nature neuroscience

SUPPLEMENTARY INFORMATION

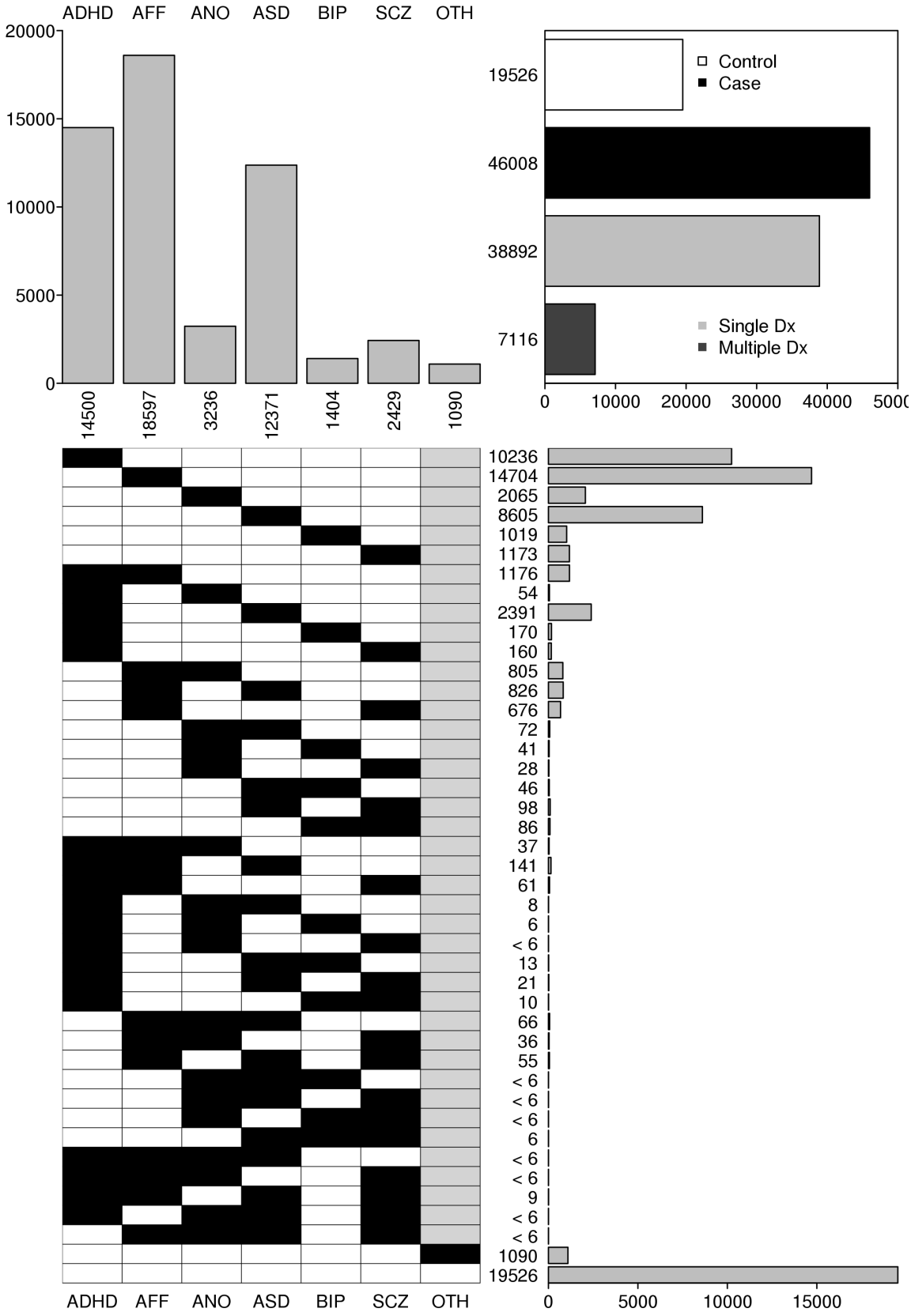
https://doi.org/10.1038/s41593-018-0320-0

In the format provided by the authors and unedited.

A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment

Andrew J. Schork[1,2], Hyejung Won[3,4,5,6,7], Vivek Appadurai[1,2], Ron Nudel[1,2], Mike Gandal[3,4,5], Olivier Delaneau[8,9,10], Malene Revsbech Christiansen[11], David M. Hougaard[2,12], Marie Bækved-Hansen[2,12], Jonas Bybjerg-Grauholm[2,12], Marianne Giørtz Pedersen[2,13,14], Esben Agerbo[2,13,14], Carsten Bøcker Pedersen[2,13,14], Benjamin M. Neale[15,16,17], Mark J. Daly[15,16,17], Naomi R. Wray[18,19], Merete Nordentoft[2,20,21], Ole Mors[2,22], Anders D. Børglum[2,23,24], Preben Bo Mortensen[2,13,14,24], Alfonso Buil[1,2], Wesley K. Thompson[1,2,25], Daniel H. Geschwind[3,4,5,26] and Thomas Werge[1,2,21*]

[1]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark. [2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Copenhagen, Denmark. [3]Department of Neurology, University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [5]Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [6]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [7]UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC, USA. [8]Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland. [9]Swiss Institute of Bioinformatics (SIB), University of Geneva, Geneva, Switzerland. [10]Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland. [11]DTU Bioinformatics, Technical University of Denmark, Lyngby, Denmark. [12]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [13]NCRR - National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [14]Centre for Integrated Register-based Research (CIRRAU), Aarhus University, Aarhus, Denmark. [15]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, USA. [16]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [17]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [18]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [19]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [20]Copenhagen Mental Health Center, Mental Health Services Capital Region of Denmark Copenhagen, Copenhagen, Denmark. [21]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [22]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. [23]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. [24]Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark. [25]Division of Biostatistics, Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA. [26]Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. *e-mail: Thomas.Werge@regionh.dk


ADHD
AFF
ANO
ASD
BIP
SCZ
OTH
20000
15000
10000
5000
0
14500
18597
3236
12371
1404
2429
1090
Control
Case
19526
46008
38892
7116
Single Dx
Multiple Dx
0
10000
20000
30000
40000
10236
14704
2065
8605
1019
1173
1176
54
2391
170
160
805
826
676
72
41
28
46
98
86
37
141
61
8
6
< 6
13
21
10
66
36
55
< 6
< 6
< 6
6
< 6
< 6
9
< 6
< 6
1090
19526
0
5000
10000
15000

Supplementary Figure 1. The Distribution of Indications in the iPSYCH GWAS Cohort. The histogram in the upper right panel depicts the counts for cases and controls as defined for the XDX GWAS. It also depicting the number of single and multiple indication cases. The lower left panel shows, in each row, a pattern of indications observed in the iPysch patient population, with filled boxed representing presence, white absence and grey untested. The OTH indication was only assigned to patients who did not have an ascertained indication, so is depicted as grey across other patterns. The histogram in the upper left panel shows the counts for each individual indication. These numbers sum to more than the total number of patients because individuals with multiple indications contribute to the counts for both. The histogram in the lower right panel depicts the counts for observed patterns of indications.

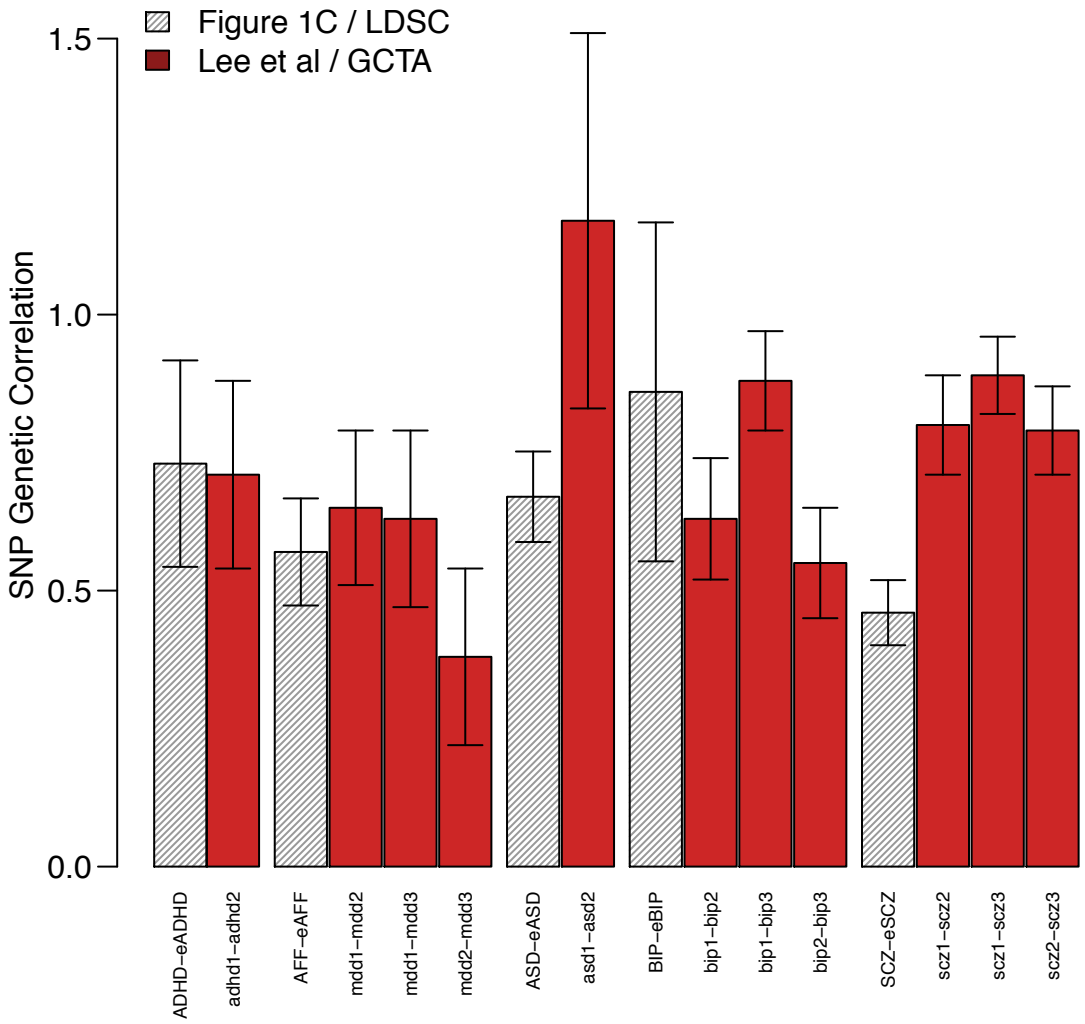
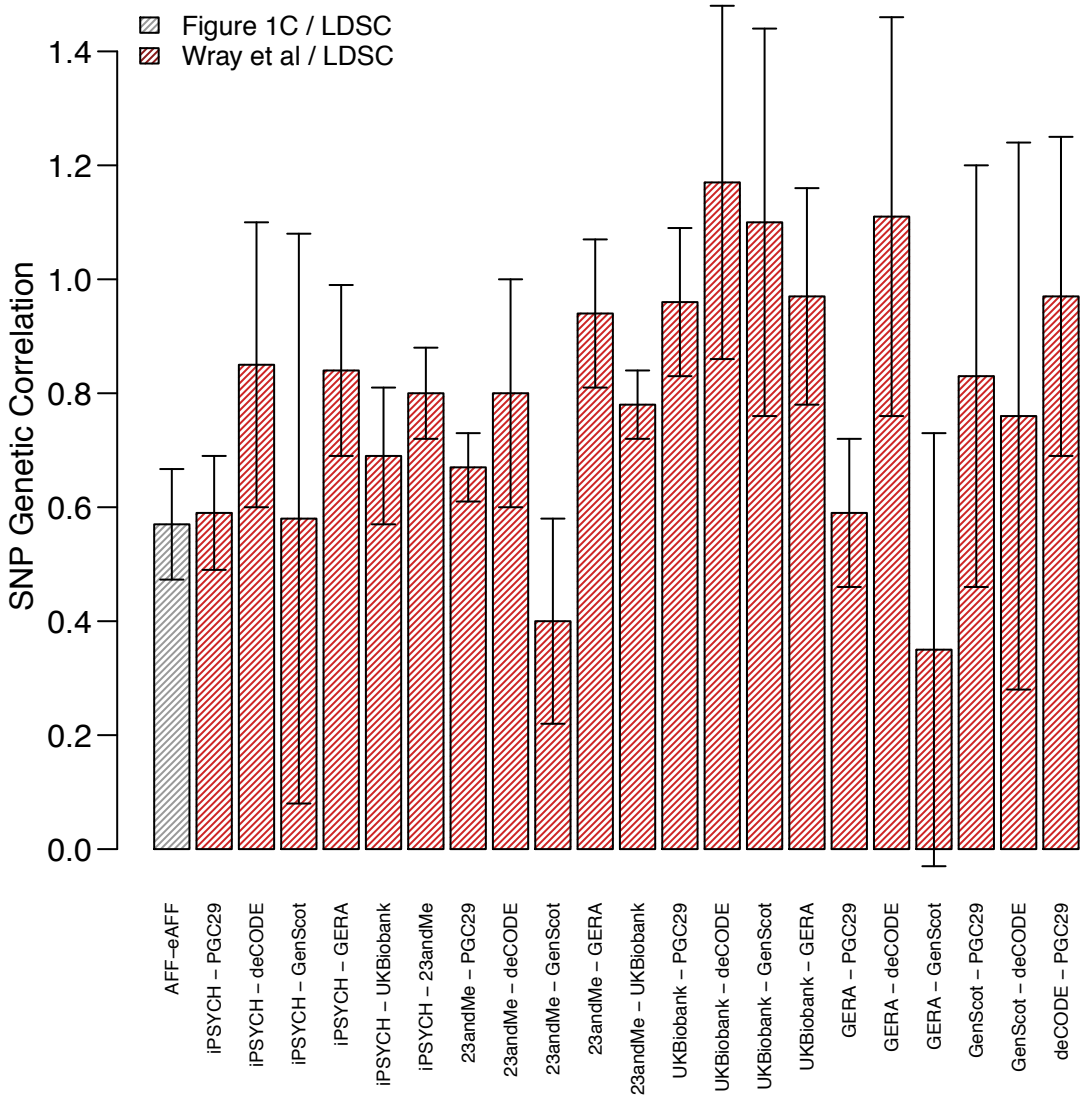


