Supplementary Online Material

COMPREHENSIVE ANALYSIS OF SCHIZOPHRENIA-ASSOCIATED LOCI HIGHLIGHTS BIOLOGICALLY RELEVANT NEURONAL PATHWAYS

Tune H Pers, Pascal Timshel, Stephan Ripke, Samantha Lent, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patrick F. Sullivan, Michael C. O'Donovan, Lude Franke and Joel N Hirschhorn

SUPPLEMENTARY NOTE

Negative control analyses

To ensure that our findings were not driven by biases in DEPICT towards genes with high postnatal expression, we first used DEPICT to prioritize genes for 29 GWAS catalog traits. From the GWAS Catalog www.genome.gov/gwastudies/ (download date: 02 January 2014) we downloaded 61 phenotypes with at least 10 fully independent (DEPICT locus definition) loci (as previously described in ref. [1]. For each phenotype, DEPICT was run using the same parameters as used for the schizophrenia analysis. Phenotypes with at least 10 significantly prioritized (false discovery rate < 0.05) were plotted. Only two traits exhibited a larger t test statistic for prenatal versus/ postnatal expression, namely ulcerative colitis and psoriasis. We found that only two traits comprised prioritized genes with more extreme postnatal- versus prenatal expression levels (Supplementary Figure 5).

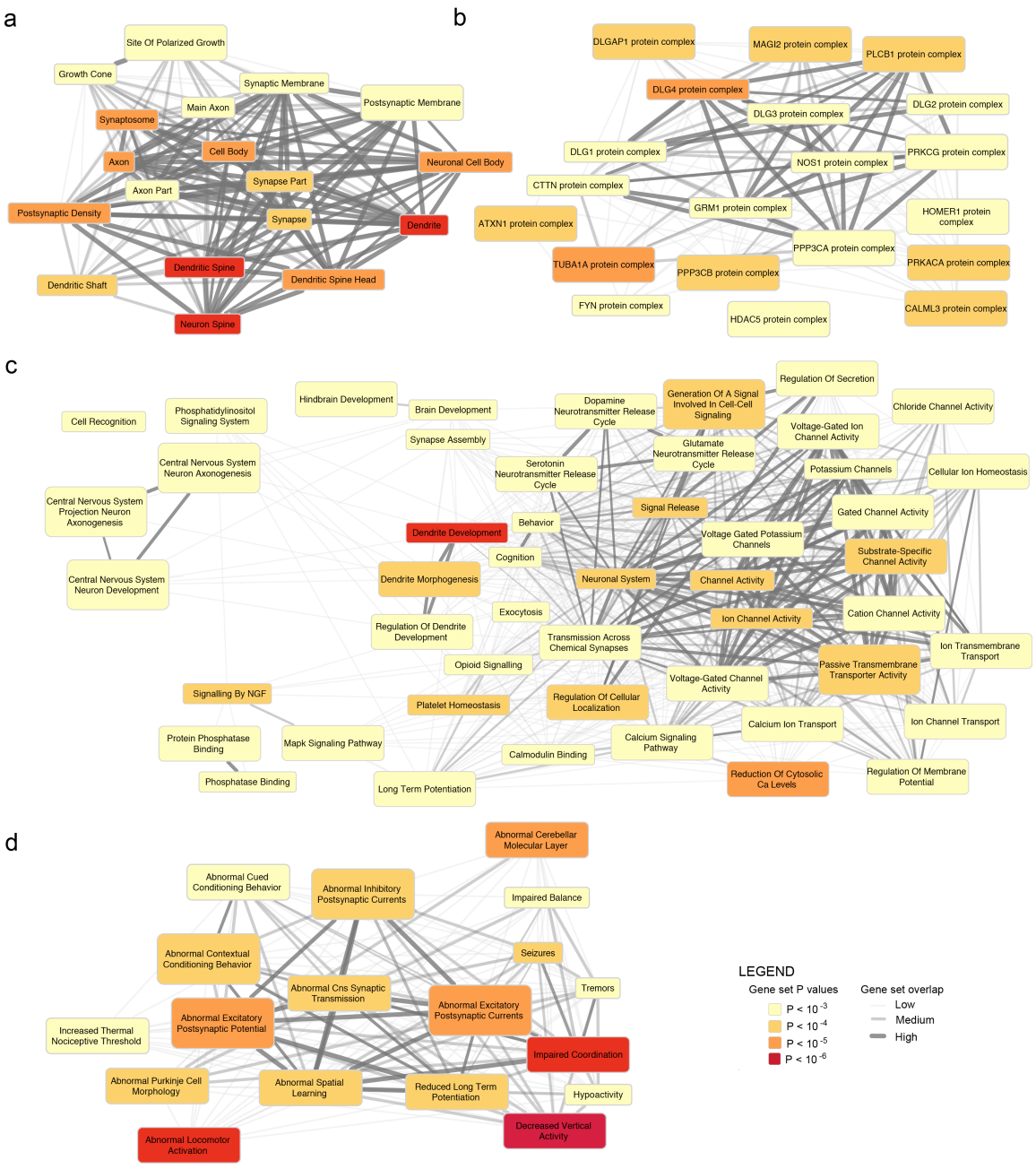
