8.5. For this study, RNA was extracted from livers of 3 ethanol-exposed and 4 control animals at PD28.

GSE1996 – To our knowledge there is no publication for this study. The GEO record states that the goal of the study was to compare the pattern of gene expression in hippocampus of rats prenatally exposed to alcohol. Three groups of pregnant Sprague-Dawley rat dams were placed on different diet regimens: a liquid diet containing 5% alcohol, an isocaloric liquid diet without alcohol and a lab chow ad libitum diet, but as it was done for Downing data set, we grouped two non-alcohol containing diet groups together. Offspring from each of these groups were allowed to grow until adulthood and placed into one of two training groups: a contextual fear conditioned group and a naive/unhandled group. Hippocampi of adult male rats were dissected at PD100 and used to isolate total RNA from each rat diet and training condition.

Quality Control Procedures

The Affymetrix data sets were first subjected to quality assessment of raw data. For Affymetrix GeneChip arrays, we computed Affymetrix standard quality metrics using Bioconductor package simpleaffy: average background, scale factor, percent present and 3'/5' ratio (Affymetix 2002).

Average background is intended to measure optical background and is typically between 20 and 100 for a good quality array. This value is usually compared across the arrays and should be comparable. We consider an array potentially troubled if its average background is 50% smaller or greater than the average value across all arrays.

The scale factor is median feature intensity on an array and is used by Affymetrix for normalization. Good quality arrays are expected to be within 3-fold of each other; arrays with scale factors outside this range are considered to be poor quality.

Percent present represent percentage of genes called present by Affymetrix detection algorithm and it is expected to be comparable across the arrays within one experiment. We consider an array potentially troubled if its percent present value is 20% smaller or greater than the average value across all arrays.

The 3'/5' ratio servers as a measure of RNA quality and is computed as the expression ratio of 3' and 5' probesets mapped to two internal control genes, β-actin and GAPDH, which are relatively long and ubiquitously expressed. A high 3' to 5' ratio may indicate degraded RNA or inefficient transcription of double-stranded cDNA. For an array of good quality, Affymetrix recommends that the 3'/5' ratio should not exceed 3 for β-actin and 1.25 for GAPDH.

In addition to quality metrics available for Affymetrix GeneChip arrays, we computed two multi-array quality metrics proposed by (Bolstad et al. 2004) and implemented in R package affyPLM (Bolstad et al. 2005). Relative Probe Expression (RLE) values are computed for each probeset by comparing the expression value on each array against the median expression value for that probeset across all arrays. The median RLE values are expected to be close to 0, assuming that most genes are not differentially expressed in a given experiment. A median RLA value that is not centered near 0 or has large interquartile range (IQR) might indicate a poor quality array. We consider an array potentially troubled if absolute value of RLE is greater than 0.05 or IRQ is more than 2-fold greater than the mean IRQ value across arrays.

The second multi-array metrics that we compute is Normalized Unscaled Standard Error (NUSE). This value is computed for each gene and on each array as a ratio of standard error estimate and the median standard error estimate for that gene across all the arrays in the experiment. The median NUSE value for each array is expected to be around one. Arrays are suspected to be of poor quality if either the median NUSE is higher than 1one or they have large IQR. We consider an array potentially troubled if NUSE value is greater than 1.05 or IRQ is more than 2-fold greater than the mean IRQ value across arrays.

All Affymetrix microarrays were scored for the described quality metrics and if an array scored poorly on at least two measures it was considered for exclusion. Based on QC we excluded one sample from Downing data set (Hopkins_E_90_Mouse430_2.CEL) and one sample from GSE1074 data set (GSM136067).

Finally, for every data set included in the meta-analysis we plotted boxplots of log2-intensities and sample correlation heatmaps. The boxplots give a summary of the distribution of probeset (or probe, in case of Illumina platform) expression values and are expected to be comparable across samples. Discordant boxplots may indicate problematic arrays. Sample correlation heatmaps show sample pair-wise correlation coefficients as a color matrix and are another approach for identifying potential outlier samples. For all the arrays sample correlation heatmaps and boxplots were examined for further evidence of outliers.

Batch Effect Correction

Batch effects are systematic non-biological differences between batches of samples in microarray experiments that can be caused by technical or environmental difference during sample extraction, preparation or scanning (Luo et al. 2010). These batch effects can be strong enough to mask or confound true biological differences and need to be removed prior to further data processing and analysis. For data sets used in this study the information about batches was not explicitly stated in the original publications and/or public GEO records, so we relied on the “scan date” of the chips extracted from the CEL files.

Three of the data sets, Downing, GSE34305 and GSE34469, were generated in multiple batches based on the scan dates. The Downing data set contained six different batches, generated during a period of two months, five of which had 12 samples on average, while the last generated batch had only one sample (Hopkins_E_70_Mouse430_2.CEL). Since batch correction cannot be

performed in the case of single-sample batches, this sample was removed from further analysis. Data sets GSE34305 and GSE34469 contained two batches each, generated a year apart, corresponding to two experiments performed in each study. The existence of batch effects was confirmed using the gt function from the R package globaltest (Goeman et al. 2004), which tests for association between batches and gene expression patterns. For all three data sets the globaltest p-value was <0.05 indicating strong associations between batches and levels of gene expression.

We used Combat (Johnson et al. 2007) to correct for the batch effect in these three data sets. After batch correction globaltest p-value was 1 for Downing and GSE34469 and 0.87 for GSE34305.

Using p-value distributions for data set quality control

The distribution of raw p-values for each data set was used for additional quality control. Under the complete null hypothesis of no differential expression, the p-value distribution is expected to be uniform on the interval (0,1). If a data set has differentially expressed genes we would expect to see an excess of small p-values, generating peak on the left (e.g. Downing distribution in Figure S1). A p-value distribution that has an unexpected shape such as a lower than expected number of small p-values (e.g. GSE43324 in Figure S1) is an indicator of some technical artifacts in the data or unexplained correlation among the samples that violates the assumption of independence of the statistical model used (Barton et al. 2013). Such violations are particularly problematic for a meta-analysis based on p-values, as Fisher's test assumes a uniform distribution of p-values under the null. Based on visual inspection and the persistence of poor distributions despite QC efforts, we excluded data sets GSE43324 and GSE1997 from the meta-analysis (Figure S1). Several other data sets were retained despite having minor deviations from the expected distributions (GSE23105, GSE23106 and GSE34469, which all have drops in p-value density close to zero). All the remaining data sets show clear evidence of differential expression (Figure S1).

Meta-analysis

For meta-analysis we used Fisher's combined probability test (Fisher 1928). Fisher's method combines p-values resulting from the individual differential expression analyses into one test statistic (F) in the following way:

$$F = -2\sum_{i=1}^{k} \ln(p_i)$$

where k is the number of data sets being integrated and p_i is the p-value from the i^{th} data set. Under the null hypothesis, the test statistic has a χ^2 distribution with 2k degrees of freedom. P-values for Fisher's test were computed based on this distribution.

In order to take into account the directionality of gene expression changes, we conducted separate meta-analyses for up-regulated and down-regulated genes. For each individual data set, we computed one-sided p-values corresponding to two alternative hypotheses (gene expression does not increase after PAE and gene expression does not decrease after PAE) and used them to

compute F statistics for each direction separately (the computed one-sided p-values for each data set and each direction of change are given at http://www.chibi.ubc.ca/faculty/paul-pavlidis/pavlidis-lab/data-and-supplementary-information/pae/).

Since data sets were generated on different platforms we used gene-level data to allow for cross-platform integration. In the case where a gene had more than one probeset assigned to it, the p-values for the probesets were Bonferroni-corrected and the lowest corrected p-value was used to represent the gene for that data set (thus if a gene had two probesets, the p-values were multiplied by 2, subject to corrected $p \leq 1.0$). The gene-level p-values used in Fisher's method are given at http://www.chibi.ubc.ca/faculty/paul-pavlidis/pavlidis-lab/data-and-supplementary-information/pae/. A gene was analyzed in a meta-analysis if it was present in at least three data sets. This requirement was relaxed to two for meta-analyses containing only four data sets ("prenatal", "chronic").

Functional enrichment analysis

We used ermineJ (Gillis et al. 2010), version 3.0, to run functional enrichment analysis on our results. ErmineJ computes enrichment of gene sets associated with genes in the input gene lists. By default, ermineJ uses Gene Ontology (Ashburner et al. 2000) gene sets from all three domains: cellular component, molecular function and biological process, but it also allows user-imported gene sets, such as Phenocarta (Portales-Casamar et al. 2013) gene set (see below). For GO enrichment analysis of our results we used only biological process ontology. We analyzed meta-results (complete list of analyzed genes with their Fisher's p-values) using precision-recall method in ermineJ. This method uses precision-recall curves to compute gene set enrichment and is mostly concerned with what is happening at the top of the input gene list. We used negative $\log_{10}$ of Fisher's p-values as gene scores, which were then converted to ranks. Only GO gene sets that had between 5 and 400 genes were considered.