Figure 1C / LDSC
Lee et al / GCTA
SNP Genetic Correlation
1.5
1.0
0.5
0.0
ADHD−eADHD
adhd1−adhd2
AFF−eAFF
mdd1−mdd2
mdd1−mdd3
mdd2−mdd3
ASD−eASD
asd1−asd2
BIP−eBIP
bip1−bip2
bip1−bip3
bip2−bip3
SCZ−eSCZ
scz1−scz2
scz1−scz3
scz2−scz3

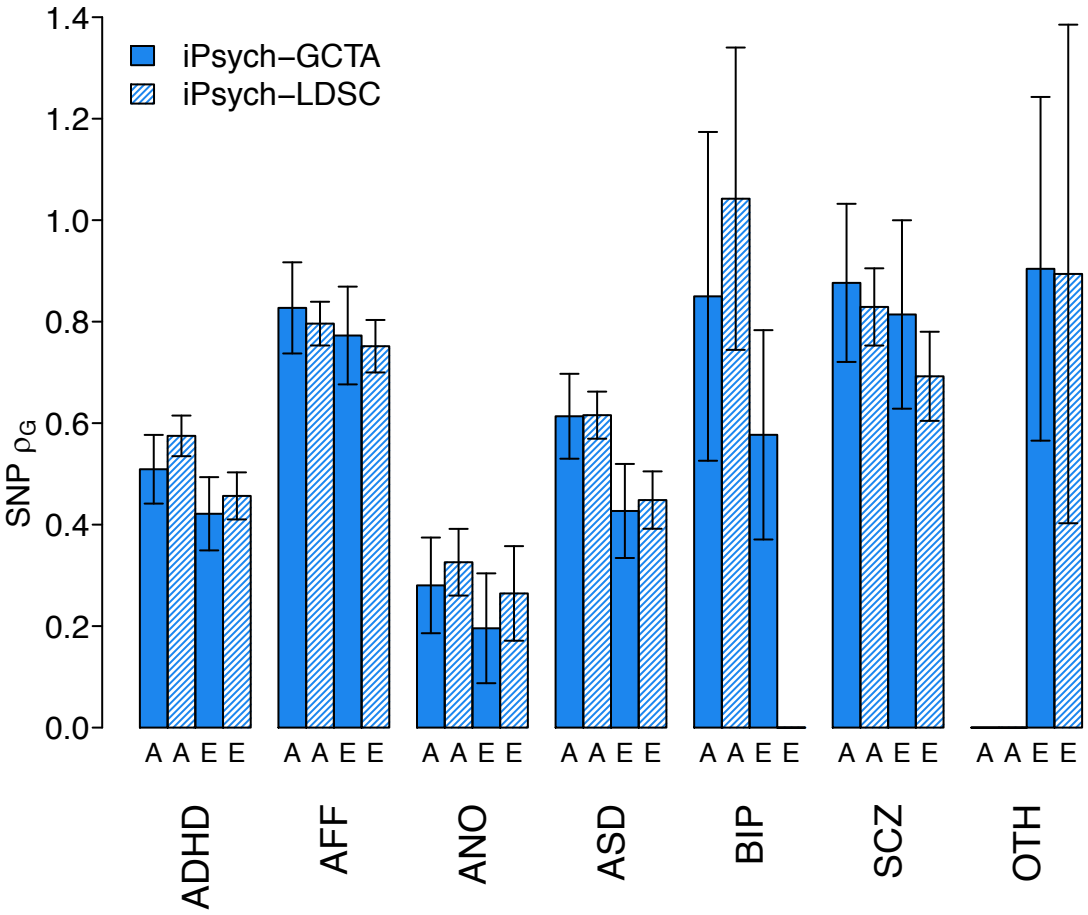
Supplementary Figure 2. Multiple iPSYCH-eGWAS genetic correlation versus other cross-cohort estimates. iPSYCH-external study genetic correlations of the same disorder are consistent with previous reports of across-cohort within-disorder genetic correlations. Here we present the genetic correlations reported in Figure 1C alongside similar estimates obtained from Supplementary Table 5B of Lee et al[1]. Bar height denotes genetic correlation point estimate, error bars are estimate standard errors. Sample Sizes: ADHD, n = cases 14500, controls 18597; eADHD, n = cases 2960, controls 9240; adhd1, n = cases 1736, controls = 1766; adhd2, n = cases 2427, controls = 10274; AFF, n = cases 18597, controls 3236; eAFF, n = cases 9240, controls 3495; mdd1, n = cases 3077, controls = 3420; mdd2, n = cases 3785, controls = 3289; mdd3, n = cases 2179, controls = 2672; ASD, n = cases 12371, controls 1404; eASD, n = cases 7387, controls 9784; asd1, n = cases 1893, controls = 1888; asd2, n = cases 1540, controls = 1540; BIP, n = cases 1404, controls 2429; eBIP, n = cases 9784, controls 34241; bip1, n = cases 2465, controls = 4058; bip2, n = cases 2540, controls = 2058; bip3, n = cases 1699, controls = 2915; SCZ, n = cases 2429, controls 33332; eSCZ, n = cases 34241, controls; scz1, n = cases 3220, controls = 3445; scz2, n = cases 2571, controls = 2419; scz3, n = cases 3296, controls = 6307.

1 Cross-Disorder Group of the Psychiatric Genomics Consortium et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45, 984-994, doi:10.1038/ng.2711 (2013).

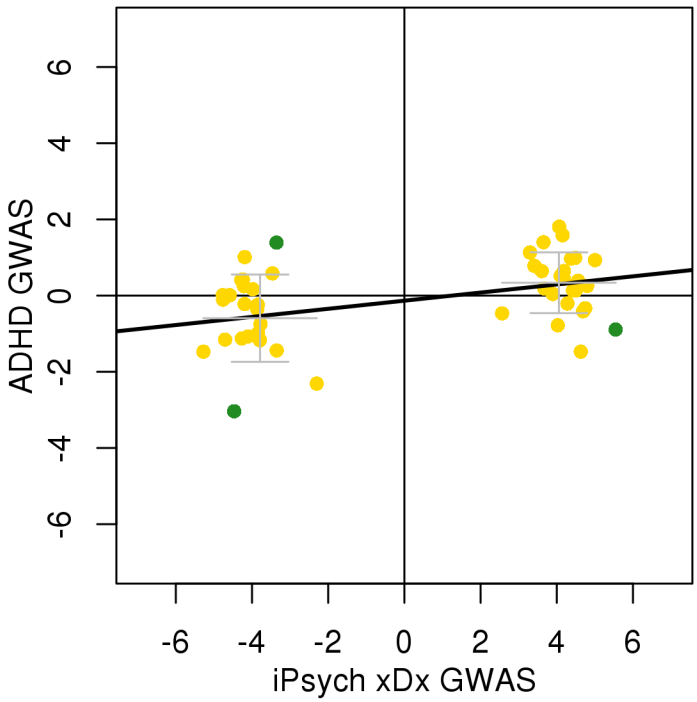


Supplementary Figure 3. iPSYCH-eMDD genetic correlation versus other cross-cohort estimates. The iPSYCH-external study genetic correlation for affective disorder is consistent with the magnitude of across-study genetic correlations described in a recent meta-analysis, which includes a broader assessment of the iPSYCH genetic correlations with multiple independent cohorts. Bar height represents LDSC genetic correlation point estimate, bars are estimate standard errors. Data was taken from Supplementary Table 3 of Wray et al[1]. Sample sizes: AFF, n = cases 18597, controls 3236; eAFF, n = cases 9240, controls 3495; iPSYCH, n = cases 18629, controls 17841; PGC29, n = cases 16823, controls 25632; deCODE, n = cases 1980, controls 9536; GenScot, n = cases 997, controls 6358; GERA, n = cases 7162, controls 38307; UKBiobank, n = cases 14260, controls 15480; 23andMe, n = cases 75607, controls 231747.

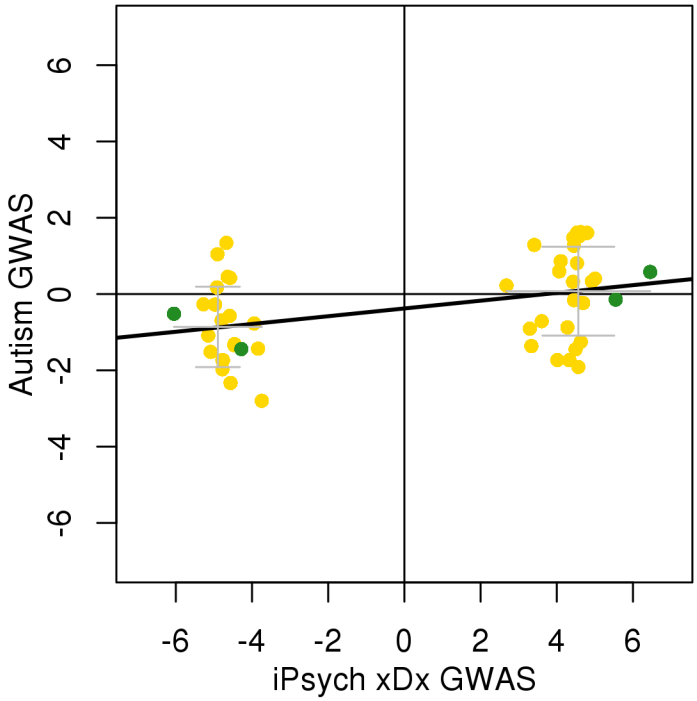
1 Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 50, 668-681, doi:10.1038/s41588-018-0090-3 (2018).


iPsych–GCTA
iPsych–LDSC
SNP ρG
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
A A E E
ADHD
AFF
ANO
ASD
BIP
SCZ
OTH

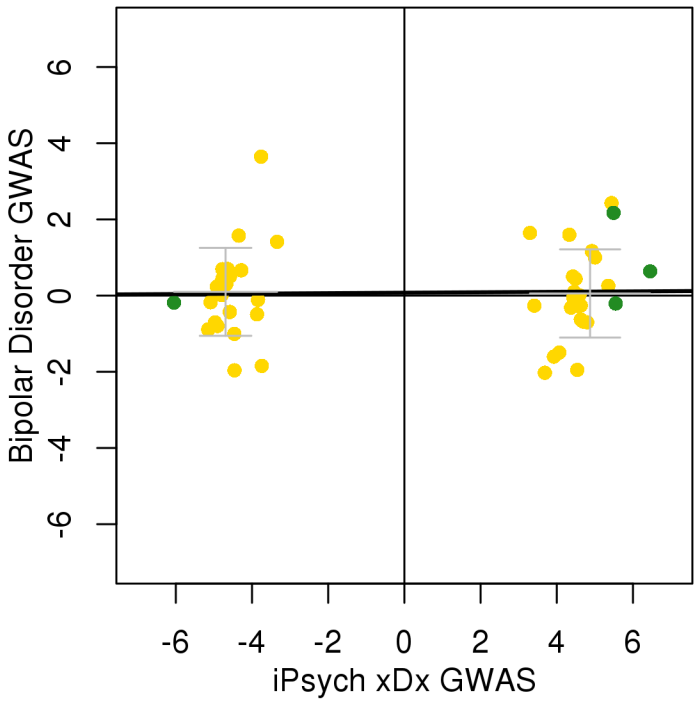
Supplementary Figure 4. Each indication shows at least moderate genetic correlation with the union of all other indications. Here we used bivariate GCTA to estimate the genetic correlation (solid bars) between all individuals (A) with a given indication and the complimentary portion of the XDX case-cohort (XDX less A). We also estimated the same genetic correlation excluding comorbid cases from the single disorder cohort (E), but keep the controls unchanged (XDX less A). The estimates were repeated using LDSC to estimate genetic correlations (striped bars). Error bars denote estimate standard errors of point estimates. Sample sizes: ADHD-A-GCTA, n = 7601 cases, 8988 controls; ADHD-E-GCTA, n = 5293 cases, 8988 controls; ADHD-XDX less A-GCTA, n = 17733 cases, 8989 controls; ADHD-A-LDSC, n = 14500 cases, 19526 controls; ADHD-E-LDSC, n = 10236 cases, 19526 controls; ADHD-XDX less A-LDSC, n = 31508 cases, 19526 controls; AFF-A-GCTA, n = 9929 cases, 8988 controls; AFF-E-GCTA, n = 7854 cases, 8988 controls; AFF-XDX less A-GCTA, n = 15405 cases, 8989 controls; AFF-A-LDSC, n = 18597 cases, 19526 controls; AFF-E-LDSC, n = 14704 cases, 19526 controls; AFF-XDX less A-LDSC, n = 27411 cases, 19526 controls; ANO-A-GCTA, n = 1837 cases, 8988 controls; ANO-E-GCTA, n = 1184 cases, 8988 controls; ANO-XDX less A-GCTA, n = 23497 cases, 8989 controls; ANO-A-LDSC, n = 3236 cases, 19526 controls; ANO-E-LDSC, n = 2065 cases, 19526 controls; ANO-XDX less A-LDSC, n = 42772 cases, 19526 controls; ASD-A-GCTA, n = 6939 cases, 8988 controls; ASD-E-GCTA, n = 4825 cases, 8988 controls; ASD-XDX less A-GCTA, n = 18395 cases, 8989 controls; ASD-A-LDSC, n = 12371 cases, 19526 controls; ASD-E-LDSC, n = 8605 cases, 19526 controls; ASD-XDX less A-LDSC, n = 33637 cases, 19526 controls; BIP-A-GCTA, n = 780 cases, 8988 controls; BIP-E-GCTA, n = 560 cases, 8988 controls; BIP-XDX less A-GCTA, n = 24554 cases, 8989 controls; BIP-A-LDSC, n = 1404 cases, 19526 controls; BIP-XDX less A-LDSC, n = 44604 cases, 19526 controls; OTH-E-GCTA, n = 1097 cases, 8988 controls; OTH-XDX less A-GCTA, n = 24237 cases, 8989 controls; OTH-E-LDSC, n = 1090 cases, 19526 controls; OTH-XDX less A-LDSC, n = 44918 cases, 19526 controls; SCZ-A-GCTA, n = 1330 cases, 8988 controls; SCZ-E-GCTA, n = 632 cases, 8988 controls; SCZ-XDX less A-GCTA, n = 24004 cases, 8989 controls; SCZ-A-LDSC, n = 2429 cases, 19526 controls; SCZ-E-LDSC, n = 1173 cases, 19526 controls; SCZ-XDX less A-LDSC, n = 43579 cases, 19526 controls



