REPORT

Pleiotropic effects of trait-associated genetic variation on DNA methylation: utility for refining GWAS loci.

Eilis Hannon,[1] Mike Weedon,[1] Nicholas Bray,[2] Michael O'Donovan,[2] Jonathan Mill[1*]

[1] University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK

[2] MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, CF24 4HQ, UK

[*] Jonathan Mill, University of Exeter Medical School, RILD Building, Royal Devon & Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK. J.Mill@exeter.ac.uk

Abstract

Most genetic variants identified in genome-wide association studies (GWAS) of complex traits are thought to act via effects on gene regulation rather than directly altering the protein product. As a consequence, the actual genes involved in disease are not necessarily the most proximal to the associated variants. By integrating data from GWAS analyses with that from genetic studies of regulatory variation, it is possible to identify variants pleiotropically-associated with both a complex trait and measures of gene regulation. In this study, we use Summary data–based Mendelian Randomization (SMR), a method developed to identify variants pleiotropically associated with both complex traits and gene expression, to identify variants that are associated with complex traits and DNA methylation. We use large DNA methylation quantitative trait loci (mQTL) datasets generated from two different tissues (blood and fetal brain) to prioritize genes for >40 complex traits with robust GWAS data, highlighting considerable overlap with the results of SMR analyses performed using expression QTL (eQTL) data. We identify multiple examples of variable DNA methylation associated with GWAS variants for a range of complex traits, demonstrating the utility of this approach for refining genetic association signals.

There has been major progress in the identification of genetic variants influencing a diverse range of complex human phenotypes including anthropometric measures (e.g. height and weight)[1; 2], cardiovascular disease[3; 4], inflammatory disorders[5], neurological diseases[6; 7] and psychiatric illness[8-10]. The challenge is now to improve our understanding of the biological effects of these genetic risk factors, especially because the actual gene(s) involved in mediating phenotypic variation are not necessarily the most proximal to the lead SNPs identified in genome-wide association studies (GWAS). Supported by the observation that GWAS variants are preferentially located in enhancers and regions of open chromatin[11; 12], the majority of common genetic risk factors are predicted to influence gene regulation rather than directly affecting the coding sequences of transcribed proteins[13].

Expression quantitative trait loci (eQTL) have been successfully used to investigate the functional consequences of GWAS variants[14; 15]. The co-localization of GWAS and eQTL variants, however, is not sufficient to show that the overlapping association signals are causally related, as the association signals may be tagging different causal variants in the same linkage disequilibrium (LD) block. Recently, an approach called Summary data–based Mendelian Randomization (SMR) was proposed as a strategy for identifying overlapping genetic signals associated with both phenotypic and transcriptional variation, subsequently distinguishing pleiotropic effects - i.e. where the same variant is influencing both outcomes, although not necessarily dependently - from those that are an artefact of LD[16]. Genetic effects on gene expression can be mediated by epigenetic processes including changes in DNA methylation, a modification to cytosine that has an essential role in mammalian development[17]. We have previously demonstrated the utility of DNA methylation QTLs (mQTLs) for interpreting GWAS findings by identifying specific examples where genetic polymorphisms associated with schizophrenia [MIM: 181500] co-localize with variants associated with DNA methylation[18; 19]. In this study we apply the SMR approach to test 35,263 DNA methylation sites against 43 complex phenotypes with robust GWAS data (**Table S1**) using mQTLs identified in our recent analysis of methylomic variation in whole blood and imputed single nucleotide polymorphism (SNP) genotypes (n = 639; mQTL $P < 1x10^{-10}$; a full description of this dataset, referred to as Phase 1, can be found here[19]) in conjunction with publicly-available summary data from a series of well-powered GWAS analyses.