To show that genes prioritized by DEPICT by design do not enrich for genes with high postnatal expression, we ran DEPICT gene prioritization 1,000 times using null associations with no genetic basis and counted the number of null runs that resulted in more significant prenatal versus postnatal expression t-statistics than the one observed for the real data. Towards this end, we performed 1,000 DEPICT analyses based on GWAS with no genetic relationship between SNPs and the phenotypes (referred to as null GWAS in this paper). Briefly, a GWAS was constructed based on genotype data from the Diabetes Genetics Initiative [2] using random phenotypes drawn from a standard normal distribution for all samples. This process was repeated 1,000 times. DEPICT was then run for each null GWAS using on average the top associated, independent (LD $r^2 > 0.5$) 133.2 SNPs (standard deviation 17.8) as input and otherwise the same parameters as used for the schizophrenia SNPs. For each DEPICT run we stored the top 62 prioritized genes (regardless of the false discovery rates, which, as expected, were mostly insignificant) and used them to assess whether DEPICT gene prioritization is biased towards genes with high postnatal expression, that is whether the 62 top genes from the null runs exhibited similar temporal expression as the 62 prioritized schizophrenia genes schizophrenia. Based on the 1000 null GWAS and DEPICT runs, expression levels of genes within null loci and the 62 prioritized genes were summarized the following way:

1. For each null run and stage we calculated the median across all structures. In this way we processed each permutation as we did for schizophrenia prioritized/associated genes and all other gene lists.
2. To plot the average across all permutations (shown as a thick line in Supplementary Figure 6) we calculated the mean of all 1000 permutations. The shaded area corresponds to the standard deviation over all 1000 permutations.

We found that in only 61 instances, top-ranked genes showed more extreme postnatal- versus prenatal expression levels (empirical P=0.061; Supplementary Figure 7).

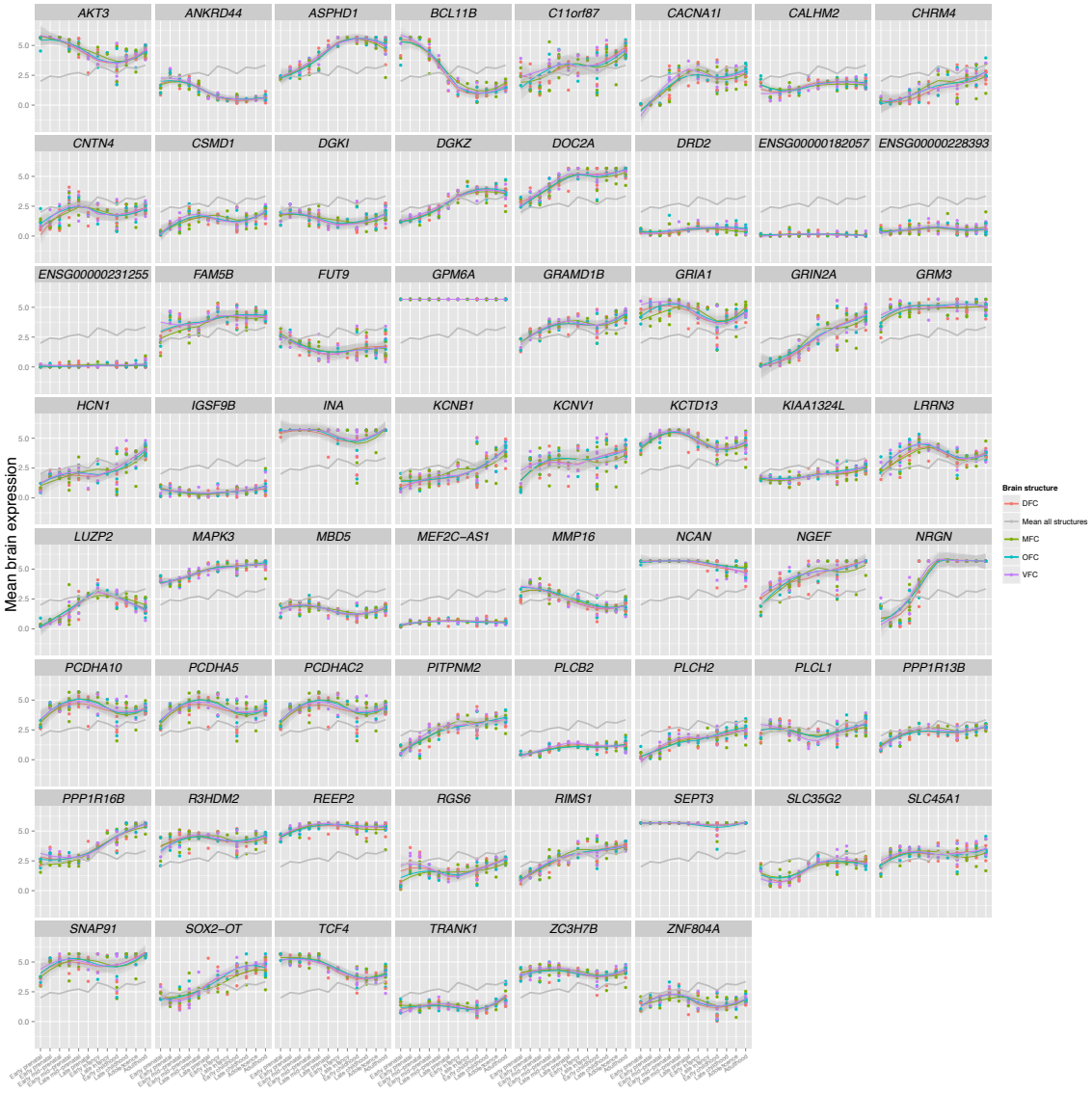
SUPPLEMENTARY FIGURES

Supplementary Figure 1 – Significantly enriched reconstituted gene sets. We report 104 significantly enriched reconstituted gene sets (FDR<0.05, Supplementary Table 3). Please note that DEPICT identified 143 significantly enriched reconstituted gene sets, but that 39 were omitted due to a potential mismatch between the reconstituted gene set identifier and the reconstituted gene set (see Methods and Supplementary Table 9). Reconstituted gene sets are represented by nodes and their overlap is represented by edges. Reconstituted gene sets are colour-coded based on their degree of enrichment in genes from genome-wide significant schizophrenia loci (the darker the more significant). Pairwise overlap between reconstituted gene sets were estimated by computing the Pearson correlation coefficient ρ between two reconstituted gene sets followed by discretization into on of three bins; $0.3 \leq \rho < 0.5$ denotes low overlap, $0.5 \leq \rho < 0.7$ denotes medium overlap, and $\rho \geq 0.7$ denotes high overlap. Edges representing overlap corresponding to $\rho < 0.3$ are not shown. **a**, 18 reconstituted cellular compartment terms from the Gene Ontology database [3] were enriched. **b**, 20 reconstituted protein complexes derived from the InWeb database [4] were enriched. **c**, 48 reconstituted canonical pathways from the KEGG and REACTOME databases [5,6] and biological process and molecular function terms from the Gene Ontology database were enriched. **d**, 18 reconstituted gene sets representing mouse phenotypes from the Mouse Genetics Initiative database [7] were enriched.