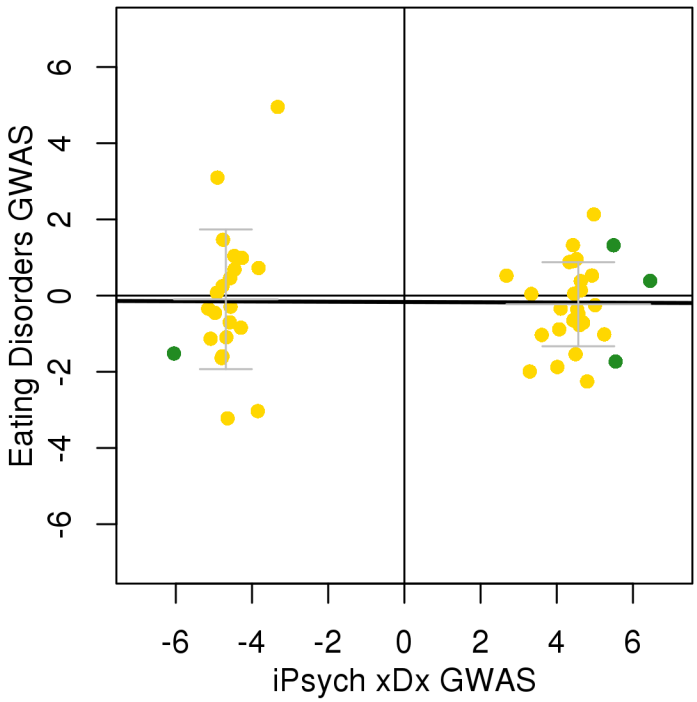
Supplementary Figure 5. Z-score concordance between iPSYCH XDX GWAS and the eADHD GWAS. The z-scores for the best proxy SNPs in the eADHD GWAS are plotted against the z-scores for the same SNPs from the XDX GWAS. Genome-wide significant XDX loci 1-4 are shown with green dots and suggestive loci 5-46 are shown with gold. Black line shows linear regression slope (N=49 z-score pairs; linear regression: $\beta = 0.11$, s.e. = 0.03; t = 3.14; two sided $p = 2.9 \times 10^{-3}$). Grey bars depict the mean and one standard deviation interval of replication effects for SNPs with positive or negative effects in the XDX GWAS. This concordance is significant after correction for multiple testing ($p < 5 \times 10^{-3}$, correcting for 10 concordance tests).



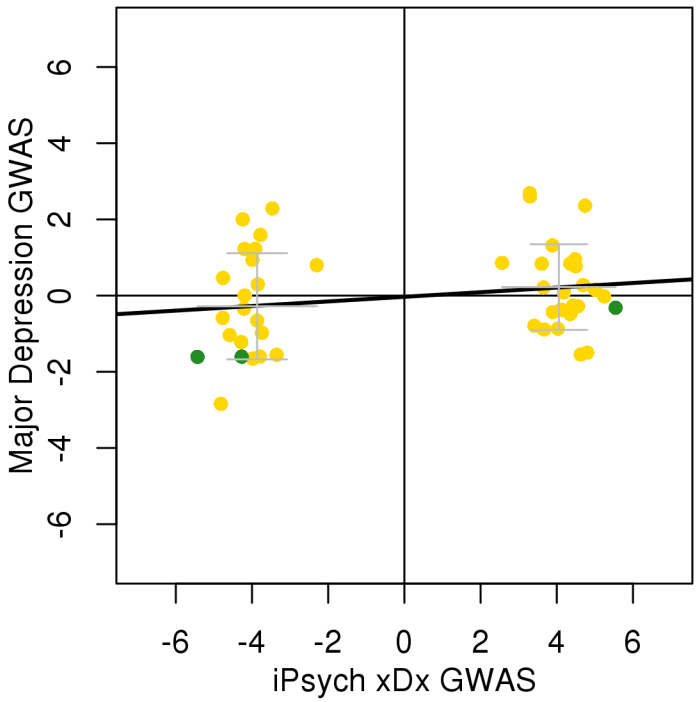
Supplementary Figure 6. Z-score concordance between iPSYCH XDX GWAS and the eASD GWAS. The z-scores for the best proxy SNPs in the eASD GWAS are plotted against the z-scores for the same SNPs from the XDX GWAS. Genome-wide significant XDX loci 1-4 are shown with green dots and suggestive loci 5-46 are shown with gold. Black line shows linear regression slope (N=50 z-score pairs; linear regression: β = 0.10, s.e. = 0.03; t = 2.93; two sided p = 5.1 x 10^{-3}). Grey bars depict the mean and one standard deviation interval of replication effects for SNPs with positive or negative effects in the XDX GWAS. This concordance is trending, but not significant after correction for multiple testing (p > 5x10^{-3}, correcting for 10 concordance tests).



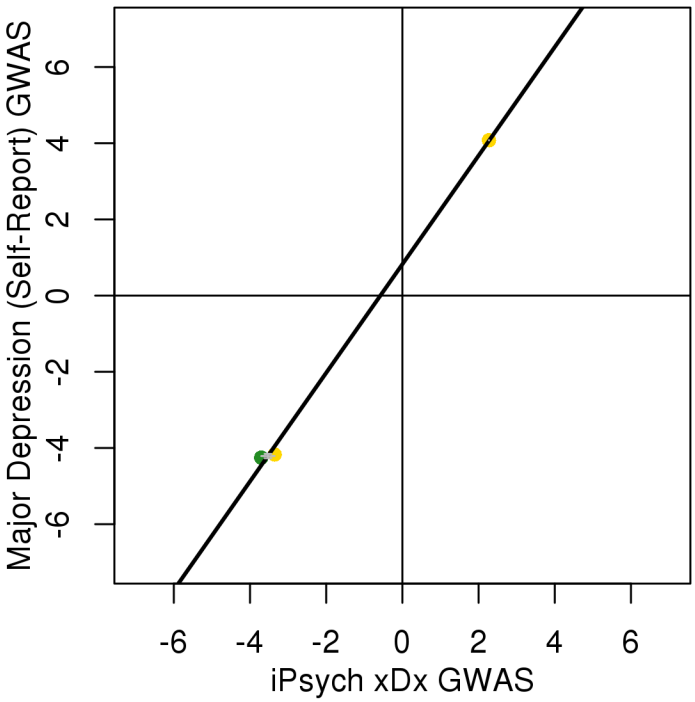
Supplementary Figure 7. Z-score concordance between iPSYCH XDX GWAS and the eBIP GWAS. The z-scores for the best proxy SNPs in the eBIP GWAS are plotted against the z-scores for the same SNPs from the XDX GWAS. Genome-wide significant XDX loci 1-4 are shown with green dots and suggestive loci 5-46 are shown with gold. Black line shows linear regression slope (N=50 z-score pairs; linear regression: β = 0.01, s.e. = 0.04; t = 0.17; two sided p = 0.87). Grey bars depict the mean and one standard deviation interval of replication effects for SNPs with positive or negative effects in the XDX GWAS.



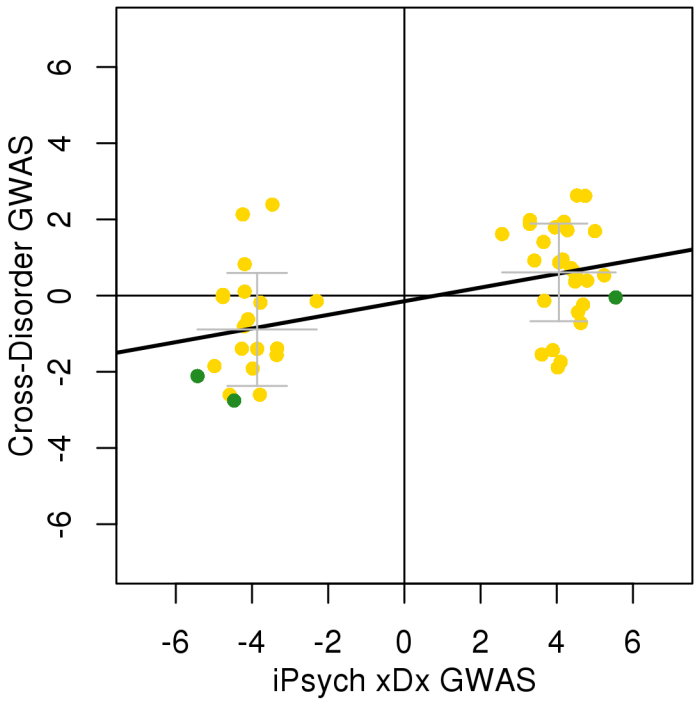
Supplementary Figure 8. Z-score concordance between iPSYCH XDX GWAS and the eANO GWAS. The z-scores for the best proxy SNPs in the eANO GWAS are plotted against the z-scores for the same SNPs from the XDX GWAS. Genome-wide significant XDX loci 1-4 are shown with green dots and suggestive loci 5-46 are shown with gold. Black line shows linear regression slope (N=50 z-score pairs; linear regression: β = -0.004, s.e. = 0.05; t = -0.08; two sided p = 0.94). Grey bars depict the mean and one standard deviation interval of replication effects for SNPs with positive or negative effects in the XDX GWAS.



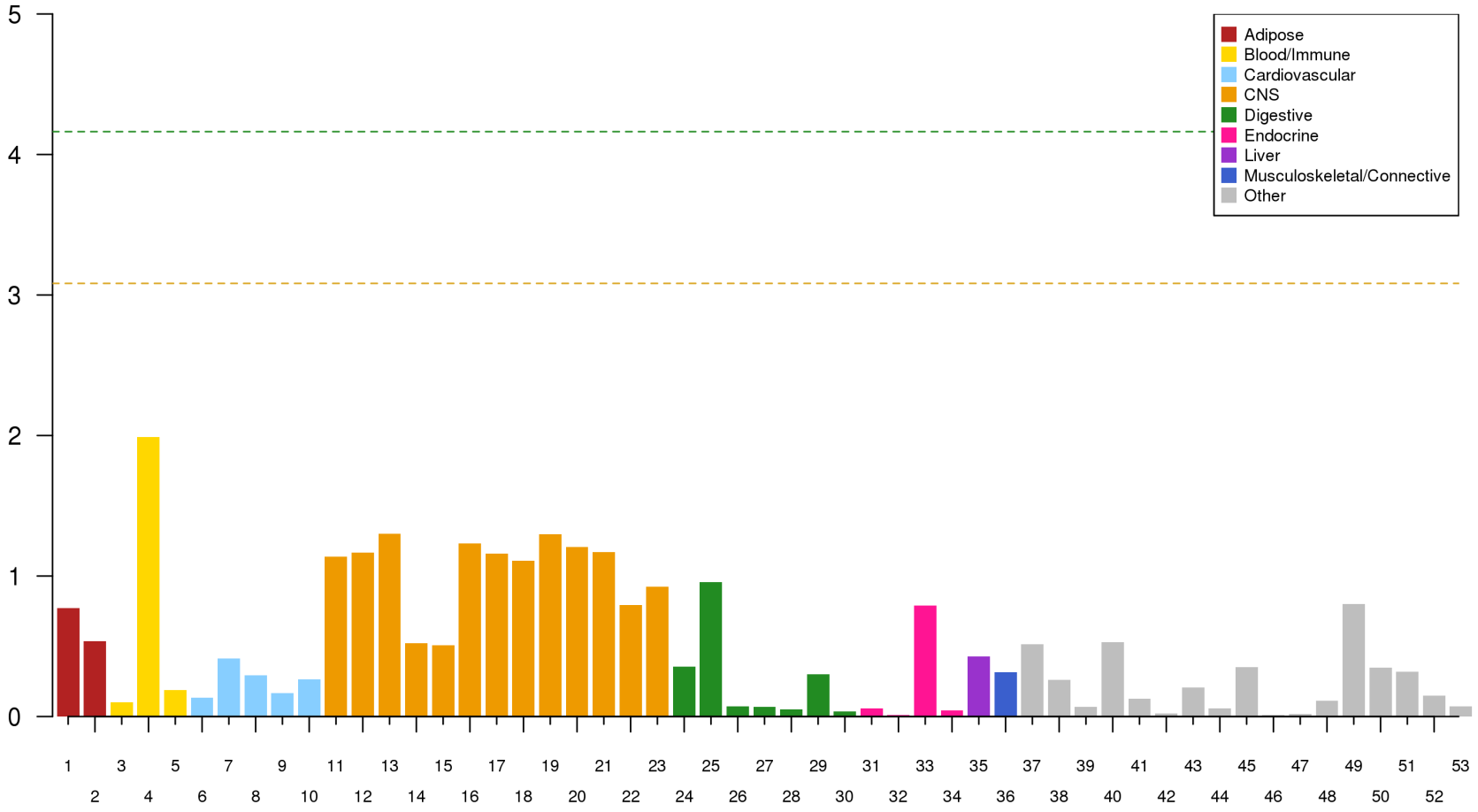
Supplementary Figure 9. Z-score concordance between iPSYCH XDX GWAS and the eAFF GWAS. The z-scores for the best proxy SNPs in the eAFF GWAS are plotted against the z-scores for the same SNPs from the XDX GWAS. Genome-wide significant XDX loci 1-4 are shown with green dots and suggestive loci 5-46 are shown with gold. Black line shows linear regression slope (N=49 z-score pairs; linear regression: $\beta = 0.06$, s.e. = 0.04; t = 1.41; two sided p = 0.17). Grey bars depict the mean and one standard deviation interval of replication effects for SNPs with positive or negative effects in the XDX GWAS.