Core signature genes were computed using jackknife procedure and they do not have p-values associated with them nor a clear ranking compared to all the other analyzed genes, thus we used ermineJ's over-representation analysis to compute functional enrichment in core signatures.

Pathway analysis using DAVID on-line resource

We also used the Database for Annotation, Visualization and Integrated Discovery (DAVID; (Huang et al. 2009)) v6.7 to compute biological pathway enrichment (based on KEGG pathways). The analyzed genes lists were our meta-signatures and background gene lists were composed of all the genes included in the particular meta-analysis.

Spliceosome pathway was found to be significantly enriched in "all", "acute" and "prenatal" down-regulated meta-signatures. This is in agreement with the results of ermineJ's GO enrichment analysis, where one of the enriched biological processes was RNA splicing. Another, unexpected, KEGG pathway that comes up as being significantly enriched in "acute" and "prenatal" meta-signatures, but also "all" down-regulated core signature is Systemic Lupus Erythematosus. This is mainly due to a number of histone genes that are shared between the lupus pathway and our gene lists.

Literature derived candidate FASD genes

Phenocarta (phenocarta.chibi.ubc.ca; previously known as Neurocarta; (Portales-Casamar et al. 2013)) is a database of gene-phenotype associations specifically focusing on neurodevelopmental diseases. Gene-phenotype associations are amalgamated across multiple resources and literature, making it one of the most comprehensive resources of this kind. We downloaded 129 known FASD candidate genes from Phenocarta (on December 2013). Non- mouse genes were converted to their mouse orthologues using HomoloGene ((NCBI Resource Coordinators 2014); http://www.ncbi.nlm.nih.gov/homologene). These genes were used to look for known FASD candidates among meta- and core signatures and for functional enrichment analysis using ermineJ (Gillis et al. 2010).

Histone genes

We downloaded mouse gene annotation file from NCBI Gene Database (ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/Mammalia/ Mus_musculus.gene_info.gz). We extracted all histone-coding genes using keyword "histone", "family" and "cluster". The nomenclature for replication-dependent histone genes is that their gene symbols are based on which histone cluster they belong to (e.g. Hist1h4a is a histone H4 gene that is found in cluster Hist1) and their full name contains phrase "histone cluster" (Marzluff et al. 2002). Replication-independent histone genes contain "histone family" in their full name. After removing all predicted and pseudo-genes we were left with 83 annotated protein-coding histone genes found in the mouse genome. These genes were used to assess the distribution of histone genes in our meta- and core signatures and to compute functional enrichment using Fisher's exact test (Table S10).

This histone enrichment was only partially detectable from the analysis based on GO, which indicated enrichment in GO categories related to nucleosome organization and chromatin assembly. However, only a subset of histone genes (currently 30 out of 83) is annotated with appropriate GO categories such as "nucleosome assembly" (GO:0006334). Other histone genes are either not annotated at all or are annotated with non-specific, high-level terms.

Network analysis

For computing local network properties of our gene signatures we used mouse and human aggregated PPI networks. Mouse PPIN was constructed using Biological General Repository for Interaction Datasets (BioGRID) (Chatr-Aryamontri et al. 2013), Molecular Interaction Database (MINT) (Chatr-aryamontri et al. 2007) and Database of Interacting Proteins (DIP) (Salwinski et al. 2004). The aggregated mouse PPIN consisted of 2,926 unique genes and 5,333 unique interactions. Since this represents a small subset of mouse genes we also used aggregated human PPIN. The human network was constructed using BioGRID, MINT, DIP, Human Protein Reference Database (HPRD) (Keshava Prasad et al. 2009), InnateDB (Breuer et al. 2013) and iRefIndex (Razick et al. 2008). This network consisted of 14,242 unique genes and 153,175 unique interactions. For the purposes of the network analysis, the two networks were aggregated together (mouse genes were converted to their human homologs using NCBI's resource HomoloGene). The final network consisted of 14,325 unique genes and 160,663 unique

interactions. Using this integrated network, we computed local network properties of our meta- and core signature genes and their statistical significance using permutation test. Permutation distributions of the average shortest path length (using Dijkstra's algorithm), average local clustering coefficient and average node degree were computed using 10,000 random gene sets sampled from the PPI network with similar size and node degree as the analyzed gene list.

Differential expression results in the original studies

Here we describe the results originally reported (if any) by the authors of the studies used, and how they were treated for comparing to our results.

GSE1074 – The original study (Green et al. 2007) reports that there were 2340 differentially expressed probesets based on the analysis of Affymetrix microarrays (the ones that we used in our study). However, neither these probesets nor corresponding genes were available for download, so we were not able to assess the overlap with our results.

GSE9545 – The original publication (Zhou et al. 2011) reports 850 and 2519 significantly differentially expressed probesets in experiments 1 and 2, respectively. These experiments correspond to data sets GSE9545.2 and GSE9545.1 in our study, respectively. The paper reports only 87 overlapping probes that are in the same direction of change (listed in Table 2 of Zhou et al. 2011). We used these probeset IDs to intersect and compare with our results.

Downing – This study (Downing et al. 2012) reported 329 probesets with significant ethanol treatment effects, but after the authors removed the ones with significant interaction between the main factors there were 283 remaining probesets. The paper does not provide a list of these probesets but a list of, presumably, corresponding genes. There are 50 unique genes listed in their Supplementary Table 4 of (Downing et al. 2012) (Ethanol main effect categories and genes). We were not able to analyze two of these genes either due to obsolete gene symbol that could not be confidently resolved (*Fswap*) or changes in genome annotation or probeset mapping criteria leading to the absence of *Prdm10* from the current gene expression platform annotation.

GSE23105 – The paper (Kaminen-Ahola et al. 2010a) reports 148 differentially expressed probes for this data set, 46 up-regulated and 102 down-regulated, albeit none of these was reported as significant at FDR<0.05. These probes are listed in Supplementary Table 3 of (Kaminen-Ahola et al. 2010a). Out of 148, 116 were present in our re-analysis (the rest of them were excluded because they did not map to a unique gene). We used probeset IDs to intersect and compare with our results.

GSE23106 – This study (Kaminen-Ahola et al. 2010b) reported 12 genes to be significantly down-regulated and 3 genes to be significantly up-regulated after ethanol exposure. The genes are listed in Supplementary Table 1 of (Kaminen-Ahola et al. 2010b). Of the 15 genes, we were not able to find an entry for two in NCBI Gene Database (LOC668047 and D14ERTD449E). Additional two genes have been renamed since the paper was published; C730007P19RIK has been changed to Sult2a2 and EG624219 has been changed to Gm6484 (predicted gene). Genes Sult2a2 and Slco1a4 were excluded from differential expression analysis in our study due to

unavailability of uniquely matching probes and low expression values, respectively. This left 11 genes to be directly compared to our results. This data set was later re-analyzed in (Kaminen-Ahola et al. 2010a), where the authors report 167 differentially expressed genes (p-value<0.01) in Supplementary Table 1, 16 of which were reported as significant at FDR<0.05. 36 of these probes were not present in our re-analysis because they did not map to a unique gene. While there was no direct overlap between the reported probes and the significantly differentially expressed probes from our analysis, the AUC for the rank comparison was 0.70 and 0.82 for up-regulated and down-regulated probes, respectively. In Table 2 we report the comparison with the original findings (Kaminen-Ahola et al. 2010b).

GSE34305 – The authors of this study (Kleiber et al. 2012) report genes that were identified in both experiments as significant and show same direction of change. There are 73 up-regulated genes listed in Table 2 of the paper and 90 down-regulated gens listed in Table 3 (Kleiber et al. 2012). Out of these 163 genes, 15 were not analyzed in our study, 12 due to multi-mapping issues (as mentioned in the main article, probesets mapping to multiple genes were excluded from the analysis) and 3 due to changes in genome annotation (they were no longer considered assayed by the GPL6246 platform based on our annotation of the platform).

GSE34469 – This data set was initially described in (Laufer et al. 2013) but no list of differentially expressed genes was provided. The same group more recently published another paper (Kleiber et al. 2013) using the same gene expression data and provided lists of significant probesets/genes in their supplemental table 1, however they did separate analyses for the two experiments (first and second trimester). The published lists of up-regulated and down-regulated probesets for early (Trimester 1; 195 probesets)) and late (Trimester 2; 231 probesets) exposure have very little overlap: there are only 6 down-regulated and none of the up-regulated probesets shared across different treatments. Since we performed our analysis on all of the samples in the data set, disregarding the time difference between the exposures, but batch-corrected for different experiments (based on the scanned date), our results are not directly comparable to the published results. We still provide overlap information in Table 2, done on both early and late lists from the paper.