The first stage of the SMR analysis identifies the most significantly associated SNP for a DNA methylation site (that is also present in the GWAS dataset) as an instrumental variable to test for an association with a phenotype by using the two-step least squares (2SLS) approach comparing the coefficients from the mQTL with those from a GWAS of the phenotype using the same SNP (**Figure S1a**). This approach identified 1,932 associations ($P < 1.42x10^{-6}$ corrected for 35,263 DNA methylation sites) between 31 complex traits and 1,354 individual DNA methylation sites (**Table S2** and **Figure S2**). Because these associations may be driven by two highly-correlated but different causal variants for the GWAS trait and DNA methylation, the second stage of the SMR approach repeats the analysis with alternative SNPs associated with DNA methylation as the instrument and performs a heterogeneity in dependent instruments (HEIDI) test for heterogeneity in the resulting association statistics. If there is a single causal variant associated with both phenotype and DNA methylation, the association statistics will be identical regardless of the selected instrument (**Figure S1b**) and the HEIDI *P*-value will be non-significant. In contrast, if there are two separate causal variants, each correlated with the instrument, there will be variation in the results from different instruments (**Figure S1c**) indicated by a significant HEIDI *P*-value. It should be noted that this approach is unable to distinguish these two scenarios if the two causal variants are in perfect LD, with power inversely proportional to the strength of the correlation between the two causal variants. Furthermore, the assumptions underlying Mendelian Randomization [20] also apply to SMR, and it is possible that variants may act through mechanisms such as horizontal pleiotropy.

By identifying non-significant heterogeneity (HEIDI $P > 0.05$), we identified a refined set of 625 associations between 28 complex traits and 440 DNA methylation sites (**Table S2**), which can be described as pleiotropic. We were able to test 581 of these associations with mQTL generated from a second independent whole blood dataset (n = 665; a description of this cohort, referred to as Phase 2, can be found here[19]). A highly significant proportion (99.2%; sign test $P = 1.47x10^{-172}$) had the same direction of association across the two datasets (**Figure S3**) with a large proportion (n = 337; 58.0%) satisfying the criteria for a pleiotropic association ($P < 1.04x10^{-6}$ and HEIDI $P > 0.05$) in the replication dataset also. Out of the GWAS traits tested, height is characterized by the most associations (n = 193), an unsurprising

observation given that is the most highly-powered GWAS with the largest number of GWAS-significant loci (n = 423). Power for SMR analysis is influenced by the power of the GWAS, which differs for each trait considered, making comparisons between traits relatively difficult.

As demonstrated in its original implementation for eQTLs, the SMR approach based on mQTLs has the potential to nominate loci that currently do not have sufficient statistical power to obtain genome-wide significance based on the GWAS data alone but which represent candidates for future genetic studies (**Table S3**). Our SMR analysis of Tanner staging of puberty, for example, identified DNA methylation sites in nine independent loci (annotated to *APEH*[MIM:102645], *SYNJ2*[MIM:609410], *IDO2*[MIM:612129], *PDZRN4*[MIM: 609730], *HTR2A*[MIM: 182135], *CTDP1*[MIM: 604927], *RAE1*[MIM: 603343], and non-genic regions on chromosome 4 and 16) that do not have a genome-wide significant ($P < 5x10^{-8}$) variant within 0.5Mb in the GWAS[21] (**Figure S4**). In some genomic regions, DNA methylation sites annotated to different genes are associated with the same phenotype; for example on chromosome 15, sites annotated to *CHRNA5*[MIM: 118505] and *PSMA4*[MIM: 176846] are associated with the number of cigarettes smoked per day (**Figure S5**), and on chromosome 17, sites annotated to *ERBB2*[MIM: 164870] and *PGAP3*[MIM: 611801] are associated with total cholesterol (**Figure S6**). Furthermore, 130 DNA methylation sites were found to be associated with multiple complex traits (**Table S4**; range = 2 to 6 traits). In many cases these overlaps are consistent with either reported phenotypic correlations (e.g. cg24631222 and cg04140906 annotated to *CHRNA5* are associated with both schizophrenia and the number of cigarettes smoked per day (**Figure S7**), two traits that are epidemiologically linked[22; 23]) or shared genetic architecture (e.g. cg10583485, annotated to *DOCK7*[MIM: 615730] (**Figure S8**), is associated with LDL, triglycerides and total cholesterol, three traits characterized by a strong genetic correlation[24]). Because genetic correlations could account for some of the overlap between traits we factored in genetic correlations derived using LD Score regression[24], showing that 30 out of the 70 pairs of traits with at least one associated DNA methylation site in common are actually characterized by a genetic correlation < 0.2 (**Figure S9**).