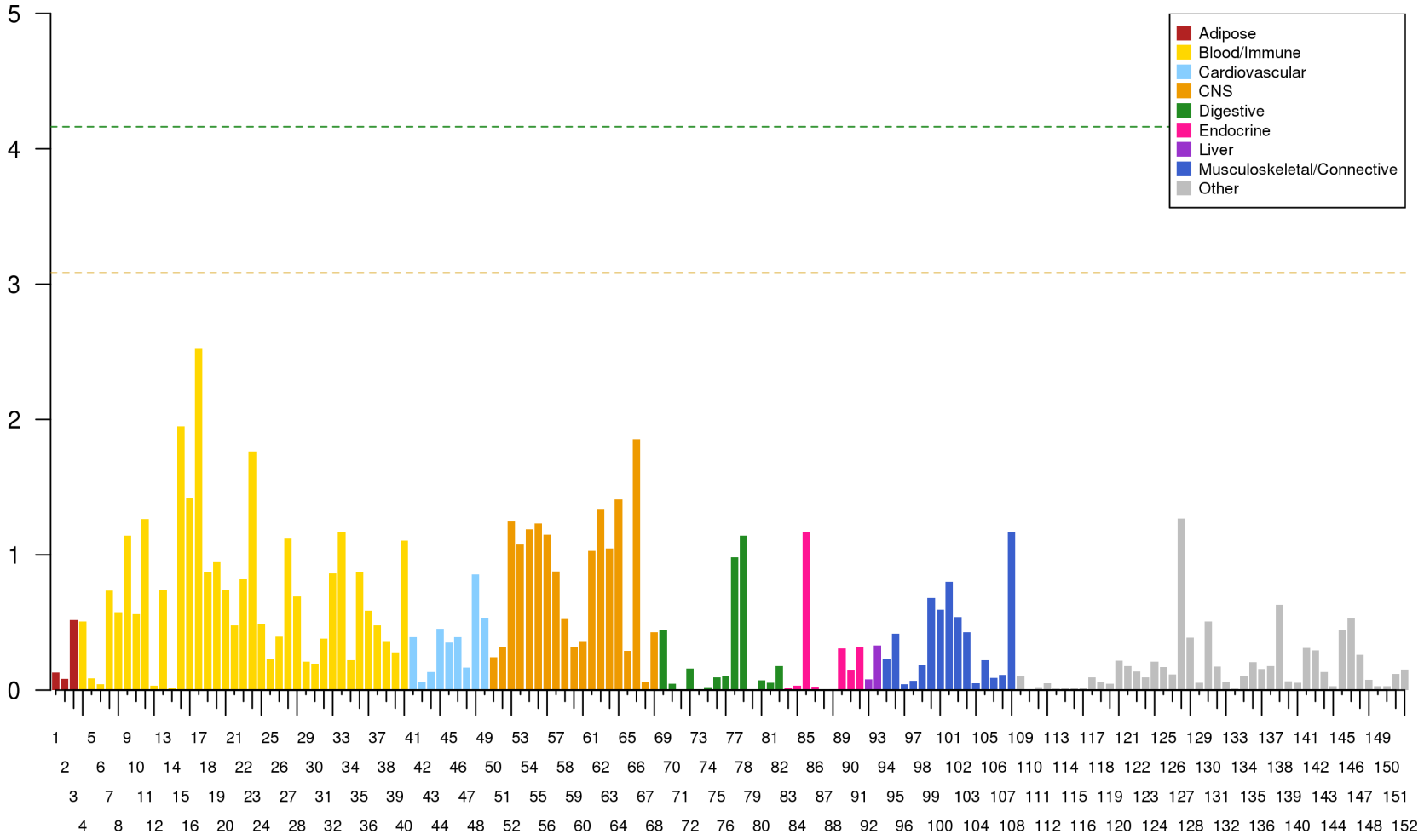
Supplementary Figure 10. Z-score concordance between iPSYCH XDX GWAS and the eSR-MDD GWAS. The z-scores for the best proxy SNPs in the eSR-MDD GWAS are plotted against the z-scores for the same SNPs from the XDX GWAS. Genome-wide significant XDX loci 1-4 are shown with green dots and suggestive loci 5-46 are shown with gold. Black line shows linear regression slope (N=3 z-score pairs; linear regression: β = 1.43, s.e. = 0.06; t = 22.63; two sided p = 0.028). Grey bars depict the mean and one standard deviation interval of replication effects for SNPs with positive or negative effects in the XDX GWAS. This concordance is not significant after correction for multiple testing ($p > 5x10^{-3}$, correcting for 10 concordance tests).



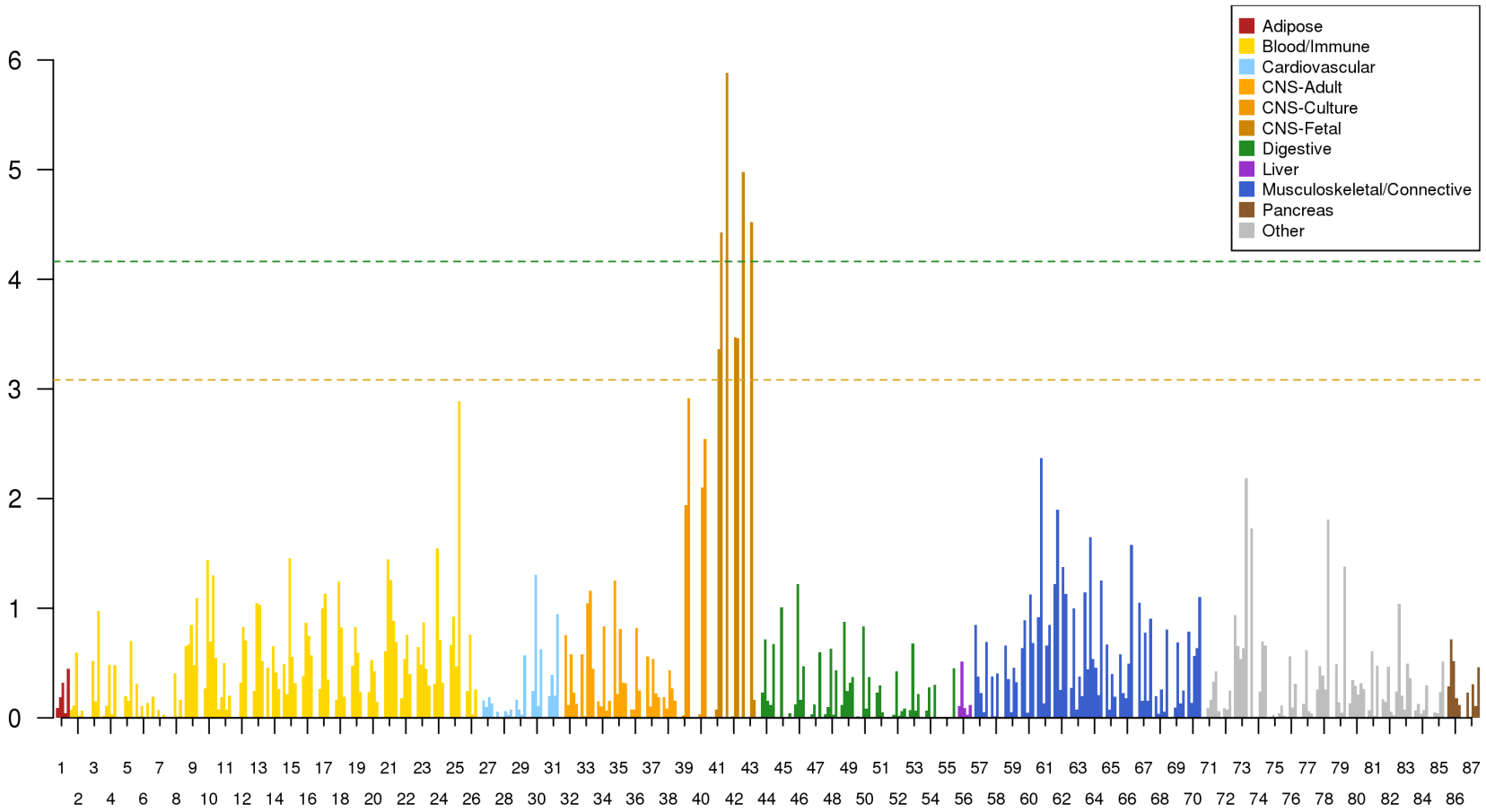
Supplementary Figure 11. Z-score concordance between iPSYCH XDX GWAS and the eXDX GWAS. The z-scores for the best proxy SNPs in the eXDX GWAS are plotted against the z-scores for the same SNPs from the XDX GWAS. Genome-wide significant XDX loci 1-4 are shown with green dots and suggestive loci 5-46 are shown with gold. Black line shows linear regression slope (N=49 z-score pairs; linear regression: $\beta = 0.18$, s.e. = 0.05; t = 3.82; two sided $p = 3.9 \times 10^{-4}$). Grey bars depict the mean and one standard deviation interval of replication effects for SNPs with positive or negative effects in the XDX GWAS. This concordance is significant after correction for multiple testing ($p < 5 \times 10^{-3}$, correcting for 10 concordance tests).


5
4
3
2
1
0
Adipose
Blood/Immune
Cardiovascular
CNS
Digestive
Endocrine
Liver
Musculoskeletal/Connective
Other
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52

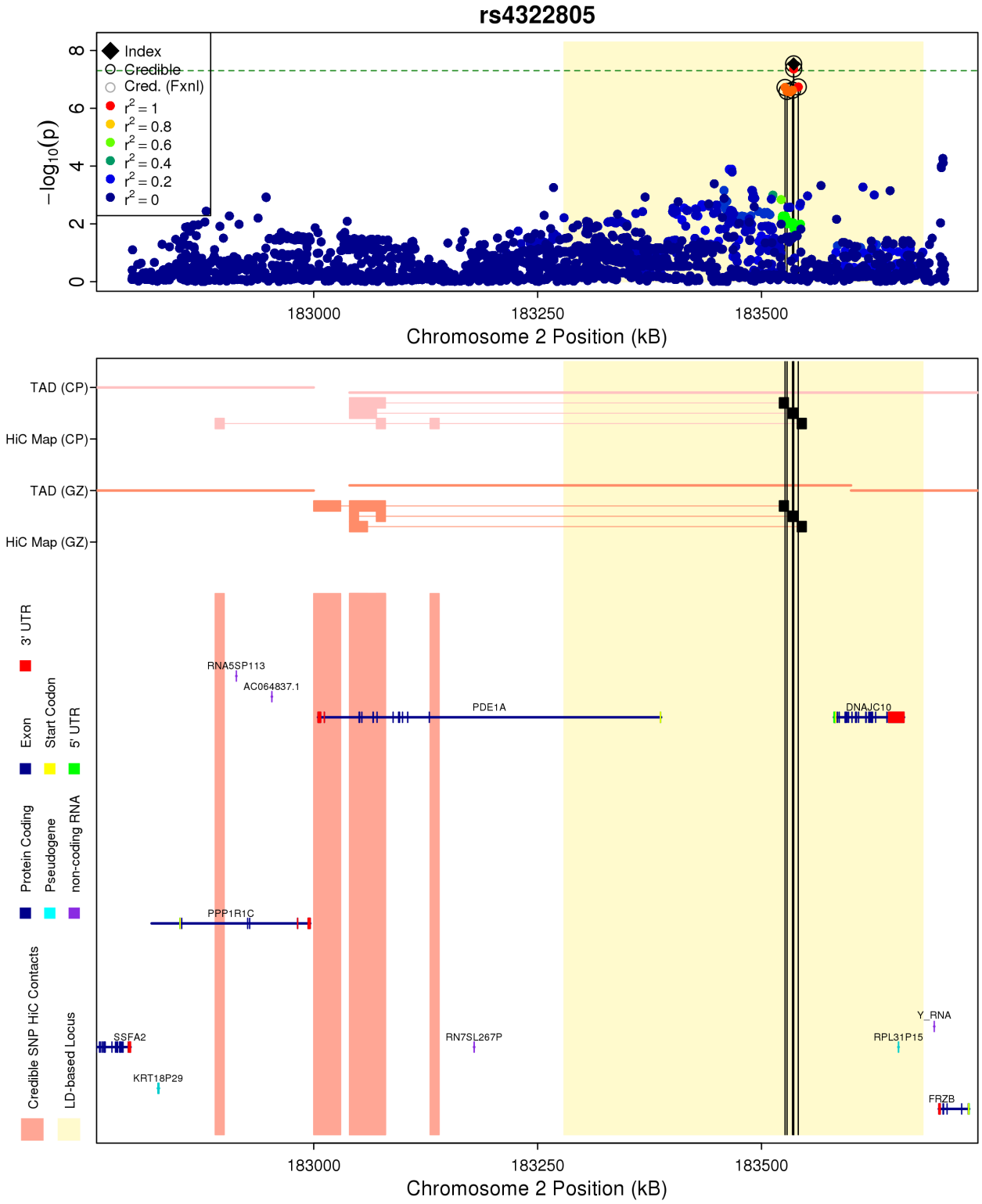
Supplementary Figure 12. LDSC-SEG with GTEx Preferential Expression Sets. Each bar represents the $-\log_{10}$(p-value) from an LDSC-SEG test of enrichment in the XDX GWAS (n=46,008 cases, 19,526 controls) for a different tissue specific expression variant annotation from the 53 in the GTEx set. The green dashed line represents significant enrichment ($p < 8.32 \times 10^{-5}$ to correct for 601 LD-scores) and gold dashed line represents suggestive enrichment (FDR < 0.05 across all 601 tests). Complete Statistics presented in supplementary table 15.