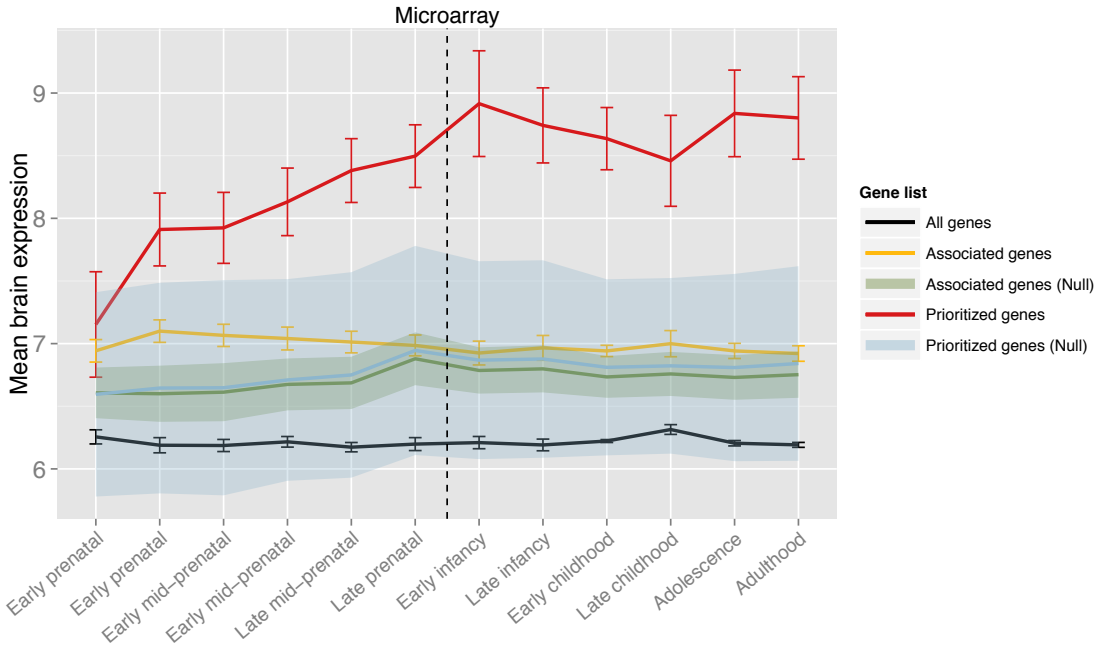

a
Site Of Polarized Growth
Growth Cone
Synaptic Membrane
Postsynaptic Membrane
Main Axon
Synaptosome
Cell Body
Axon
Neuronal Cell Body
Synapse Part
Axon Part
Postsynaptic Density
Synapse
Dendrite
Dendritic Spine
Dendritic Spine Head
Dendritic Shaft
Neuron Spine
b
DLGAP1 protein complex
MAGI2 protein complex
PLCB1 protein complex
DLG4 protein complex
DLG2 protein complex
DLG3 protein complex
DLG1 protein complex
PRKCG protein complex
NOS1 protein complex
CTTN protein complex
GRM1 protein complex
HOMER1 protein complex
ATXN1 protein complex
PPP3CA protein complex
TUBA1A protein complex
PRKACA protein complex
PPP3CB protein complex
FYN protein complex
CALML3 protein complex
HDAC5 protein complex
c
Regulation Of Secretion
Hindbrain Development
Dopamine Neurotransmitter Release Cycle
Generation Of A Signal Involved In Cell-Cell Signaling
Chloride Channel Activity
Cell Recognition
Phosphatidylinositol Signaling System
Brain Development
Voltage-Gated Ion Channel Activity
Synapse Assembly
Glutamate Neurotransmitter Release Cycle
Central Nervous System Neuron Axonogenesis
Serotonin Neurotransmitter Release Cycle
Potassium Channels
Cellular Ion Homeostasis
Central Nervous System Projection Neuron Axonogenesis
Signal Release
Gated Channel Activity
Behavior
Dendrite Development
Voltage Gated Potassium Channels
Substrate-Specific Channel Activity
Cognition
Central Nervous System Neuron Development
Dendrite Morphogenesis
Neuronal System
Channel Activity
Exocytosis
Ion Channel Activity
Cation Channel Activity
Regulation Of Dendrite Development
Transmission Across Chemical Synapses
Ion Transmembrane Transport
Opioid Signalling
Passive Transmembrane Transporter Activity
Voltage-Gated Channel Activity
Signalling By NGF
Regulation Of Cellular Localization
Platelet Homeostasis
Calcium Ion Transport
Ion Channel Transport
Protein Phosphatase Binding
Mapk Signaling Pathway
Calcium Signaling Pathway
Calmodulin Binding
Reduction Of Cytosolic Ca Levels
Regulation Of Membrane Potential
Phosphatase Binding
Long Term Potentiation
d
Abnormal Cerebellar Molecular Layer
Abnormal Cued Conditioning Behavior
Abnormal Inhibitory Postsynaptic Currents
Impaired Balance
Abnormal Contextual Conditioning Behavior
Seizures
Abnormal Cns Synaptic Transmission
Tremors
Abnormal Excitatory Postsynaptic Potential
Abnormal Excitatory Postsynaptic Currents
Increased Thermal Nociceptive Threshold
Impaired Coordination
Abnormal Purkinje Cell Morphology
Abnormal Spatial Learning
Reduced Long Term Potentiation
Hypoactivity
Decreased Vertical Activity
Abnormal Locomotor Activation
LEGEND
Gene set P values
P < 10 -3
P < 10 -4
P < 10 -5
P < 10 -6
Gene set overlap
Low
Medium
High