Multiple DNA methylation sites can be annotated to a single gene, and in total we identified 337 gene-trait pleiotropic associations with a mean of 1.46 sites associated per gene (range = 1-11). These overlapping

associations between a particular complex trait and a gene would not be expected to be associated in the same direction as correlation of DNA methylation across a gene is not always positive, and were not for 20 out of the 31 gene–trait associations involving genes with multiple annotated DNA methylation sites. To add further support to the genes prioritised at GWAS loci using blood mQTL data, we aligned these results with SMR analyses performed on publically available whole blood eQTL data (n = 5,311; $P < 5 \times 10^{-8}$) described in detail in a recent paper by Westra and colleagues[15]. We identified an overlapping set of 2,724 genes that were i) annotated to DNA methylation sites influenced by significant mQTLs (involving 7,722 distinct DNA methylation sites) and ii) also transcriptionally influenced by variation at significant eQTLs (involving 2,770 gene expression microarray probes), making them suitable for testing in the SMR framework. It should be noted that one limitation to assessing the relationship between mQTLs and eQTLs is that DNA methylation sites, like SNPs, are annotated to genes based on their location; therefore a lack of overlap in the associations with a particular gene from the SMR analyses of DNA methylation and gene expression should not necessarily be interpreted as inconsistent evidence. Furthermore, the differences in the sample sizes used to generate the mQTL and eQTL datasets may result in different statistical power to detect QTLs. 86 of the 337 (25.5%) pleiotropic gene-trait associations identified with mQTLs were also tested with eQTLs in the SMR framework (**Figure S10**). Of these 27 (31.4%) also meet the criteria to be defined as representing pleiotropic associations between the trait and gene expression (SMR $P < 8.38\text{x}10^{-6}$ corrected for 5,966 gene expression probes and HEIDI $P > 0.05$) involving 17 complex traits associated with expression at 16 genes (**Table S5**). An example of an overlapping mQTL and eQTL signal for *RNASET2* [MIM: 612944] on chromosome 6 is presented in **Figure 1;** both *RNASET2* expression (SMR $P = 6.04\text{x}10^{-8}$) and DNA methylation at two CpG sites in the first intron of the gene (cg25258033: SMR $P = 2.84\text{x}10^{-10}$; cg25258033: SMR $P = 2.50\text{x}10^{-10}$) are associated with Crohn's disease [MIM: 266600].

Given the tissue-specific and developmentally-dynamic nature of gene regulation we were next interested in examining the consistency of our findings in a different tissue, repeating the SMR analysis on mQTL identified in our recent analysis of human fetal brain (n = 166; mQTL $P < 1\text{x}10^{-8}$; a detailed description of this dataset can be found here[18]). The majority (75.4%) of SNP-DNA methylation relationships identified for SMR analysis in whole blood are characterized by a consistent direction of effect when tested in fetal brain

(sign test $P = 4.94x10^{-324}$; **Figure S11**). Despite the strong concordance of mQTL effects across tissues, the smaller number of samples used to generate the fetal brain dataset (n = 166) means that only a subset (4,691 (13.3%)) of these mQTL associations passed our mQTL significance threshold ($P < 1x10^{-8}$) and were included in the subsequent SMR analyses; almost all of these (96.0%; sign test $P < 2.2x10^{-308}$) were characterized by the same direction of effect in both tissues (**Figure S12**). 84 (13.5%) of the 625 pleiotropic associations between identified with whole blood mQTLs involved a DNA methylation site that also had a significant fetal brain mQTL ($P < 1x10^{-8}$) meaning it could be tested with the SMR framework (**Figure S13**). 35 of these 84 (41.7%) pleiotropic associations met the criteria (i.e. SMR $P < 5.40 \times 10^{-6}$ corrected for 9,265 DNA methylation sites tested and HEIDI $P > 0.05$) to be defined as also having a pleiotropic association in fetal brain involving9 complex traits (**Table S6**). While six (17.1%) of the site-trait associations involve brain-related phenotypes (5 for schizophrenia and 1 for migraine [MIM: 157300]), the majority (82.9%) involve traits which are presumed to affect other tissues (e.g. total cholesterol and Crohn's disease), suggesting that effects are common across tissues. **Figure 2** summarizes SMR analysis across the *HEY2-NOCA7* region on chromosome 6 implicated in a recent GWAS of migraine[25]. Manhattan plots for the genetic analysis of cg05901451 located in the 5'UTR of *HEY2* [MIM: 604674] in whole blood and fetal brain show a highly comparable profile to the GWAS of migraine, consistent with overlapping genetic signals influencing DNA methylation in both tissues and migraine [25].