GSE34549 – This data set was initially described in (Laufer et al. 2013) but no list of differentially expressed genes was provided. The same group recently published another paper using the same gene expression data and provided a list of significant probesets/genes in the supplemental table 1 (Trimester 3). There are 376 probesets (374 unique genes) on the list, 15 up-regulated after ethanol exposure and 361 down-regulated (the paper reports 336 probesets in total, which seems to be a typo). We used probeset IDs to intersect and compare with our results.

GSE1996 – Currently there is no publication for this study and thus we were not able to compare our results with the results from the original study.

References

Affymetix, 2002. GeneChip Expression Analysis Data Analysis Fundamentals.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25 (1), 25–29.

Barton, S. J., Crozier, S. R., Lillycrop, K. A., Godfrey, K. M., and Inskip, H. M., 2013. Correction of unexpected distributions of P values from analysis of whole genome arrays by rectifying violation of statistical assumptions. *BMC Genomics*, 14 (1), 161.

Bolstad, B. M., Collin, F., Brettschneider, J., Simpson, K., Cope, L., Irizarry, R. A., and Speed, T. P., 2005. Quality Assessment of Affymetrix GeneChip Data. *In*: Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A., and Dudoit, S., eds. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* [online]. Springer New York, 33–47. Available from: http://link.springer.com/chapter/10.1007/0-387-29362-0_3 [Accessed 12 Sep 2013].

Bolstad, B. M., Collin, F., Simpson, K. M., Irizarry, R. A., and Speed, T. P., 2004. Experimental design and low-level analysis of microarray data. *International review of neurobiology*, 60, 25–58.

Breuer, K., Foroushani, A. K., Laird, M. R., Chen, C., Sribnaia, A., Lo, R., Winsor, G. L., Hancock, R. E. W., Brinkman, F. S. L., and Lynn, D. J., 2013. InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic acids research*, 41 (Database issue), D1228–1233.

Chatr-Aryamontri, A., Breitkreutz, B.-J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L., Reguly, T., Breitkreutz, A., Sellam, A., Chen, D., Chang, C., Rust, J., Livstone, M., Oughtred, R., Dolinski, K., and Tyers, M., 2013. The BioGRID interaction database: 2013 update. *Nucleic acids research*, 41 (Database issue), D816–823.

Chatr-aryamontri, A., Ceol, A., Palazzi, L. M., Nardelli, G., Schneider, M. V., Castagnoli, L., and Cesareni, G., 2007. MINT: the Molecular INTeraction database. *Nucleic Acids Research*, 35 (Database issue), D572–D574.

Downing, C., Flink, S., Florez-McClure, M. L., Johnson, T. E., Tabakoff, B., and Kechris, K. J., 2012. Gene Expression Changes in C57BL/6J and DBA/2J Mice Following Prenatal Alcohol Exposure. *Alcoholism, clinical and experimental research*, 36 (9), 1519–1529.

Fisher, S. R. A., 1928. *Statistical Methods for Research Workers*. Oliver and Boyd.

Gillis, J., Mistry, M., and Pavlidis, P., 2010. Gene function analysis in complex data sets using ErmineJ. *Nature protocols*, 5 (6), 1148–1159.

Goeman, J. J., Geer, S. A. van de, Kort, F. de, and Houwelingen, H. C. van, 2004. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, 20 (1), 93–99.

Green, M. L., Singh, A. V., Zhang, Y., Nemeth, K. A., Sulik, K. K., and Knudsen, T. B., 2007. Reprogramming of genetic networks during initiation of the Fetal Alcohol Syndrome. *Developmental Dynamics*, 236 (2), 613–631.

Huang, D. W., Sherman, B. T., and Lempicki, R. A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4 (1), 44–57.

Johnson, W. E., Li, C., and Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8 (1), 118–127.

Kaminen-Ahola, N., Ahola, A., Flatscher-Bader, T., Wilkins, S. J., Anderson, G. J., Whitelaw, E., and Chong, S., 2010a. Postnatal growth restriction and gene expression changes in a

mouse model of fetal alcohol syndrome. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 88 (10), 818–826.

Kaminen-Ahola, N., Ahola, A., Maga, M., Mallitt, K.-A., Fahey, P., Cox, T. C., Whitelaw, E., and Chong, S., 2010b. Maternal Ethanol Consumption Alters the Epigenotype and the Phenotype of Offspring in a Mouse Model. *PLoS Genet*, 6 (1), e1000811.

Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D. S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C. J., Kanth, S., Ahmed, M., Kashyap, M. K., Mohmood, R., Ramachandra, Y. L., Krishna, V., Rahiman, B. A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R., and Pandey, A., 2009. Human Protein Reference Database--2009 update. *Nucleic acids research*, 37 (Database issue), D767–772.

Kleiber, M. L., Laufer, B. I., Wright, E., Diehl, E. J., and Singh, S. M., 2012. Long-term alterations to the brain transcriptome in a maternal voluntary consumption model of fetal alcohol spectrum disorders. *Brain Research*, 1458, 18–33.

Kleiber, M. L., Mantha, K., Stringer, R. L., and Singh, S. M., 2013. Neurodevelopmental alcohol exposure elicits long-term changes to gene expression that alter distinct molecular pathways dependent on timing of exposure. *Journal of Neurodevelopmental Disorders*, 5 (1), 6.

Laufer, B. I., Mantha, K., Kleiber, M. L., Diehl, E. J., Addison, S. M. F., and Singh, S. M., 2013. Long-lasting alterations to DNA methylation and ncRNAs could underlie the effects of fetal alcohol exposure in mice. *Disease models & mechanisms*, 6 (4), 977–992.

Luo, J., Schumacher, M., Scherer, A., Sanoudou, D., Megherbi, D., Davison, T., Shi, T., Tong, W., Shi, L., Hong, H., Zhao, C., Elloumi, F., Shi, W., Thomas, R., Lin, S., Tillinghast, G., Liu, G., Zhou, Y., Herman, D., Li, Y., Deng, Y., Fang, H., Bushel, P., Woods, M., and Zhang, J., 2010. A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *The Pharmacogenomics Journal*, 10 (4), 278–291.

Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J., and Maltais, L. J., 2002. The human and mouse replication-dependent histone genes. *Genomics*, 80 (5), 487–498.

NCBI Resource Coordinators, 2014. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*, 42 (Database issue), D7–17.

Petryszak, R., Burdett, T., Fiorelli, B., Fonseca, N. A., Gonzalez-Porta, M., Hastings, E., Huber, W., Jupp, S., Keays, M., Kryvych, N., McMurry, J., Marioni, J. C., Malone, J., Megy, K., Rustici, G., Tang, A. Y., Taubert, J., Williams, E., Mannion, O., Parkinson, H. E., and Brazma, A., 2013. Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Research*, gkt1270.

Portales-Casamar, E., Ch'ng, C., Lui, F., St-Georges, N., Zoubarev, A., Lai, A. Y., Lee, M., Kwok, C., Kwok, W., Tseng, L., and Pavlidis, P., 2013. Neurocarta: aggregating and sharing disease-gene relations for the neurosciences. *BMC Genomics*, 14 (1), 129.

Razick, S., Magklaras, G., and Donaldson, I. M., 2008. iRefIndex: a consolidated protein interaction database with provenance. *BMC bioinformatics*, 9, 405.

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D., 2004. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research*, 32 (suppl 1), D449–D451.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T., 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, 6 (7), e21800.

Zhou, F. C., Zhao, Q., Liu, Y., Goodlett, C. R., Liang, T., McClintick, J. N., Edenberg, H. J., and Li, L., 2011. Alteration of gene expression by alcohol exposure at early neurulation. *BMC Genomics*, 12 (1), 124.

Zoubarev, A., Hamer, K. M., Keshav, K. D., McCarthy, E. L., Santos, J. R. C., Van Rossum, T., McDonald, C., Hall, A., Wan, X., Lim, R., Gillis, J., and Pavlidis, P., 2012. Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data. *Bioinformatics (Oxford, England)*, 28 (17), 2272–2273.