80

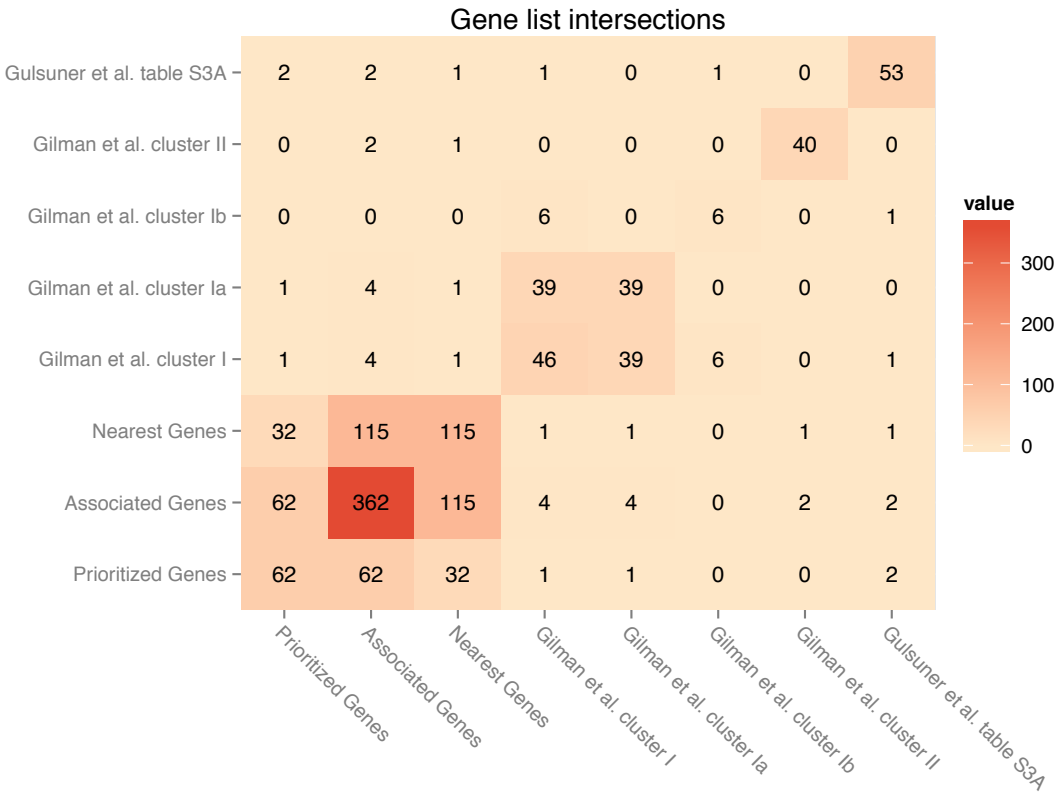
Supplementary Figure 2 – BrainSpan Developing Brain expression trajectories of prioritized genes. The 62 prioritized genes' expression trajectories from early prenatal to adulthood for four prefrontal cortex tissues (dorsolateral prefrontal cortex, DFC; anterior cingulate