Finally, comparing the SMR results across multiple complex traits gives a potential insight into shared pleiotropic associations between pairs of traits. We performed hierarchical clustering of SMR results for 38 complex traits, selected because they were tested against a minimum of 20,000 DNA methylation sites, to identify consistent signatures (**Figure S14)**. **Figure 3**, for example, depicts the association statistics for 43 DNA methylation sites associated with Crohn's disease (SMR $P < 1.42x10^{-6}$) across all 38 phenotypes; interestingly, we observe a highly concordant profile between Crohn's disease and ulcerative colitis across all associated sites, consistent the strong genetic correlation between these traits (**Figure S9**); the SMR results may highlight which genes are characterized by shared effects between traits. There is also a notable overlap with BMI, waist and hip circumference at specific loci (i.e. *ATP2A1*[MIM: 108730], *SULT1A2*

[MIM: 601292] and *SBK1*[MIM: 300374]), an interesting observation given the negligible genetic correlations between these traits and Crohn's disease.

Taken together, these analyses demonstrate the utility of the SMR approach for identifying instances where complex traits and variable DNA methylation are pleiotropically-associated with genetic variation. This approach may facilitate our understanding of the functional consequences of genetic risk variants for a range of complex traits and may facilitate the localization and prioritization of specific genes within genomic regions identified by GWAS.

Supplemental Data

Supplemental Data include twelve figures and six tables and can be found with this article online at XXXXXXX.

Acknowledgements

This work was funded by a grant from the UK Medical Research Council (MRC; MR/K013807/1) to JM.

None of the authors have any competing financial interests to declare.

Web Resources

ChunkChromosome, http://genome.sph.umich.edu/wiki/ChunkChromosome

Fetal brain mQTL data, http://epigenetics.essex.ac.uk/mQTL/All_Imputed_BonfSignificant_mQTLs.csv.gz

Online Mendelian Inheritance in Man, http://www.omim.orgSMR, http://cnsgenomics.com/software/smr/download.html

Whole blood eQTL data, http://cnsgenomics.com/software/smr/westra_eqtl_hg19.zip

Whole blood mQTL data, http://epigenetics.essex.ac.uk/schizophrenia/BloodmQTL.csv.tgz

References

1. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197-206.
2. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46, 1173-1186.
3. CARDIoGRAMplusC4D Consortium, Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 47, 1121-1130.
4. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C., et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 43, 333-338.
5. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 47, 979-986.
6. Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45, 1452-1458.
7. van Rheenen, W., Shatunov, A., Dekker, A.M., McLaughlin, R.L., Diekstra, F.P., Pulit, S.L., van der Spek, R.A., Võsa, U., de Jong, S., Robinson, M.R., et al. (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet.
8. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H., Boomsma, D.I., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 18, 497-511.
9. Schizophrenia Working Group of the PGC, Ripke, S., Neale, B., Corvin, A., Walters, J., Farh, K., Holmans, P., Lee, P., Bulik-Sullivan, B., Collier, D., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-+.
10. Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 43, 977-983.
11. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. Genome Res 22, 1748-1759.
12. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43-49.
13. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190-1195.
14. Li, M., Jaffe, A.E., Straub, R.E., Tao, R., Shin, J.H., Wang, Y., Chen, Q., Li, C., Jia, Y., Ohi, K., et al. (2016). A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. Nat Med 22, 649-656.
15. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 45, 1238-1243.
16. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48, 481-487.
17. Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev 16, 6-21.
18. Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T.M., Troakes, C., Turecki, G., O'Donovan, M.C., Schalkwyk, L.C., et al. (2015). Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci.