Supplementary Figures

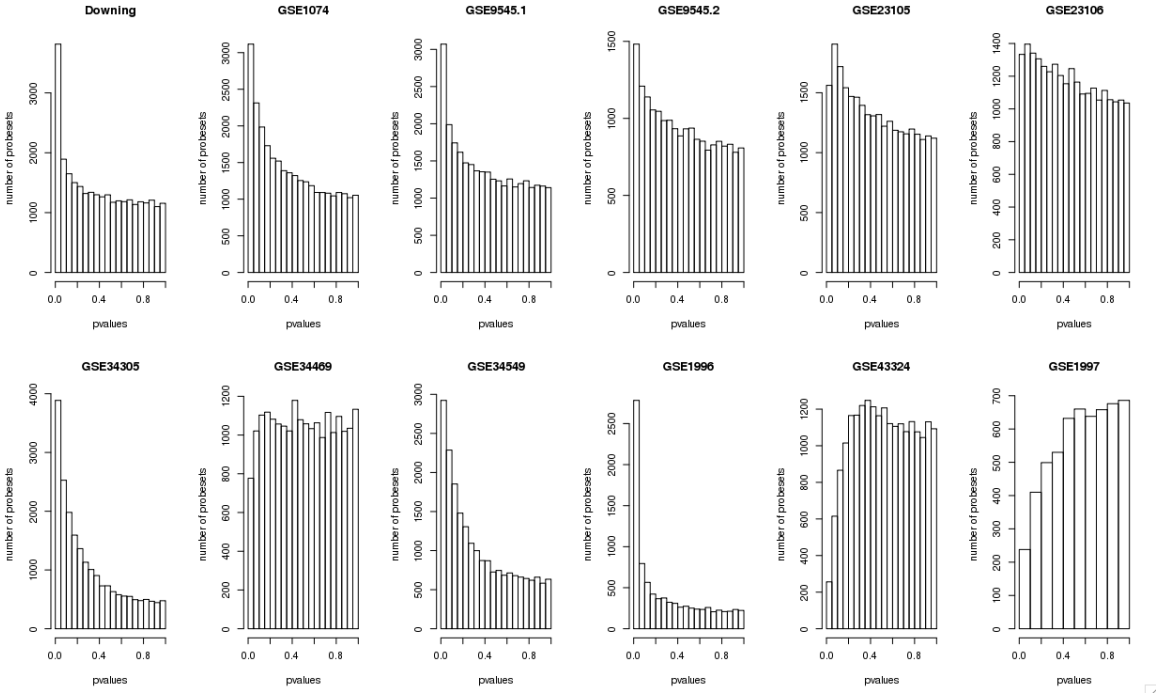


Figure S1: Distributions of p-values for all considered data sets. Data sets GSE43324 and GSE1997 were excluded based on their p-value distributions.

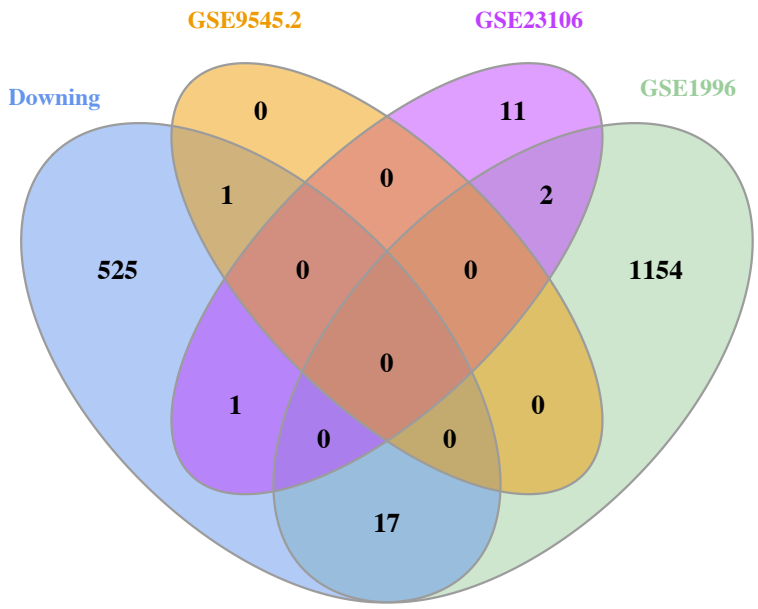


Figure S2: Overlaps of significantly differentially expressed genes obtained from the re-analysis of the individual data sets. Data set GSE9545.1 is not included since it does not have any overlaps with any other data set. The gene lists for each of the above data sets were obtained by considering unique set of genes associated with probesets differentially expressed at FDR<0.05. The direction of change was taken into account.

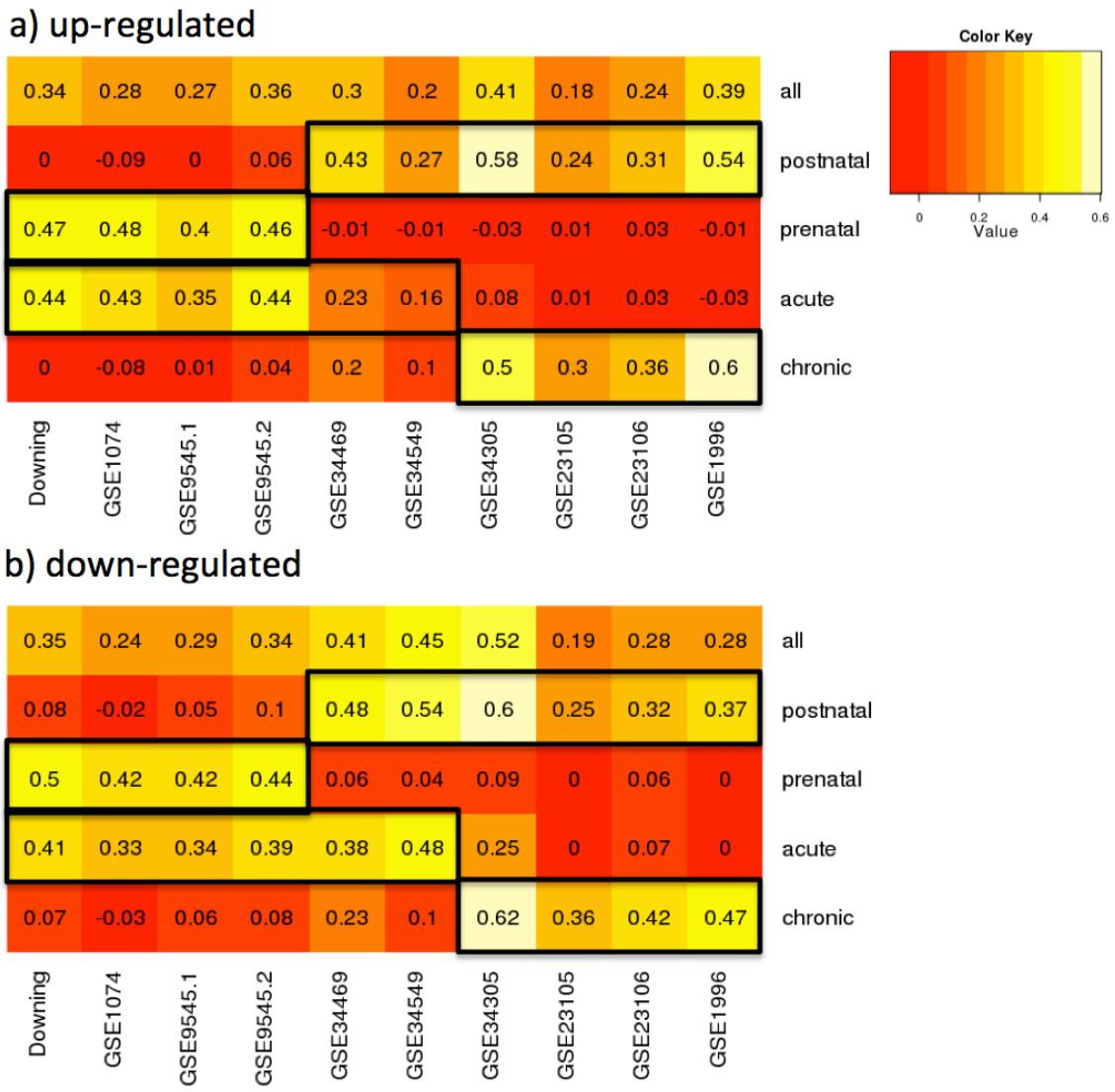
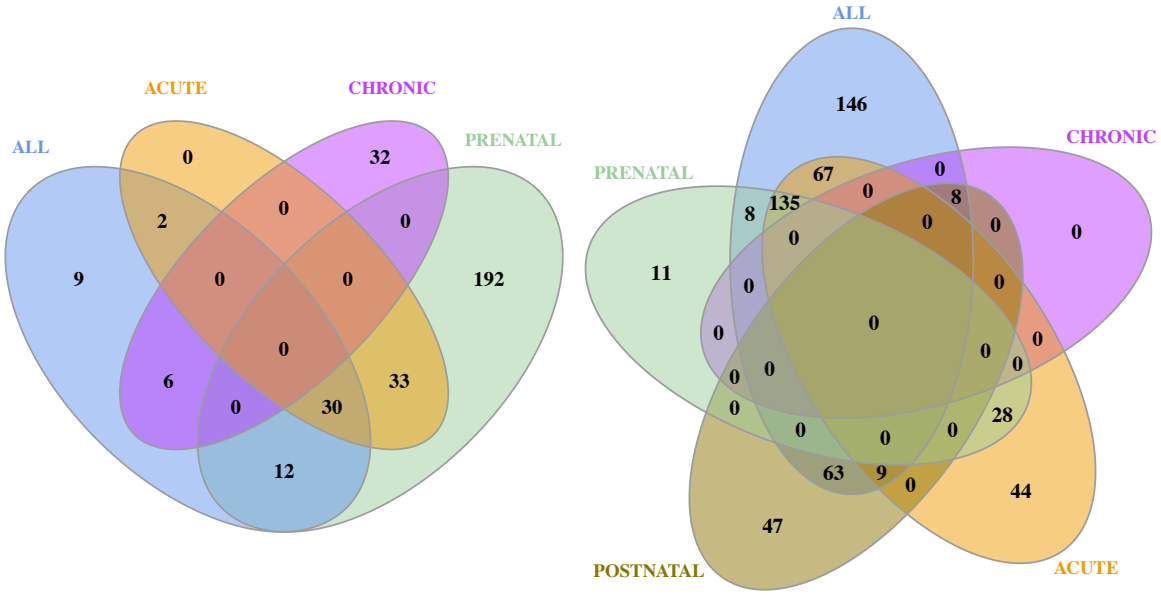