5
4
3
2
1
0
Adipose
Blood/Immune
Cardiovascular
CNS
Digestive
Endocrine
Liver
Musculoskeletal/Connective
Other
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97 101 105 109 113 117 121 125 129 133 137 141 145 149
2 6 10 14 18 22 26 30 34 38 42 46 50 54 58 62 66 70 74 78 82 86 90 94 98 102 106 110 114 118 122 126 130 134 138 142 146 150
3 7 11 15 19 23 27 31 35 39 43 47 51 55 59 63 67 71 75 79 83 87 91 95 99 103 107 111 115 119 123 127 131 135 139 143 147 151
4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96 100 104 108 112 116 120 124 128 132 136 140 144 148 152

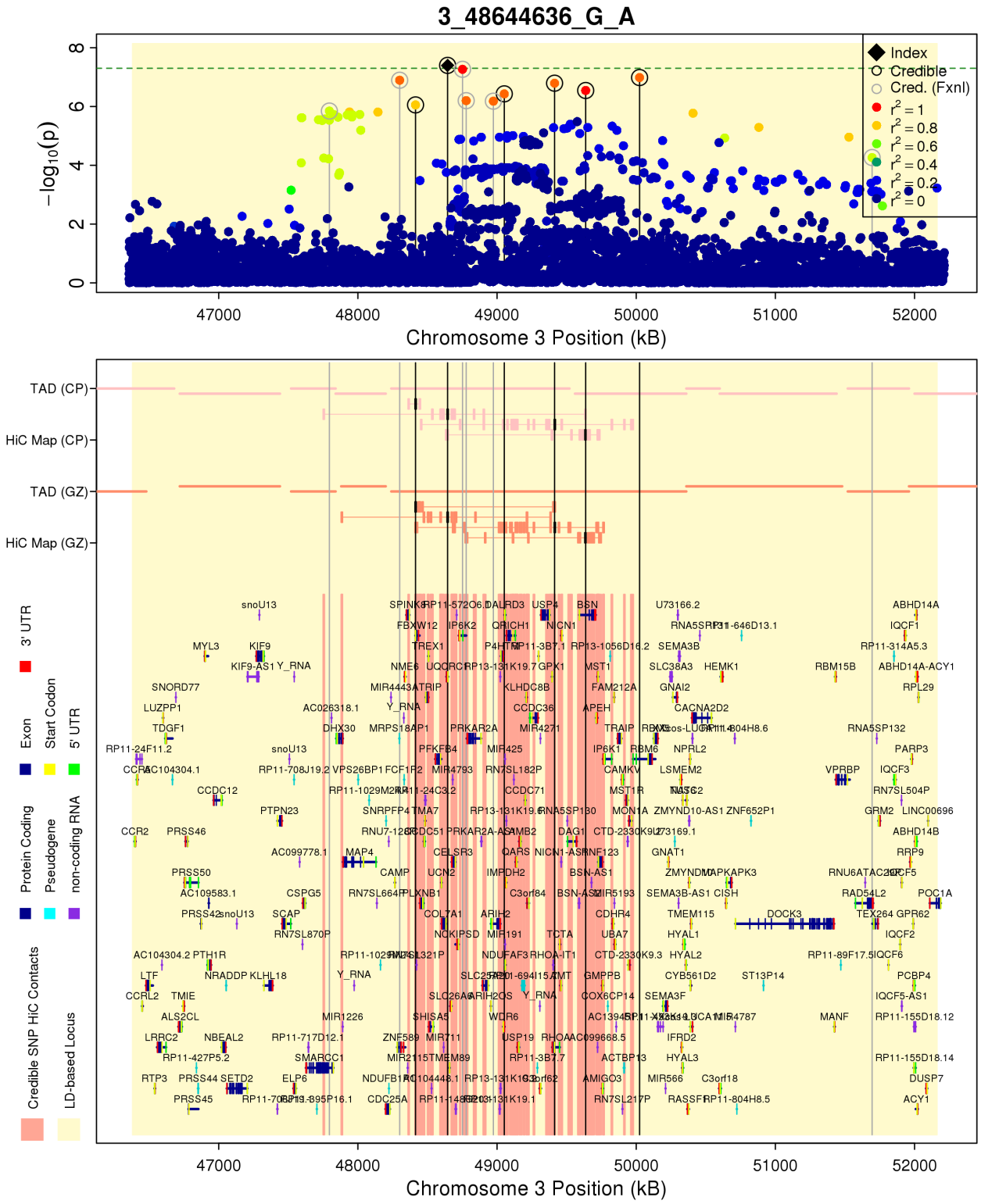
Supplementary Figure 13. LDSC-SEG with DEPICT Preferential Expression Sets. Each bar represents the $-\log_{10}$(p-value) from an LDSC-SEG test of enrichment in the XDX GWAS (n=46,008 cases, 19,526 controls) for a different tissue specific expression variant annotation from the 152 in the DEPICT set. The green dashed line represents significant enrichment ($p < 8.32 \times 10^{-5}$ to correct for 601 LD-scores) and gold dashed line represents suggestive enrichment (FDR < 0.05 across all 601 tests). Complete Statistics presented in supplementary table 16.


Adipose
Blood/Immune
Cardiovascular
CNS-Adult
CNS-Culture
CNS-Fetal
Digestive
Liver
Musculoskeletal/Connective
Pancreas
Other
0
1
2
3
4
5
6
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86

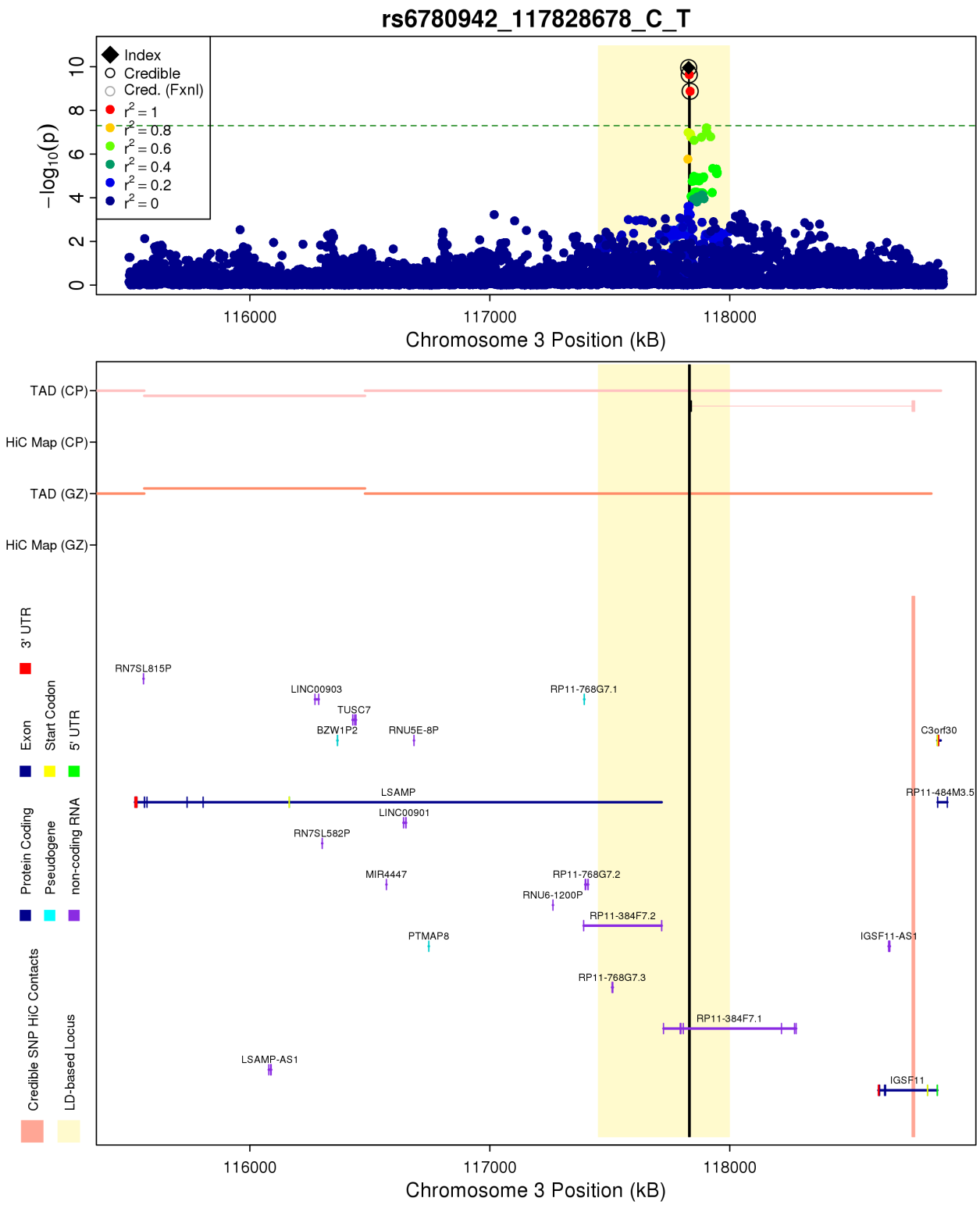
Supplementary Figure 14. LDSC-SEG with RoadMap Chromatin Marks. Each bar represents the $-\log_{10}$(p-value) from an LDSC-SEG test of enrichment in the XDX GWAS (n=46,008 cases, 19,526 controls) for a different tissue specific expression variant annotation from the 396 in the RoadMap set. The green dashed line represents significant enrichment ($p < 8.32 \times 10^{-5}$ to correct for 601 LD-scores) and gold dashed line represents suggestive enrichment (FDR < 0.05 across all 601 tests). Complete Statistics presented in supplementary table 18.


rs4322805
Index
Credible
Cred. (Fxnl)
r2 = 1
r2 = 0.8
r2 = 0.6
r2 = 0.4
r2 = 0.2
r2 = 0
−log10(p)
0
2
4
6
8
183000
183250
183500
Chromosome 2 Position (kB)
TAD (CP)
HiC Map (CP)
TAD (GZ)
HiC Map (GZ)
3' UTR
Exon
Start Codon
5' UTR
Protein Coding
Pseudogene
non-coding RNA
Credible SNP HiC Contacts
LD-based Locus
RNA5SP113
AC064837.1
PDE1A
DNAJC10
PPP1R1C
SSFA2
KRT18P29
RN7SL267P
Y_RNA
RPL31P15
FRZB
183000
183250
183500
Chromosome 2 Position (kB)

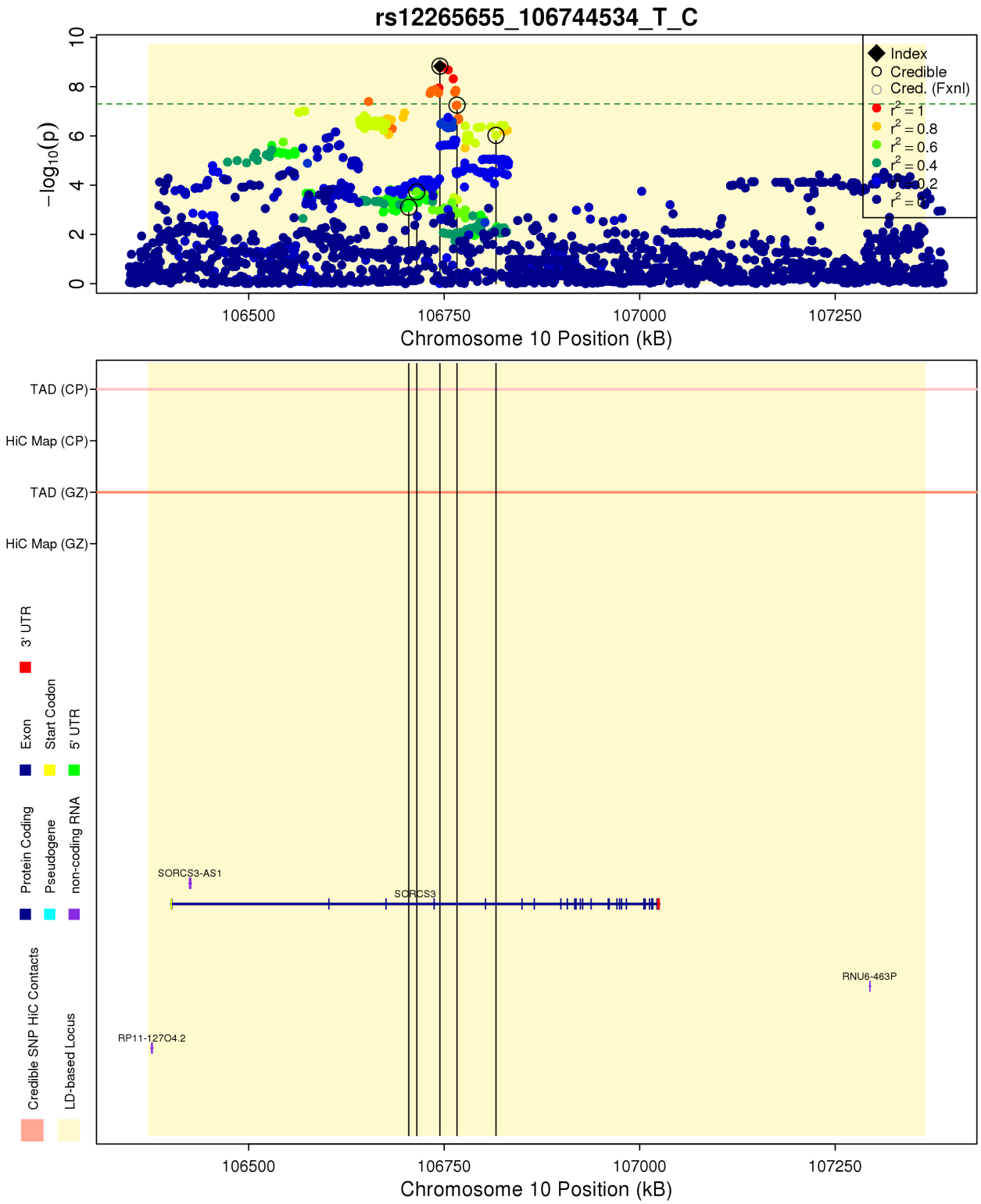
Supplementary Figure 15. Annotated Local Manhattan Plot for Locus 1. The top panel depicts the $-\log_{10}$(two-sided p-values) from the logistic regression association tests performed for the XDX GWAS (n=46,008 cases, 19,526 controls) for variants in the region of interest surrounding the implicated locus. The extent of the linkage disequilibrium (LD) defined locus implicated by the index SNP (black diamond) is contained within the pale-yellow background. The color coding of each variant represents the strength of LD with the index SNP. Credible SNPs are encircled and their position is indicated by a vertical black bar. The bottom panel shows a selected set of functional elements within the region of interest. Topological associated domains (TAD) are depicted by thick horizontal lines, offset slightly to note changes, for two fetal tissues: Cortical Plate (CO) and Germinal Zone (GZ). 10 kilo-base regions containing credible SNPs are depicted as black boxes and their HiC contact regions in each fetal tissue are defined by thin line-connected colored rectangles. Genes and functional elements from the longest transcript are depicted below the interaction schematics, and the union of credible SNP contact regions in the two tissues are depicted by the intermediately colored pink background.