cortex, MFC; orbital frontal cortex, OFC; ventrolateral frontal cortex, VFC). The average expression of prioritized genes (gray line) and structure-specific expression (colored lines) were Loess smoothed.



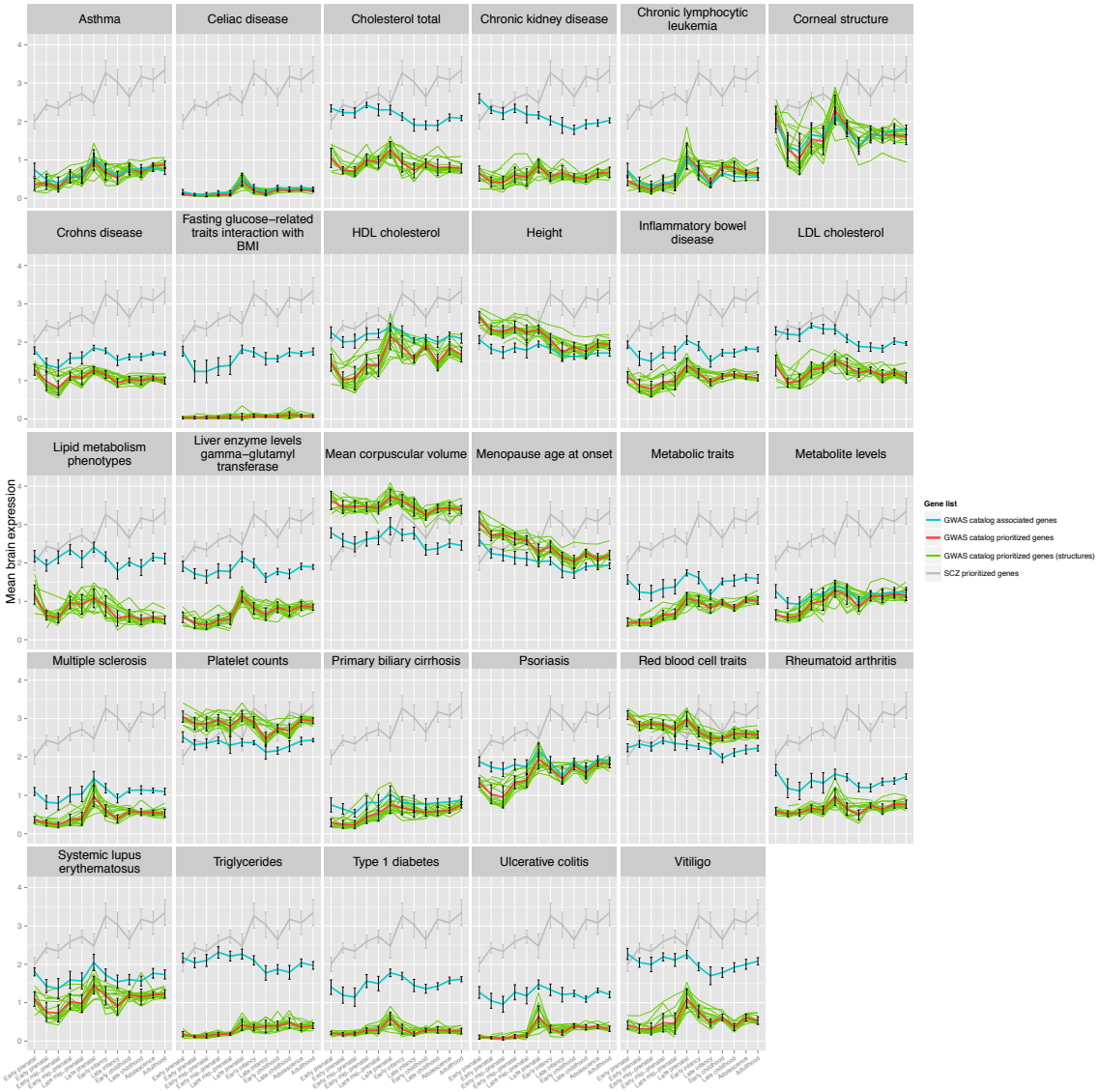
Supplementary Figure 3 – Microarray-based trajectories with gene lists