Figure S3: Heatmap/table showing correlation between meta-results and results of DE analysis of the individual data sets. Spearman correlation was calculated using a full list of Fisher's p-values (meta-results) and p-values resulting from the DE analysis of the individual data sets. In general, higher correlation is, as expected, reflected in a data set's inclusion in a particular meta-analysis (data sets included in a particular analysis are grouped and framed together). Positive correlations with each individual data set are also evident in the "all" meta-analysis, suggesting a degree of concordance across all the studies.



UP-REGULATED DOWN-REGULATED

Figure S4: Overlap between up-regulated and down-regulated meta-signature.

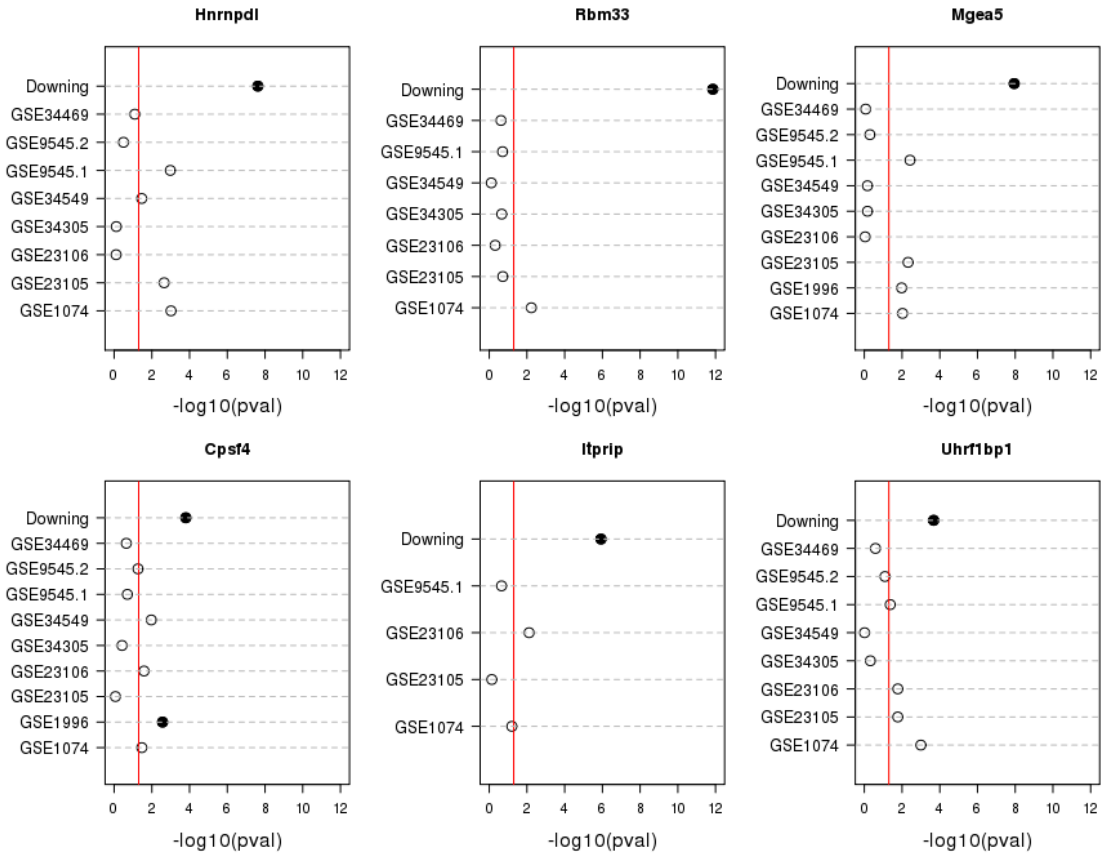


Figure S5: Support for top six up-regulated genes in "all" meta-analysis. Data points represent log-transformed one-sided p-values from the individual differential expression analyses of all data sets. Full circles signify the p-values that remained significant after multiple test correction (FDR<0.05). The red vertical line is added for a reference at a nominal (uncorrected) p-value of 0.05.

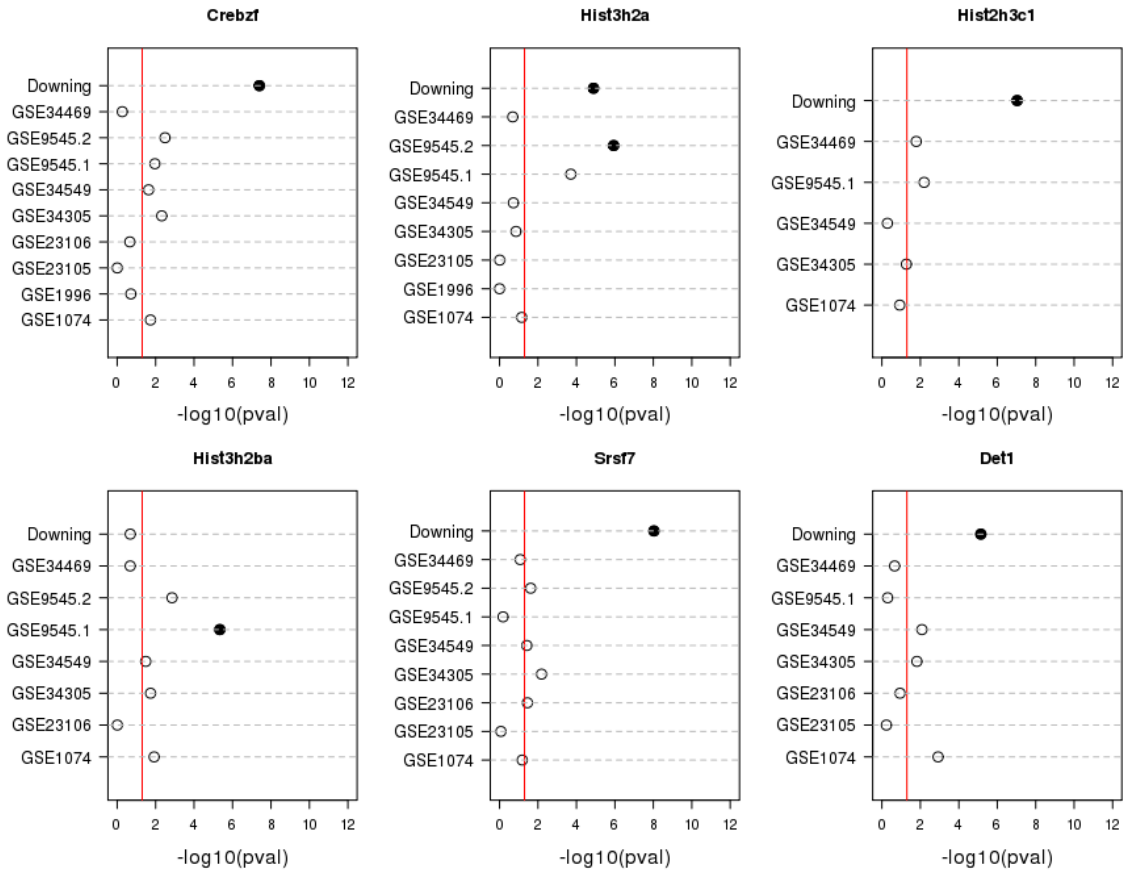


Figure S6: Support for top six down-regulated genes in "all" meta-analysis. See legend to Figure S5 for explanation.