19. Hannon, E., Dempster, E., Viana, J., Burrage, J., Smith, A.R., Macdonald, R., St Clair, D., Mustard, C., Breen, G., Therman, S., et al. (2016). An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. Genome Biol 17, 176.
20. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32, 1-22.
21. Cousminer, D.L., Stergiakouli, E., Berry, D.J., Ang, W., Groen-Blokhuis, M.M., Körner, A., Siitonen, N., Ntalla, I., Marinelli, M., Perry, J.R., et al. (2014). Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. Hum Mol Genet 23, 4452-4464.
22. de Leon, J., and Diaz, F.J. (2005). A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. Schizophr Res 76, 135-157.
23. McClave, A.K., McKnight-Eily, L.R., Davis, S.P., and Dube, S.R. (2010). Smoking characteristics of adults with selected lifetime mental illnesses: results from the 2007 National Health Interview Survey. Am J Public Health 100, 2464-2472.
24. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236-1241.
25. Gormley, P., Anttila, V., Winsvold, B.S., Palta, P., Esko, T., Pers, T.H., Farh, K.H., Cuenca-Leon, E., Muona, M., Furlotte, N.A., et al. (2016). Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet 48, 856-866.

Figure Legends

Figure 1. Summary data–based Mendelian Randomization (SMR) analysis using quantitative trait loci associated with DNA methylation (mQTL) and gene expression (eQTL) implicates a role for *RNASET2* in Crohn's disease. Shown is a genomic region on chromosome 6 (hg19:167243095- 167565882) identified in a recent GWAS of Crohn's disease performed by Liu et al.[5]. Genes located in this region are shown at the top, with exons indicated by thicker bars and the red arrows indicating the direction of transcription. DNA methylation sites interrogated by the Illumina 450K array are indicated by solid vertical lines underneath the genes. The four bottom panels depict the –log10 *P* value (y-axis) against genomic location (x-axis) from **A)** SMR analysis (where squares represent Illumina 450K array DNA methylation sites and triangles represent gene expression probes, with green and red highlighting those with non-significant HEIDI test for DNA methylation and gene expression respectively), **B)** blood mQTL (n = 639) for the DNA methylation site cg25258033 (outlined in black in panel **A**), **C**) blood eQTL (n = 5,311) results for ILMN1671565 (outlined in black in panel **A**)), and **D)** the GWAS of Crohn's disease performed by Liu et al.[5].

Figure 2. Summary data–based Mendelian Randomization (SMR) analysis using whole blood and fetal brain mQTL data implicates a role for the *HEY2* locus in migraine. Shown is a genomic region on chromosome 6 (hg19: 125970800-126170800) identified in a recent GWAS of migraine performed by Gormley et al.[25]. Genes located in this region are shown at the top, with exons indicated by thicker bars and the red arrows indicating the direction of transcription. The four bottom panels depict the –log10 *P* value (y-axis) against genomic location (x-axis) from **A)** SMR analysis (where points represent DNA methylation sites interrogated by the Illumina 450K array; black and grey indicates SMR tests from blood (squares) or fetal brain (diamonds) mQTL respectively with green squares and blue diamonds highlighting those with a non-significant HEIDI test for blood and fetal brain respectively), mQTL results for the DNA methylation site cg05901451 (outlined in black in panel **A**)) in **B)** blood (n = 639) and **C**) fetal brain (n = 166) and **D)** the GWAS of migraine performed by Gormley et al (n = 59,674 cases and 316,078 controls)[25].

Figure 3. Heat-map of the SMR results for 32 DNA methylation sites associated with Crohn's disease (SMR $P < 1.38x10^{-6}$ and HEIDI $P > 0.05$) across 38 GWAS datasets. Each square in the heat-map represents the t-statistic (b_SMR/se_SMR) of the GWAS trait (columns) for a DNA methylation site (row; n = 32) associated with Crohn's Disease. Only phenotypes (n = 38) tested against at least 20,000 DNA methylation sites were included in this comparison.

Table S1. Details of all publically available GWAS results included in SMR analysis.

Table S2. Results from SMR analysis for all significant associations ($P < 1.42x10^{-6}$) between complex traits and DNA methylation.

Table S3. Associations from SMR analysis where no SNP within 0.5Mb has obtained genome-wide significance ($P < 5x10^{-8}$) in GWAS. These loci therefore represent associations worthy of further investigation for that complex trait.

Table S4. Table of overlapping associations with specific DNA methylation sites between complex traits; entries above the diagonal detail the specific Illumina 450K array probe and entries below the diagonal detail the gene that probe is located within.

Table S5. Table of gene-trait associations identified with SMR analysis using whole blood mQTL and blood eQTL.