3_48644636_G_A
Index
Credible
Cred. (Fxnl)
r2 = 1
r2 = 0.8
r2 = 0.6
r2 = 0.4
r2 = 0.2
r2 = 0
−log10(p)
0
2
4
6
8
47000
48000
49000
50000
51000
52000
Chromosome 3 Position (kB)
TAD (CP)
HiC Map (CP)
TAD (GZ)
HiC Map (GZ)
3' UTR
Exon
Start Codon
5' UTR
Protein Coding
Pseudogene
non-coding RNA
Credible SNP HiC Contacts
LD-based Locus
Chromosome 3 Position (kB)

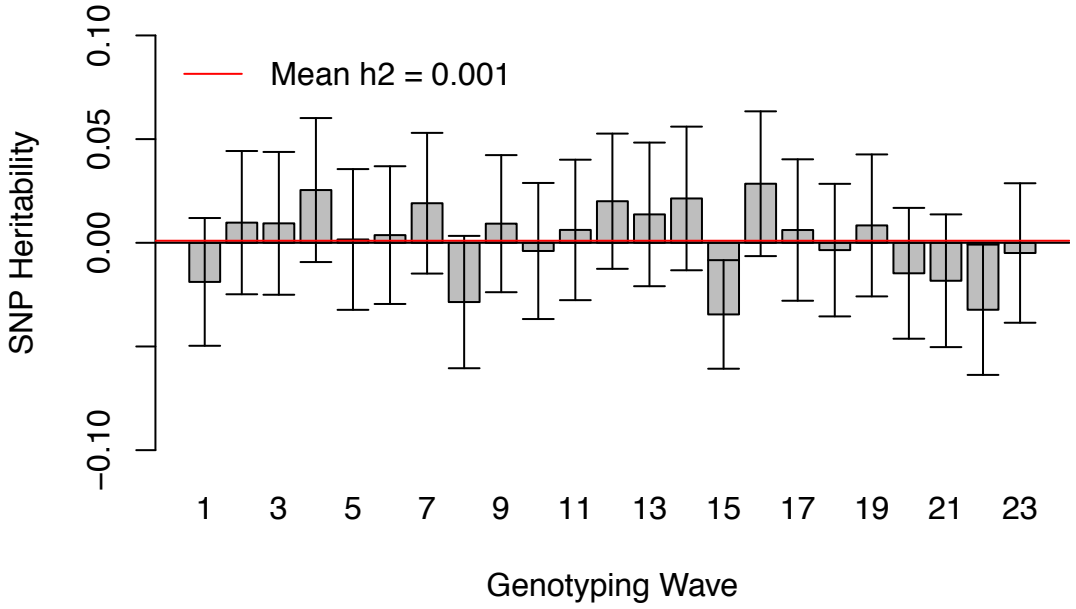
Supplementary Figure 16. Annotated Local Manhattan Plot for Locus 2. The top panel depicts the $-\log_{10}$(two-sided p-values) from the logistic regression association tests performed for the XDX GWAS (n=46,008 cases, 19,526 controls) for variants in the region of interest surrounding the implicated locus. The extent of the linkage disequilibrium (LD) defined locus implicated by the index SNP (black diamond) is contained within the pale-yellow background. The color coding of each variant represents the strength of LD with the index SNP. Credible SNPs are encircled and their position is indicated by a vertical black bar. The bottom panel shows a selected set of functional elements within the region of interest. Topological associated domains (TAD) are depicted by thick horizontal lines, offset slightly to note changes, for two fetal tissues: Cortical Plate (CO) and Germinal Zone (GZ). 10 kilo-base regions containing credible SNPs are depicted as black boxes and their HiC contact regions in each fetal tissue are defined by thin line-connected colored rectangles. Genes and functional elements from the longest transcript are depicted below the interaction schematics, and the union of credible SNP contact regions in the two tissues are depicted by the intermediately colored pink background.


rs6780942_117828678_C_T
Index
Credible
Cred. (Fxnl)
r2 = 1
r2 = 0.8
r2 = 0.6
r2 = 0.4
r2 = 0.2
r2 = 0
−log10(p)
0
2
4
6
8
10
116000
117000
118000
Chromosome 3 Position (kB)
TAD (CP)
HiC Map (CP)
TAD (GZ)
HiC Map (GZ)
3' UTR
Exon
Start Codon
5' UTR
Protein Coding
Pseudogene
non-coding RNA
Credible SNP HiC Contacts
LD-based Locus
RN7SL815P
LINC00903
TUSC7
BZW1P2
RNU5E-8P
RP11-768G7.1
C3orf30
LSAMP
LINC00901
RN7SL582P
RP11-484M3.5
MIR4447
RP11-768G7.2
RNU6-1200P
RP11-384F7.2
PTMAP8
IGSF11-AS1
RP11-768G7.3
RP11-384F7.1
LSAMP-AS1
IGSF11
116000
117000
118000
Chromosome 3 Position (kB)

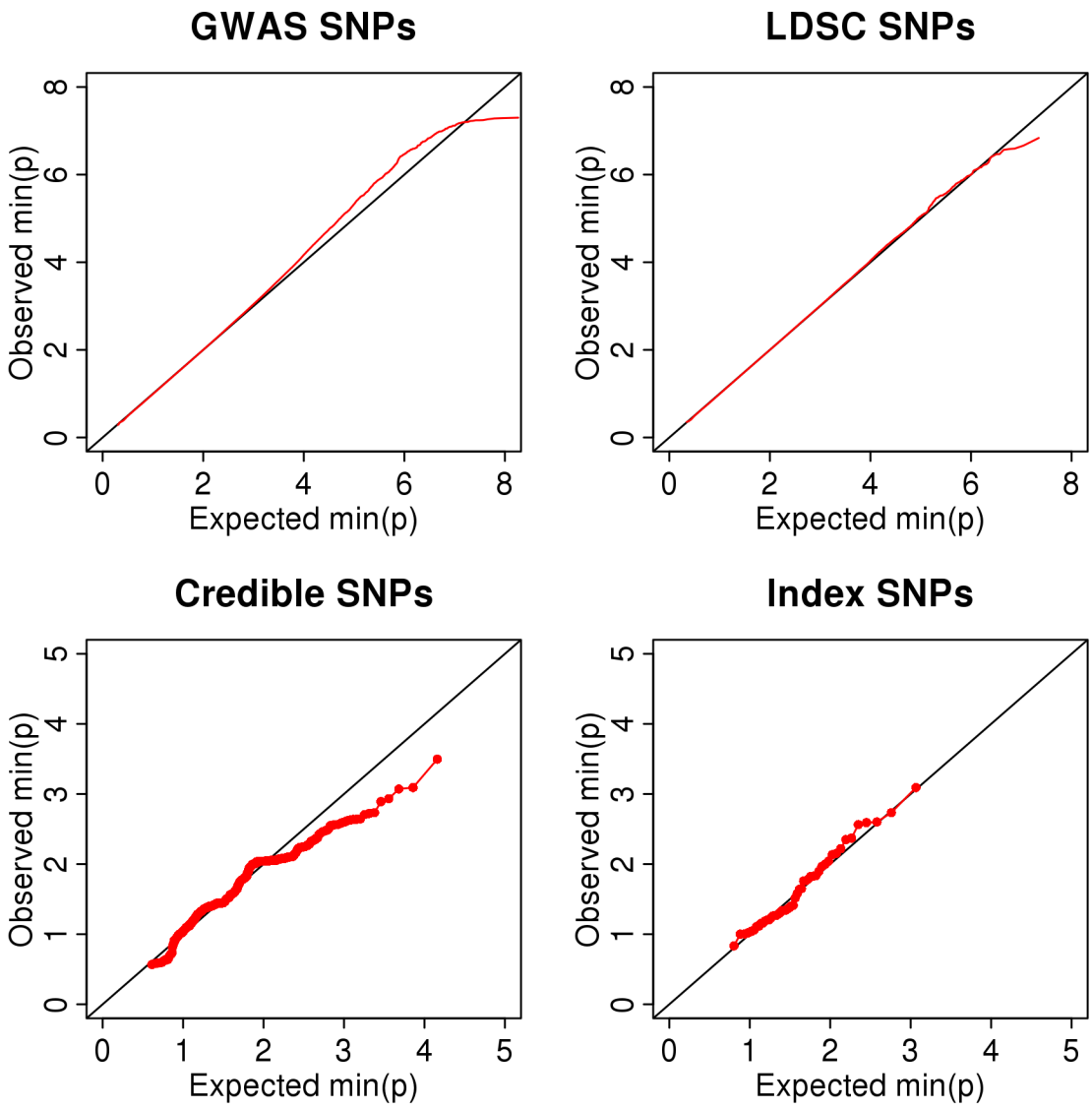
Supplementary Figure 17. Annotated Local Manhattan Plot for Locus 3. The top panel depicts the $-\log_{10}$(two-sided p-values) from the logistic regression association tests performed for the XDX GWAS (n=46,008 cases, 19,526 controls) for variants in the region of interest surrounding the implicated locus. The extent of the linkage disequilibrium (LD) defined locus implicated by the index SNP (black diamond) is contained within the pale-yellow background. The color coding of each variant represents the strength of LD with the index SNP. Credible SNPs are encircled and their position is indicated by a vertical black bar. The bottom panel shows a selected set of functional elements within the region of interest. Topological associated domains (TAD) are depicted by thick horizontal lines, offset slightly to note changes, for two fetal tissues: Cortical Plate (CO) and Germinal Zone (GZ). 10 kilo-base regions containing credible SNPs are depicted as black boxes and their HiC contact regions in each fetal tissue are defined by thin line-connected colored rectangles. Genes and functional elements from the longest transcript are depicted below the interaction schematics, and the union of credible SNP contact regions in the two tissues are depicted by the intermediately colored pink background.


rs12265655_106744534_T_C
Index
Credible
Cred. (Fxnl)
r2 = 1
r2 = 0.8
r2 = 0.6
r2 = 0.4
r2 = 0.2
r2 = 0
−log10(p)
0
2
4
6
8
10
106500
106750
107000
107250
Chromosome 10 Position (kB)
TAD (CP)
HiC Map (CP)
TAD (GZ)
HiC Map (GZ)
3' UTR
Exon
Start Codon
5' UTR
Protein Coding
Pseudogene
non-coding RNA
Credible SNP HiC Contacts
LD-based Locus
SORCS3-AS1
SORCS3
RNU6-463P
RP11-127O4.2
106500
106750
107000
107250
Chromosome 10 Position (kB)

Supplementary Figure 18. Annotated Local Manhattan Plot for Locus 4. The top panel depicts the $-\log_{10}$(two-sided p-values) from the logistic regression association tests performed for the XDX GWAS (n=46,008 cases, 19,526 controls) for variants in the region of interest surrounding the implicated locus. The extent of the linkage disequilibrium (LD) defined locus implicated by the index SNP (black diamond) is contained within the pale-yellow background. The color coding of each variant represents the strength of LD with the index SNP. Credible SNPs are encircled and their position is indicated by a vertical black bar. The bottom panel shows a selected set of functional elements within the region of interest. Topological associated domains (TAD) are depicted by thick horizontal lines, offset slightly to note changes, for two fetal tissues: Cortical Plate (CO) and Germinal Zone (GZ). 10 kilo-base regions containing credible SNPs are depicted as black boxes and their HiC contact regions in each fetal tissue are defined by thin line-connected colored rectangles. Genes and functional elements from the longest transcript are depicted below the interaction schematics, and the union of credible SNP contact regions in the two tissues are depicted by the intermediately colored pink background. Here, we did not identify a functional connection via our criteria, but implicate SORCS3 because of the substantial overlap with the gene body.