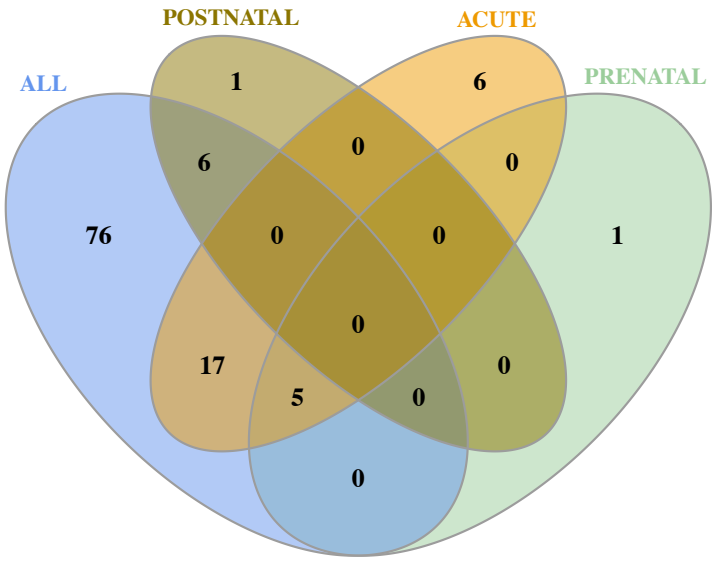


Figure S7: Overlaps among down-regulated core signatures.

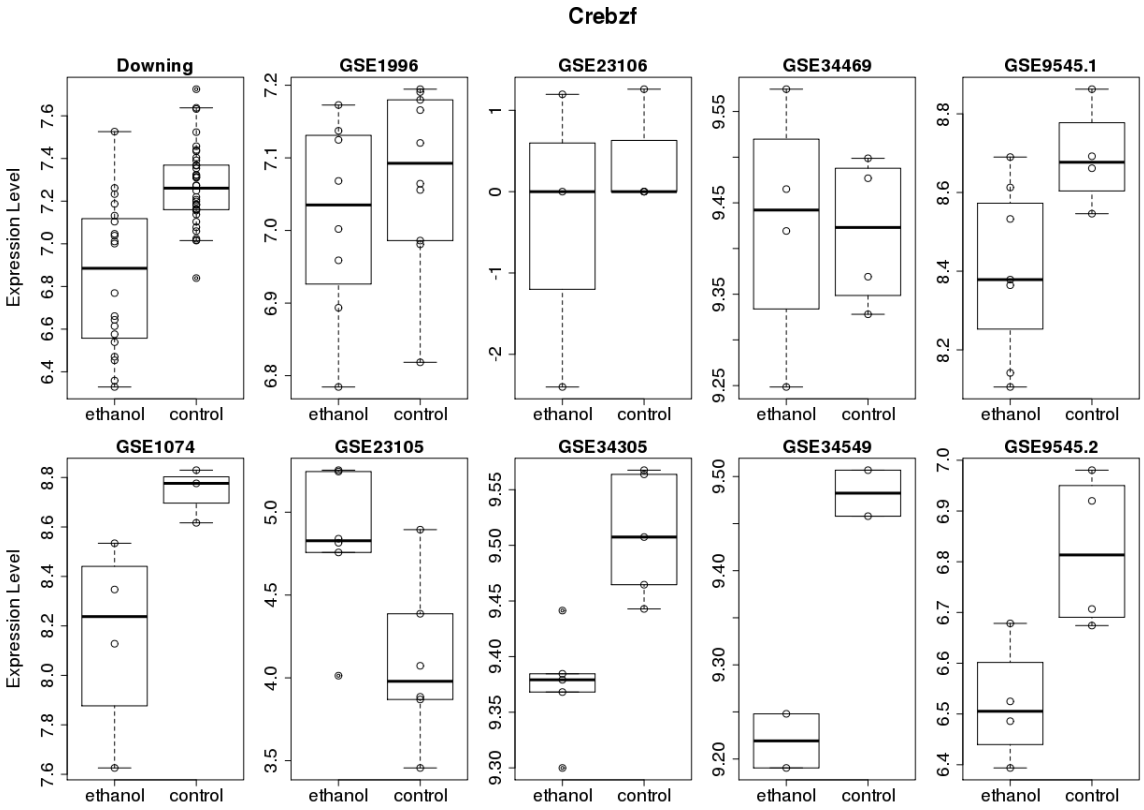


Figure S8: Example of expression pattern of core signature gene – *Crebzf*. Normalized, batch corrected expression values are plotted for each data set. Note the differences in the ordinate of each plot, which reflect expression levels as measured in each data set. Apparent absence of expression in data set GSE23106 agrees with previous observations of low *Crebzf* expression in mouse liver tissue (data from Expression Atlas; (Petryszak et al. 2013)).

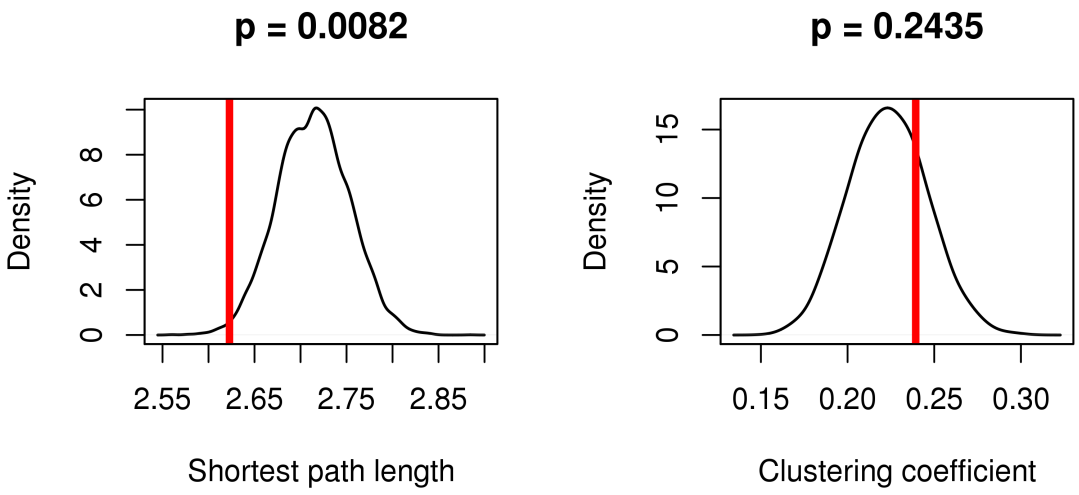


Figure S9: Network analysis - permutation distributions for "all" down-regulated meta-signature. Permutation distributions of the average shortest path length and average local clustering coefficient were computed using 10,000 random gene sets sampled from the PPI network with similar size and node degree as the analyzed gene list. Red vertical line represents obtained values for the examined genes signature list.

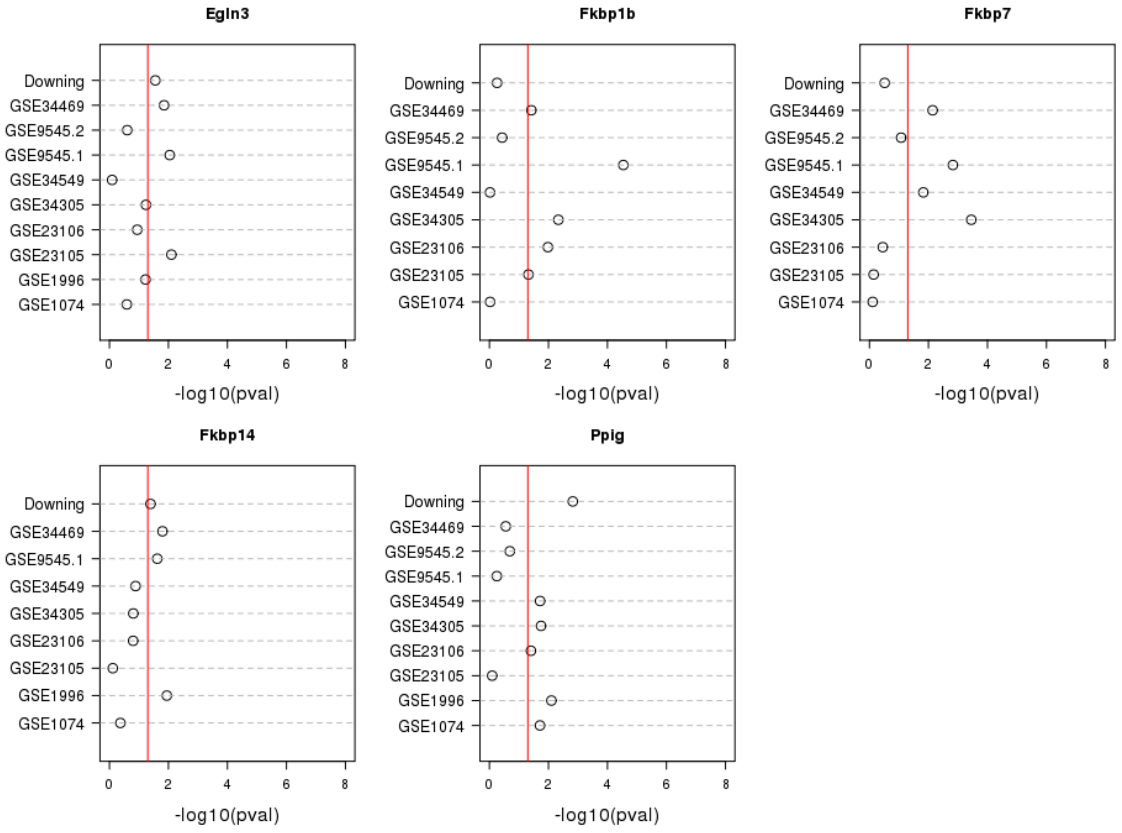


Figure S10: Support for "all" down-regulated core signature genes annotated with peptidyl-proline modification GO term. See legend to Figure S5 for explanation.