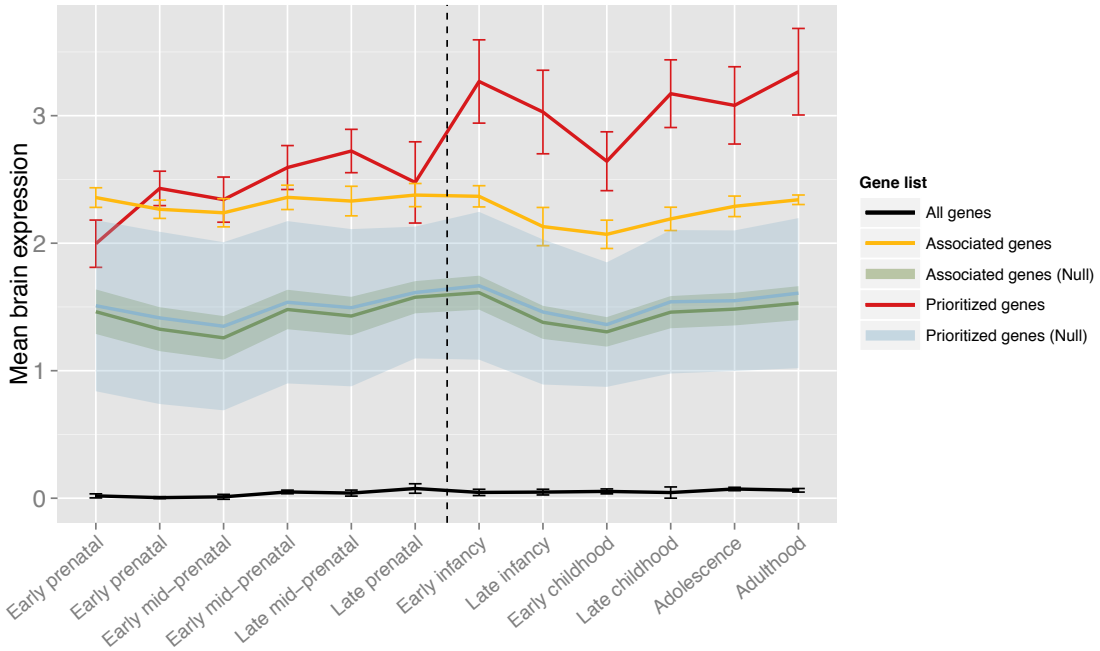
Supplementary Figure 4 – Overlap of gene lists