Supplementary Figure 19. SNP heritability estimated for each genotyping wave. iPSYCH subjects were genotyped in 23 waves, grouped according to birth year. To assess the potential of subtle genotyping batch effects, we computed the SNP heritability for each genotype wave. We used GREML in GCTA to estimate the SNP heritability (bar height) by considering all unrelated design control subjects. For a given wave, we considered all subjects genotyped in that wave as cases and all subjects from all other ways as controls. We repeated this analysis for each genotyping wave. To prevent upward bias in the average SNP heritability due to censoring of out of bounds (negative) variance component estimates (which may be expected when estimating variance components near 0) we used the –reml-no-constrain option. The average genotyping wave heritability is very close to 0 (0.001) and the variability appears qualitatively consistent with sampling variance around 0. Error bars denote estimate standard errors. Sample Sizes: Wave 1, n = 821 cases, 16200 controls; Wave 2, n = 2128 cases, 14893 controls; Wave 3, n = 794 cases, 16227 controls; Wave 4, n = 785 cases, 16236 controls; Wave 5, n = 741 cases, 16280 controls; Wave 6, n = 735 cases, 16286 controls; Wave 7, n = 642 cases, 16379 controls; Wave 8, n = 810 cases, 16211 controls; Wave 9, n = 805 cases, 16216 controls; Wave 10, n = 571 cases, 16450 controls; Wave 11, n = 555 cases, 16466 controls; Wave 12, n = 591 cases, 16430 controls; Wave 13, n = 593 cases, 16428 controls; Wave 14, n = 594 cases, 16427 controls; Wave 15, n = 635 cases, 16386 controls; Wave 16, n = 644 cases, 16377 controls; Wave 17, n = 631 cases, 16390 controls; Wave 18, n = 653 cases, 16368 controls; Wave 19, n = 688 cases, 16333 controls; Wave 20, n = 584 cases, 16437 controls; Wave 21, n = 640 cases, 16381 controls; Wave 22, n = 660 cases, 16361 controls; Wave 23, n = 721 cases, 16300 controls.


GWAS SNPs
LDSC SNPs
Credible SNPs
Index SNPs
Observed min(p)
Expected min(p)
0
1
2
3
4
5
6
8

Supplementary Figure 20. QQ-plots describing the per SNP genotyping wave associations. For each of 23 genotyping waves we performed a GWAS, treating individuals genotyped in one wave as cases and those genotyped in all other waves as controls. Association was tested via logistic regression and p-values were two-tailed. For each SNP, we selected the minimum p-value from the 23 wave GWAS, and create QQ-plots comparing the observed -log10(minimum p-value) on the y-axis, against the expected –log10(minimum p-value). We use the inverse of the cumulative distribution for the minimum of 23 independent uniform distributions to compute the distribution of expected minimum p-values. The CDF for the minimum of a draw from 23 identically and independently distributions, Y=min(p_1, p_2, …, p_{23}) is:

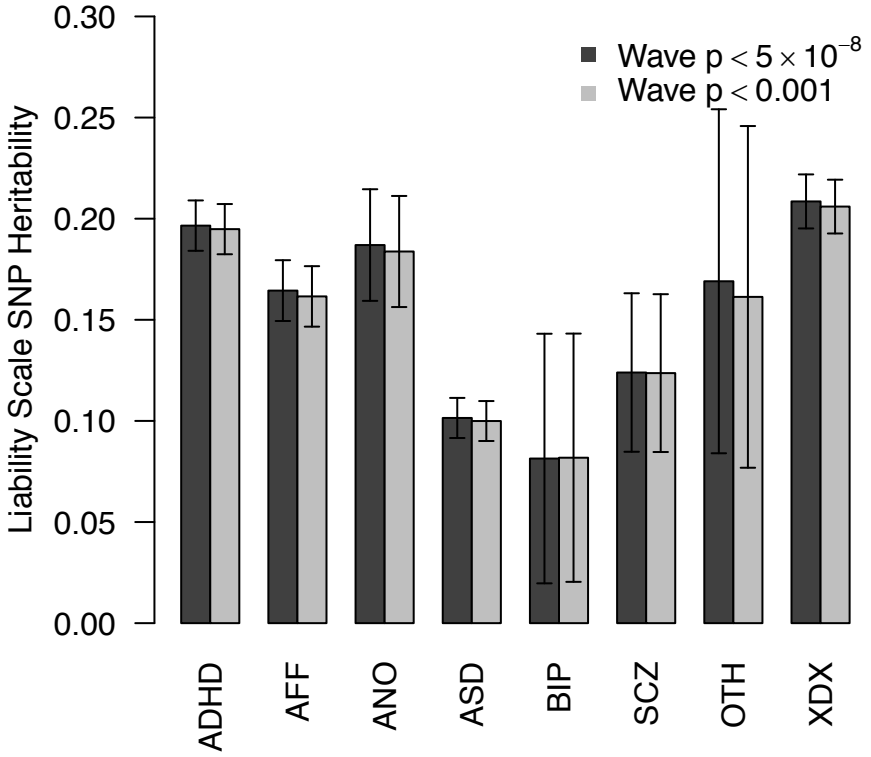
$$CDF(Y) = P(Y \leq y) = 1 - (1 - y)^n$$

Define $p_{(i)}$ as the i^{th} of m ordered minimum p-values, then the CDF of $p_{(i)}$ is i/(m). We estimate the expected value of $y=p_{(i)}$ as:

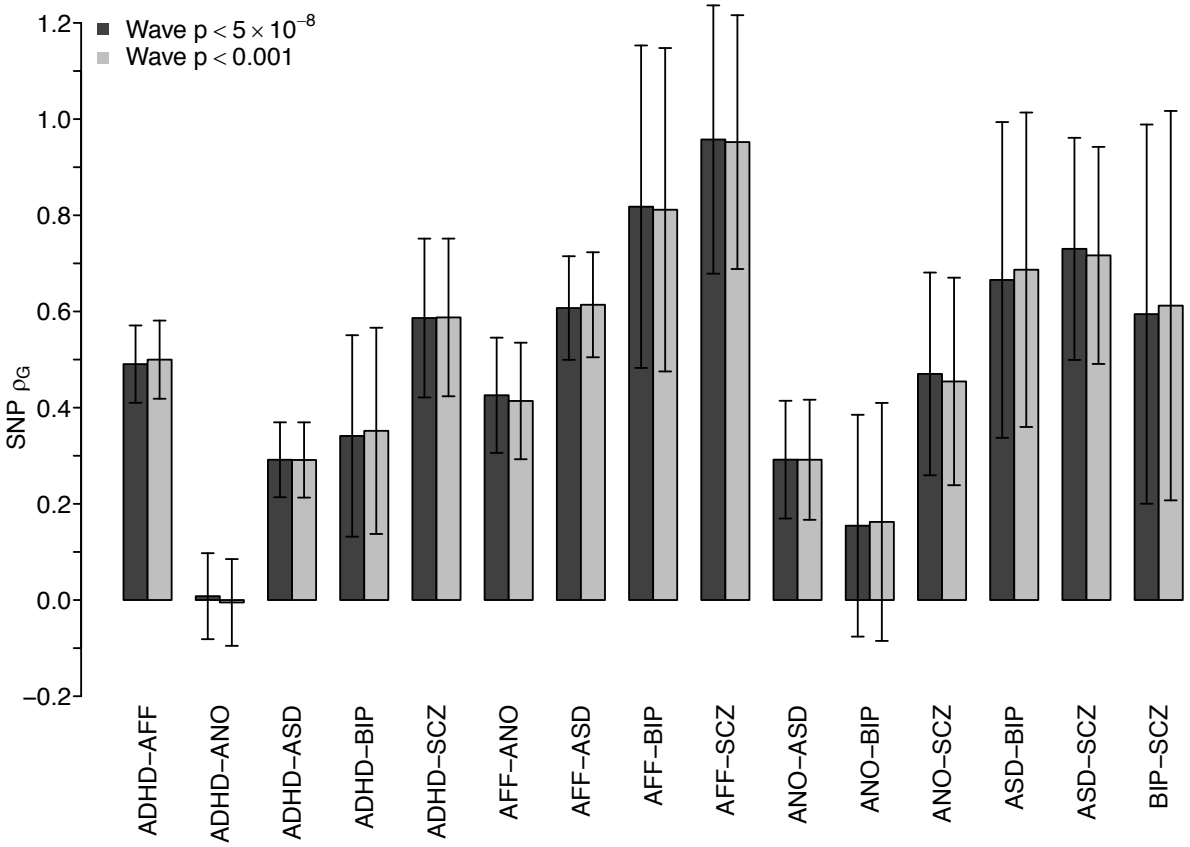
$$y = i^{th} \text{ ordered } \min(p) = p_{(i)} = 1 - (1 - \frac{i}{m})^{1/23}$$

The QQ-plots suggest there may be a slight abundance of wave associated SNPs, even after our initial removal of those with minimum p-values less than $5x10^{-8}$. This enrichment, however, does not seem present in the SNPs used for LDSC analysis, the credible set implicated by our XDX GWAS (m=627), nor the XDX index SNPs (m=50).

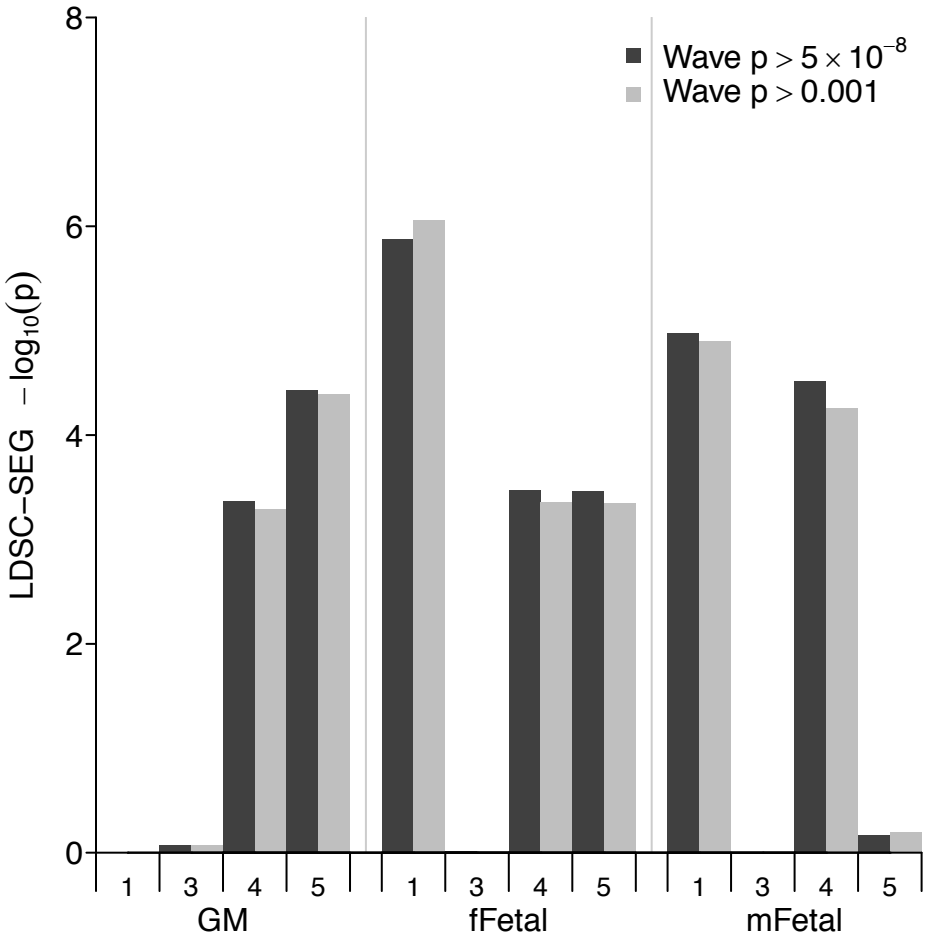
Sample Sizes: Wave 1, n = 1051 cases, 21900 controls; Wave 2, n = 2772 cases, 20179 controls; Wave 3, n = 1060 cases, 21891 controls; Wave 4, n = 1043 cases, 21908 controls; Wave 5, n = 1013 cases, 21938 controls; Wave 6, n = 999 cases, 21952 controls; Wave 7, n = 892 cases, 22059 controls; Wave 8, n = 1123 cases, 21828 controls; Wave 9, n = 1066 cases, 21885 controls; Wave 10, n = 758 cases, 22193 controls; Wave 11, n = 792 cases, 22159 controls; Wave 12, n = 813 cases, 22138 controls; Wave 13, n = 811 cases, 22140 controls; Wave 14, n = 790 cases, 22161 controls; Wave 15, n = 858 cases, 22093 controls; Wave 16, n = 836 cases, 22115 controls; Wave 17, n = 810 cases, 22141 controls; Wave 18, n = 911 cases, 22040 controls; Wave 19, n = 919 cases, 22032 controls; Wave 20, n = 829 cases, 22122 controls; Wave 21, n = 876 cases, 22075 controls; Wave 22, n = 914 cases, 22037 controls; Wave 23, n = 1015 cases, 21936 controls.