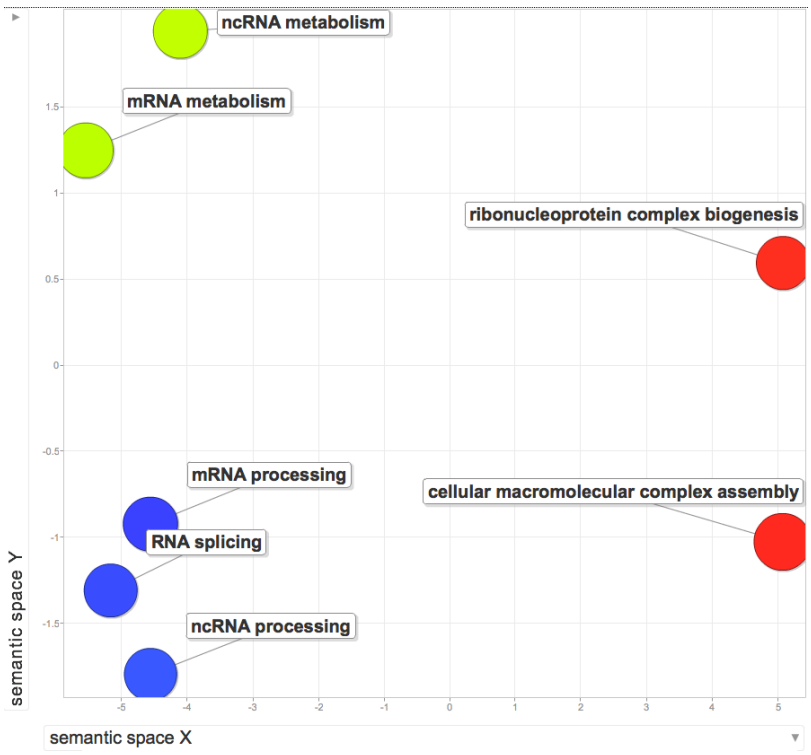


Figure S11: Scatter plot of GO term semantic similarity for statistically significant "all" meta-analysis results. We initially grouped the terms in three general categories based on their closeness in the graph. The graph was produced by REVIGO (Supek et al. 2011), a Web server that summarizes lists of GO terms using a simple clustering algorithm that relies on semantic similarity measures. In addition, it uses multidimensional scaling on a matrix of the GO terms' semantic similarities to plot GO terms in a two-dimensional space grouping more similar terms together. Since the RNA processing is a child term of RNA metabolism, we further group these GO terms into RNA metabolism and macromolecular complex biogenesis.

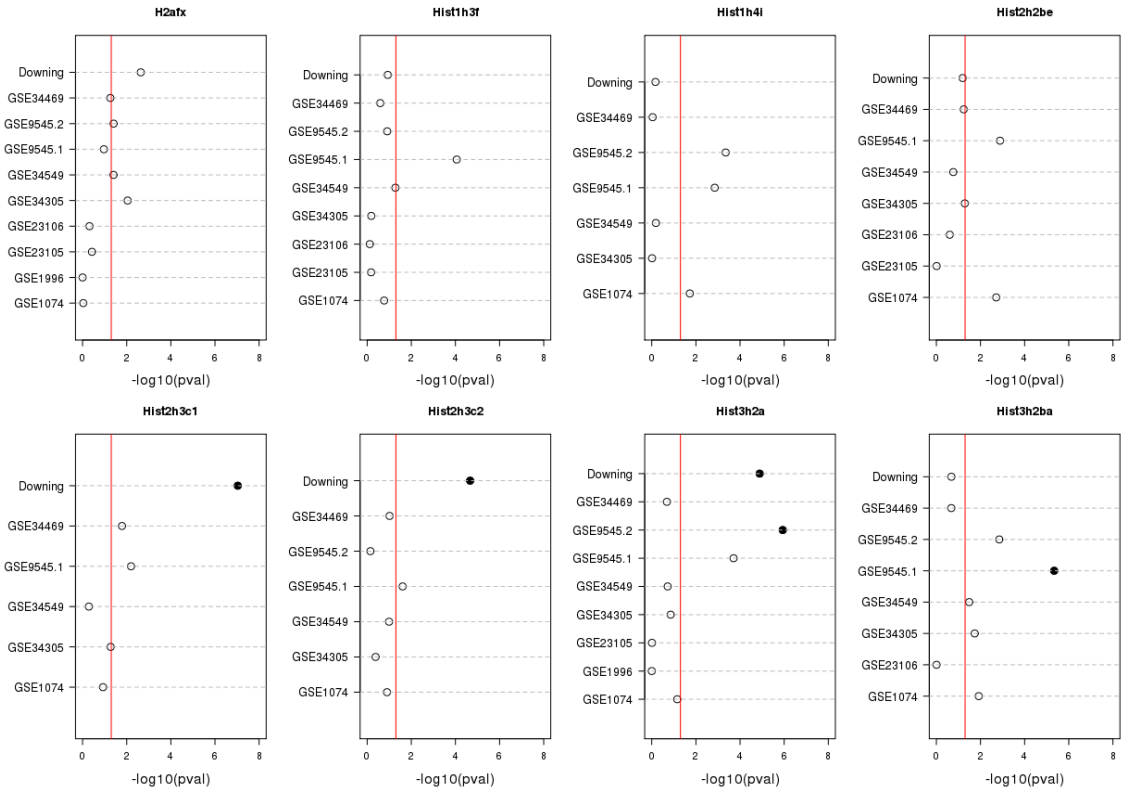


Figure S12: Support for 8 histone-coding genes that are present in meta-signatures. See legend to Figure S5 for explanation.

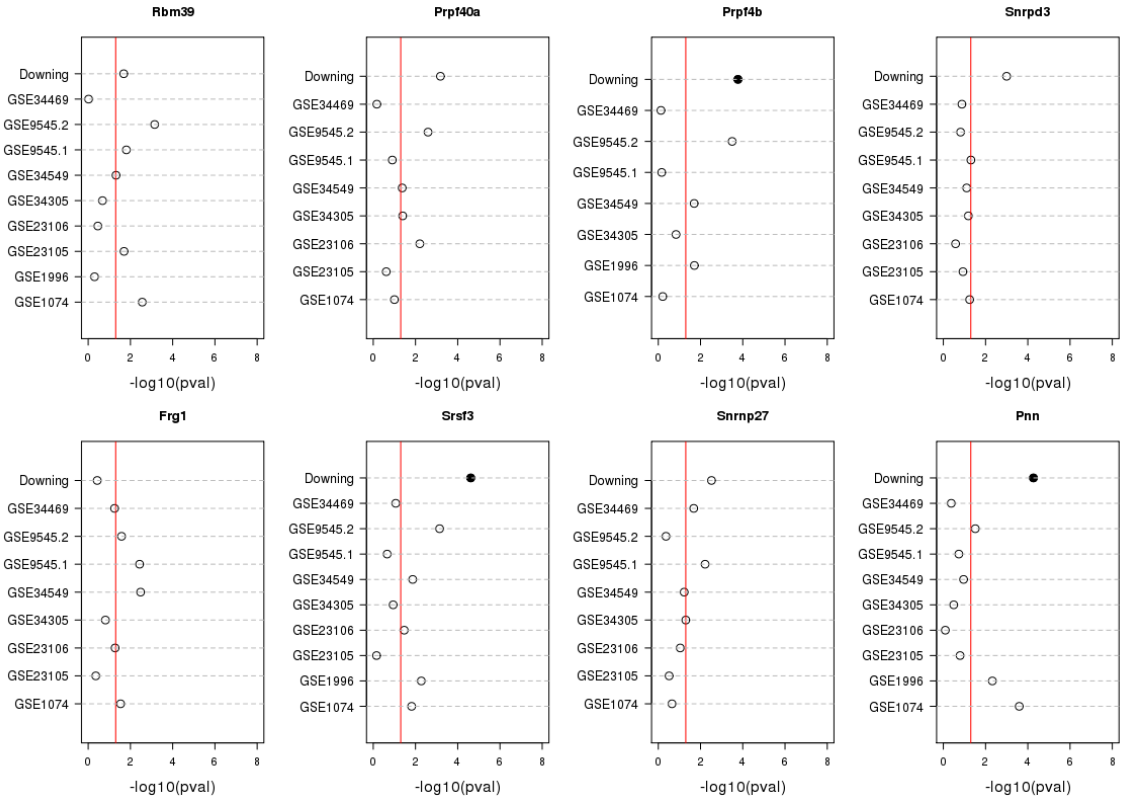


Figure S13: Support for RNA splicing associated genes found in the down-regulated "all" core signature. See legend to Figure S5 for explanation.