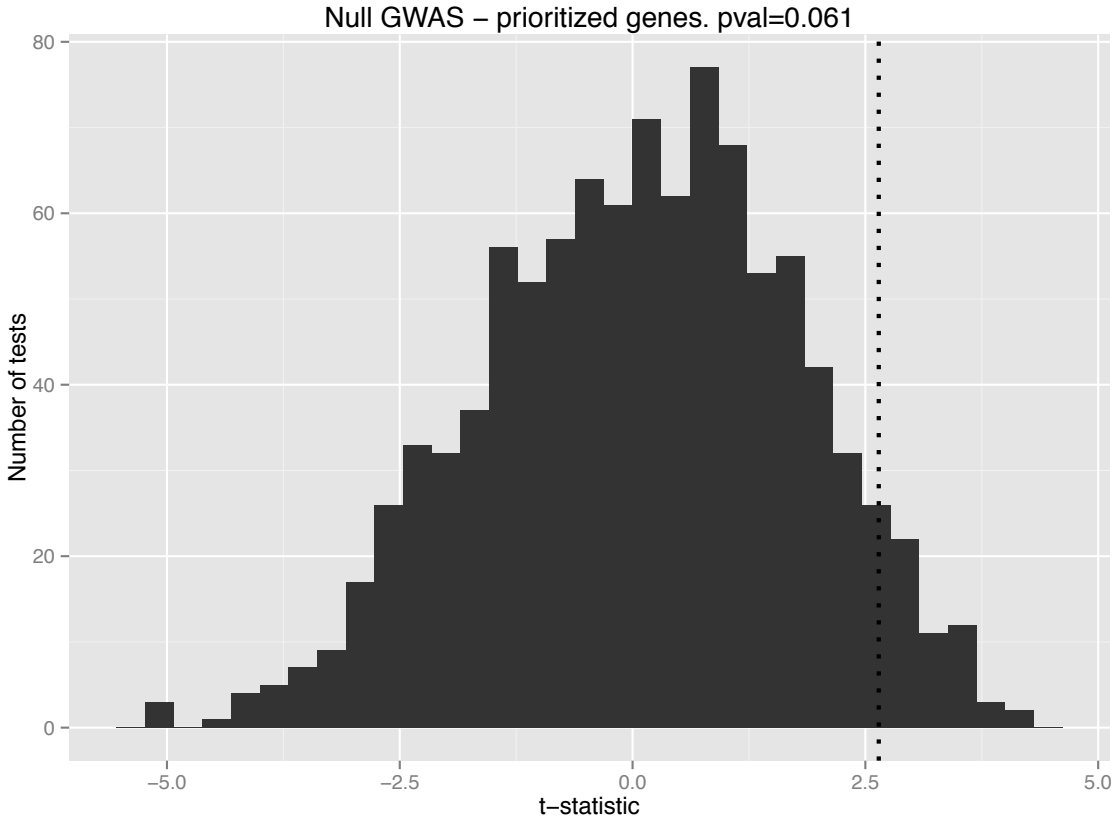
Supplementary Figure 5 – GWAS catalog traits' trajectories



Supplementary Figure 6 – RNA-Seq-based trajectories with null results



Supplementary Figure 7 – RNA-Seq-based trajectories with null results



References

1. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).

2. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316,** 1331–6 (2007).

3. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–9 (2000).

4. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25,** 309–16 (2007).

5. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40,** D109–14 (2012).

6. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** D691–7 (2011).

7. Blake, J. A., Bult, C. J., Eppig, J. T., Kadin, J. A. & Richardson, J. E. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res.* **42,** D810–7 (2014).