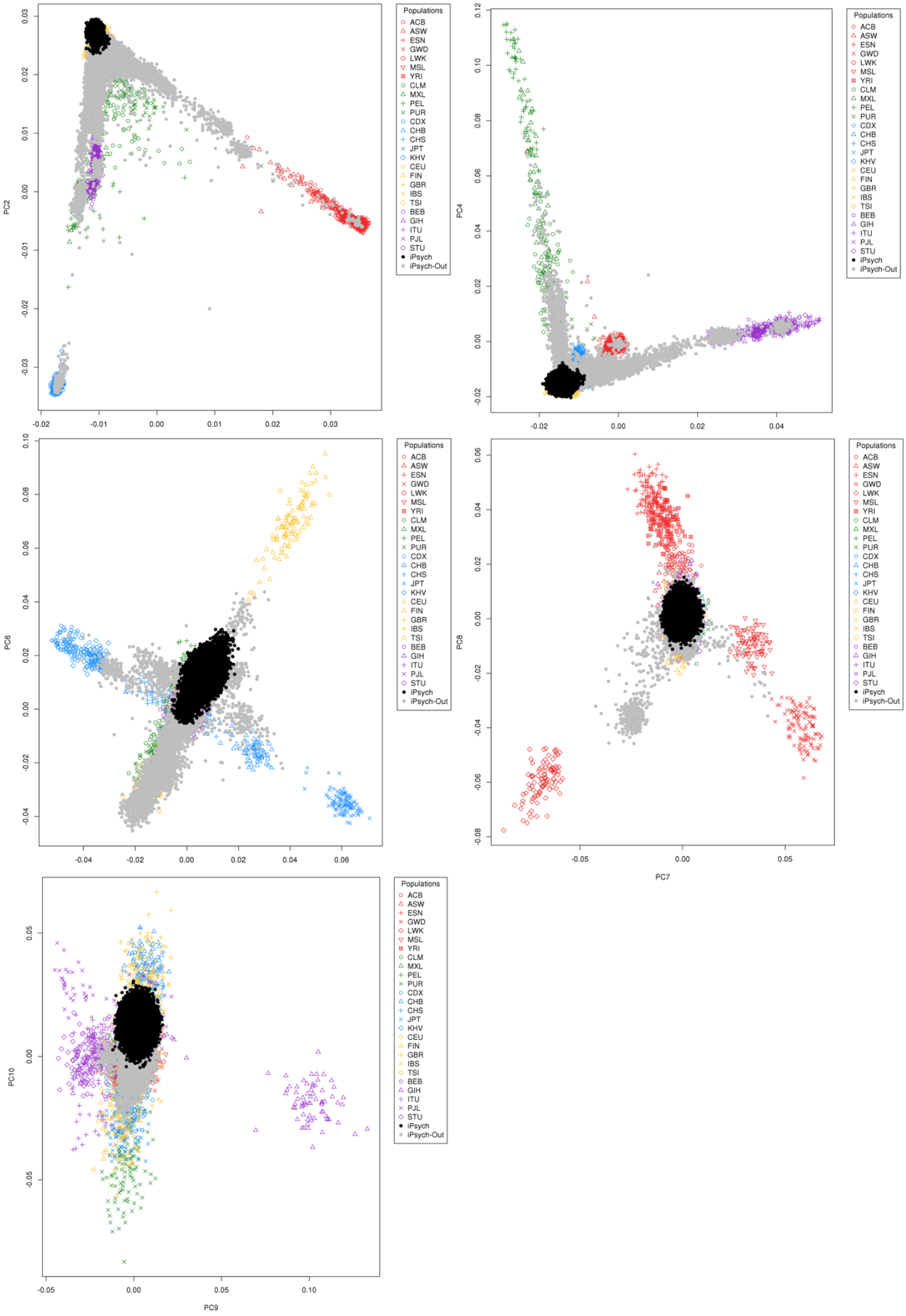
Supplementary Figure 21. Estimates of SNP heritability are not changed by censoring SNPs more stringently for wave effects. Here we repeated the GCTA SNP-heritability analysis for each individual indication more strictly censoring SNPs with potential genotyping wave effects (light grey bars), but do not observe an appreciable difference from the original heritability estimates (dark grey bars) presented in Figure 1A. Error bars denote standard errors of heritability point estimates. Samples sizes are as described for figure 1A in supplementary table 3.



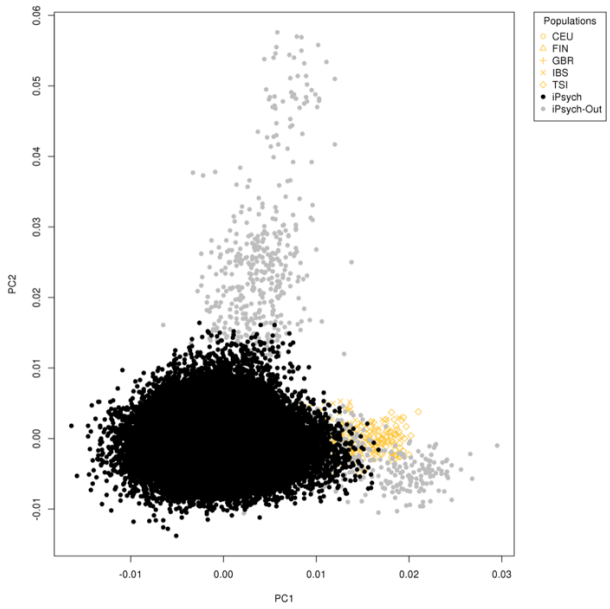
Supplementary Figure 22. Estimates of SNP genetic correlation are not changed by censoring SNPs more stringently for wave effects. Here we repeated the GCTA SNP-genetic correlation analysis more strictly censoring SNPs with potential genotyping wave effects (light grey bars), but do not observe an appreciable difference from the original heritability estimates (dark grey bars) presented in Figure 1D. Error bars denote standard errors of heritability point estimates. Samples sizes are as described for figure 1D in supplementary table 6.

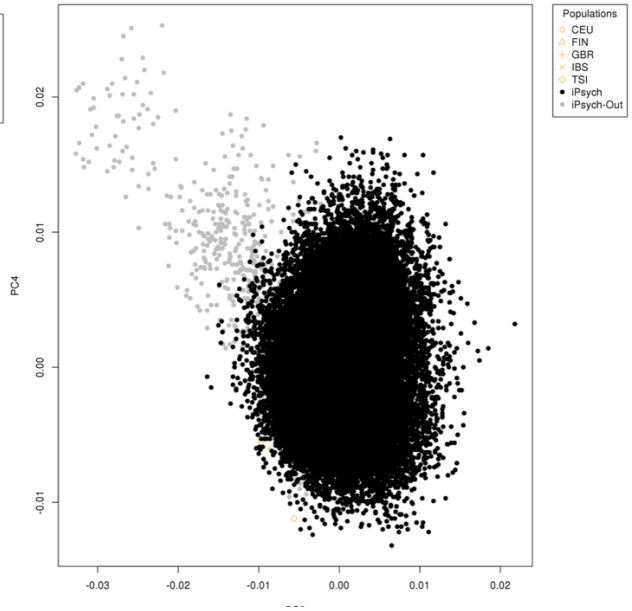


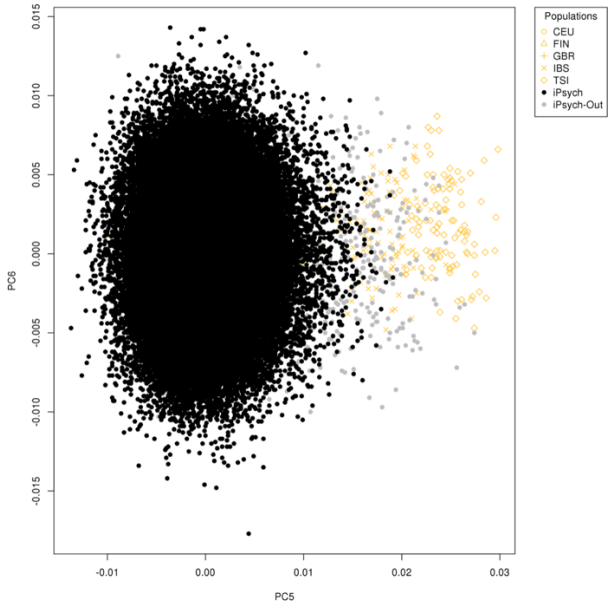
Supplementary Figure 23. Estimates of fetal brain heritability enrichment via LDSC-SEG are not changed by censoring SNPs more stringently for wave effects. Here we repeated the LDSC-SEG enrichment tests in the XDX GWAS (n=46,008 cases, 19,526 controls) for the fetal brain annotations, more strictly censoring SNPs with potential genotyping wave effects (light grey bars), but do not observe an appreciable difference from the original heritability estimates (dark grey bars) presented in Figure 4A.

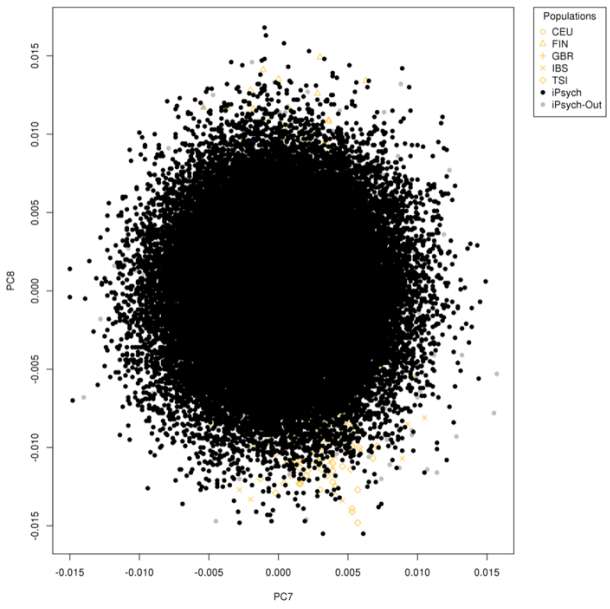

PC2
PC4
PC6
PC8
PC7
PC10
PC9
Populations
ACB
ASW
ESN
GWD
LWK
MSL
YRI
CLM
MXL
PEL
PUR
CDX
CHB
CHS
JPT
KHV
CEU
FIN
GBR
IBS
TSI
BEB
GIH
ITU
PJL
STU
iPsych
iPsych-Out

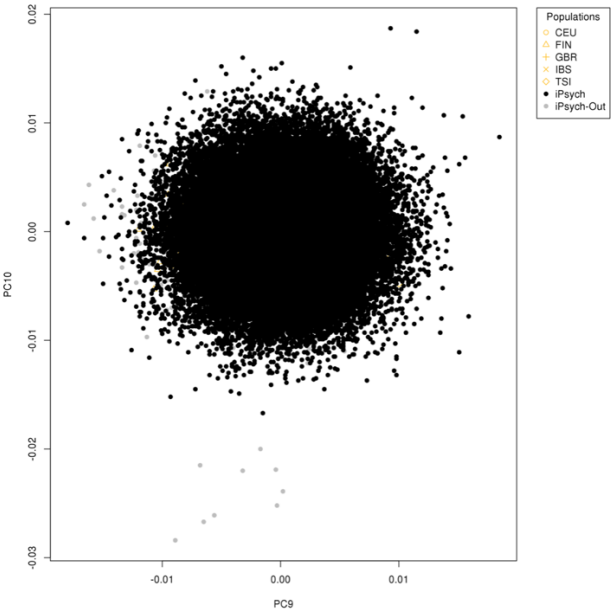
Supplementary Figure 24. First ten PCs of genetic similarity for the iPSYCH sample passing basic QC. Each panel plots iPSYCH individuals (N=77,686) according to their estimated principal components. Black dots are subjects which pass our global ancestry QC (N=71,212) and grey dots denote excluded iPSYCH subjects (N=6,474). To determine outliers we computed the mean and standard deviation for each of the first 10 PCs plotted above, using only the subjects with both parents and all four grandparents born in Denmark (N=47,586). For each subject (N=77,686) we computed the sample's Mahalanobis distance from the multivariate mean of the joint distribution of the first 10 PCs in the Danish birth group. Assuming multivariate normality, the Mahalanobis distance follows a chi-square distribution with degrees of freedom equal to the number of dimensions (here χ_{10}^2). We flagged subjects as global ancestry outliers if their distance had a probability less than 5.73×10^{-7} under the chi-square, ten degree of freedom model (the same probability as 5 standard deviations from the mean for a single normal variable). Colored symbols are reference individuals taken from the 1000 genomes project (N=2,447) with known ancestry for whom PCs were estimated according to the same procedure.


Populations
CEU
FIN
GBR
IBS
TSI
iPsych
iPsych-Out
PC1
PC2


Populations
CEU
FIN
GBR
IBS
TSI
iPsych
iPsych-Out
PC4


Populations
CEU
FIN
GBR
IBS
TSI
iPsych
iPsych-Out
PC5
PC6


Populations
CEU
FIN
GBR
IBS
TSI
iPsych
iPsych-Out
PC7
PC8


Populations
CEU
FIN
GBR
IBS
TSI
iPsych
iPsych-Out
PC9
PC10

Supplementary Figure 25. First ten PCs of genetic similarity for the iPSYCH sample passing our global ancestry QC. Each panel plots iPSYCH individuals (N=71,212) according to their estimated principal components. Black dots are subjects which pass our local ancestry QC (N=71,212) and grey dots denote excluded iPSYCH subjects (N=689). To determine outliers we computed the mean and standard deviation for each of the first 10 PCs plotted above, using only the subjects with both parents and all four grandparents born in Denmark (N=47,586). For each subject (N=71,212) we computed the sample's Mahalanobis distance from the multivariate mean of the joint distribution of the first 10 PCs in the Danish birth group. Assuming multivariate normality, the Mahalanobis distance follows a chi-square distribution with degrees of freedom equal to the number of dimensions (here χ_{10}^2). We flagged subjects as global ancestry outliers if their distance had a probability less than 5.73×10^{-7} under the chi-square, ten degree of freedom model (the same probability as 5 standard deviations from the mean for a single normal variable). Colored symbols are European reference individuals taken from the 1000 genomes project (N=483) with known ancestry for whom PCs were estimated according to the same procedure.