Supplementary Tables

GEO Platform ID	Platform Name	# of probesets	# of uniquely mapping probesets	# of unique genes
GPL1261	Affy GeneChip Mouse Genome 430A 2.0	45101	28536	17456
GPL339	Affy GeneChip Mouse Expression Array 430A	22690	19039	12559
GPL6246	Affy Mouse Gene 1.0 ST Array	3557	21045	19443
GPL341	Affy GeneChip Rat Expression Set 230 Array	15923	8778	7717
GPL6887	Illumina MouseWG-6 v2.0 Expression Beadchip	45281	32212	19169

Table S1: Information about microarray platforms used in the gene expression studies. The mapping and gene counts are from annotations obtained via Gemma.

Data set	**Factors**	**# of significant upregulated probesets**	**# of significant downregulated probesets**
Downing	treatment; strain	301	321
GSE1074	treatment; strain	0	0
GSE9545.1	treatment	0	2
GSE9545.2	treatment	0	1
GSE34469	treatment	0	0
GSE34549	treatment	0	0
GSE34305	treatment	0	0
GSE23105	treatment	0	0
GSE23106	treatment	7	7
GSE1996	treatment; strain	935	351

Table S2: Results of differential expression analysis. The numbers shown represent the number of probesets significant at FDR<0.05.

	all		prenatal		postnatal		acute		chronic	
Direction	UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN
# of genes in meta-signatures	59	436	267	182	0	127	65	283	38	8

Table S3: Results of meta-analyses for "all", "prenatal", "postnatal", "acute" and "chronic" data sets. A meta-signature consists of genes that Fisher's test found to be significant at FDR<0.05.

		UP-REGULATED				
		all	**prenatal**	**postnatal**	**acute**	**chronic**
DOWN-REGULATED	**all**		0.68	0.72	0.77	0.65
	prenatal	0.63		-0.02	0.92	-0.02
	postnatal	0.82	0.08		0.16	0.93
	acute	0.84	0.79	0.47		0.04
	chronic	0.67	0.07	0.82	0.17	

Table S4: Correlation among meta-results. Spearman correlation was calculated for each pair of performed meta-analyses (for the same direction of expression) using a full list of Fisher's p-values (meta-results). The values above the diagonal correspond to up-regulated meta-analysis results and the values below the diagonal to down-regulated.

	all		prenatal		postnatal		acute		chronic	
Direction	UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN
# of genes in core signatures	0	104	10	6	0	7	0	28	1	0

Table S5: Results of jackknife procedure for "all", "prenatal", "postnatal", "acute" and "chronic" meta-analyses. The procedure and computation of core signatures are described in Materials and Methods.

	all		prenatal		postnatal		acute		chronic	
Direction	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**
# of genes in meta-signatures	59	436	267	182	0	127	65	283	38	8
overlap with Phenocarta		Gpx1 Pax6 Ntf3 Neurod1 Neurog2	Adh1	Neurog2 Gap43 Mapk1 *		Pax6		Ntf3 Gpx1 Mapk1 Neurog2		Pax6 *

Table S6: Overlap of meta-signatures with known FASD candidate genes from Phenocarta. The asterisk symbolizes statistically significant overlap.

	all		prenatal		postnatal		acute		chronic	
Direction	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**	**UP**	**DOWN**
# of genes in core signatures	0	104	10	6	0	7	0	28	1	0
overlap with Phenocarta		Gpx1						Gpx1 Ntf3 *		

Table S7: Overlap of core signatures with known FASD candidate genes from Phenocarta. The asterisk symbolizes statistically significant overlap.

GO Term ID	GO Term Description	Number of Genes	Corrected p-value
GO:0018208	peptidyl-proline modification	29 (5)	2.14E-03
GO:0033866	nucleoside bisphosphate biosynthetic process	7 (3)	0.015
GO:0018193	peptidyl-amino acid modification	396 (11)	0.015
GO:0008380	RNA splicing	230 (8)	0.046
GO:0033865	nucleoside bisphosphate metabolic process	13 (3)	0.047
GO:0006457	protein folding	123 (6)	0.048

Table S8: Results of functional enrichment analysis for "all" down-regulated core signature. Number of genes given corresponds to the total number of genes annotated with the corresponding GO term; the number in parentheses indicates the number of core-signature genes annotated with the corresponding GO term. The values given for each meta-analysis correspond to corrected p-values from ermineJ analysis (FDR<0.1). The values are displayed only for the significant GO terms.

GO Term ID	GO Term Description	Number of Genes	all	acute	prenatal
GO:0006323	DNA packaging	86		0.042	0.056
GO:0006333	chromatin assembly or disassembly	68		0.026	0.052
GO:0006334	nucleosome assembly	42		0.021	0.035
GO:0006397	mRNA processing	292	5.31E-09	5.25E-09	5.24E-09
GO:0008380	RNA splicing	222	2.65E-09	2.62E-09	2.62E-09
GO:0016071	mRNA metabolic process	341	1.77E-09	1.75E-09	1.75E-09
GO:0016568	chromatin modification	366		0.029	0.030
GO:0022613	ribonucleoprotein complex biogenesis	183	1.33E-09	0.017	0.021
GO:0031497	chromatin assembly	53		0.026	0.042
GO:0034470	ncRNA processing	182	1.06E-09	0.023	
GO:0034622	cellular macromolecular complex assembly	289	8.85E-10	1.31E-09	1.31E-09
GO:0034660	ncRNA metabolic process	240	0.015	0.015	
GO:0034728	nucleosome organization	60		0.045	0.039
GO:0042254	ribosome biogenesis	135		0.039	0.052
GO:0065004	protein-DNA complex assembly	57		0.040	0.035
GO:0071103	DNA conformation change	114		0.056	
GO:0071824	protein-DNA complex subunit organization	75		0.021	0.038

Table S9: Results of functional enrichment analysis for “all”, “acute” and “prenatal” down-regulated meta-analyses’ results. Number of genes given corresponds to the number of genes annotated with the corresponding GO term. The values given for each meta-analysis correspond to corrected p-values from ermineJ analysis (FDR<0.1). The values are displayed only for the significant GO terms.

Gene Symbol	all	acute	prenatal	allCORE	acuteCORE
H2afx	+	+	-	-	-
Hist1h3f	-	+	+	-	-
Hist1h4i	+	+	+	-	-
Hist2h2be	+	+	+	+	+
Hist2h3c1	+	+	+	+	-
Hist2h3c2	+	+	+	-	-
Hist3h2a	+	+	+	+	+
Hist3h2ba	+	+	+	+	+
enrichment p-value	2.37E-05	5.21E-09	6.66E-09	5.65E-05	9.82E-06

Table S10: Presence of histone genes in meta- and core signatures. The first three columns represent meta-signatures and the last two core signatures. '+' indicates a presence of particular histone gene in a signature. The p-values are computed using Fisher's exact test.

	gene signature	number of genes	shortest path	node degree	clustering coefficient
meta-signatures	all UP	51	0.07	0.88	0.30
	all DOWN	356	**0.0003**	0.81	0.16
	prenatal UP	205	**0**	0.63	**0.003**
	prenatal DOWN	148	0.40	0.76	0.47
	postnatal DOWN	98	0.12	0.65	0.08
	acute UP	53	0.06	0.42	0.15
	acute DOWN	232	0.05	0.67	0.64
	chronic UP	30	0.96	0.45	0.59
core	all DOWN	91	**0.008**	0.59	0.24
	acute DOWN	25	**0.04**	0.76	0.11

Table S11: Results of gene network analysis for meta- and core signatures. The first column shows the number of signature genes in the integrated network. The following three columns show permutation test p-values for local network properties. The significant p-values (at significance threshold α=0.05) are bolded.