Table S6. Table of DNA methylation sites associated through mQTL identified in both whole blood and fetal brain.

Figure 1

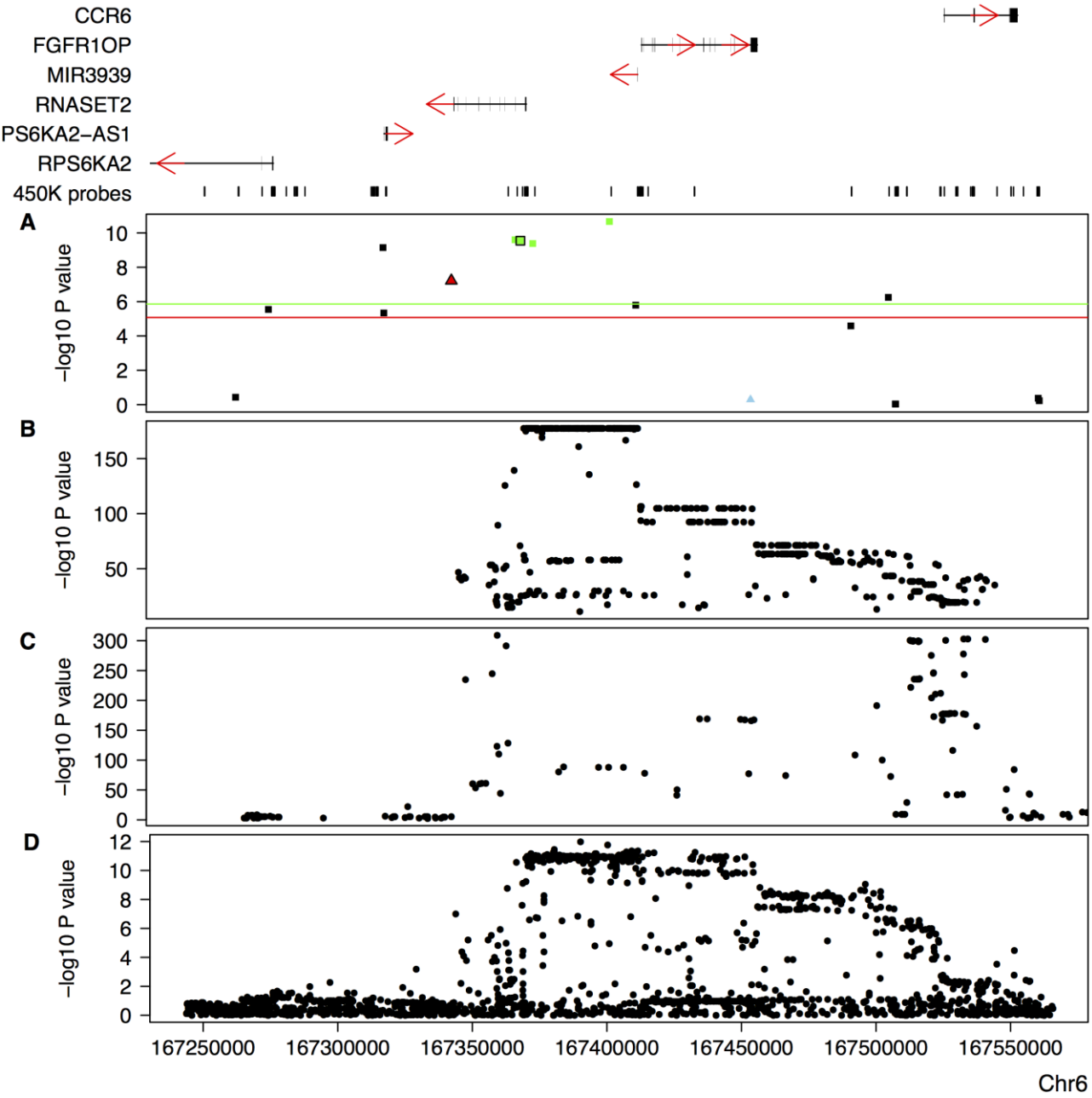

CCR6
FGFR1OP
MIR3939
RNASET2
PS6KA2−AS1
RPS6KA2
450K probes
A
B
C
D
−log10 P value
167250000
167300000
167350000
167400000
167450000
167500000
167550000
Chr6

Figure 2

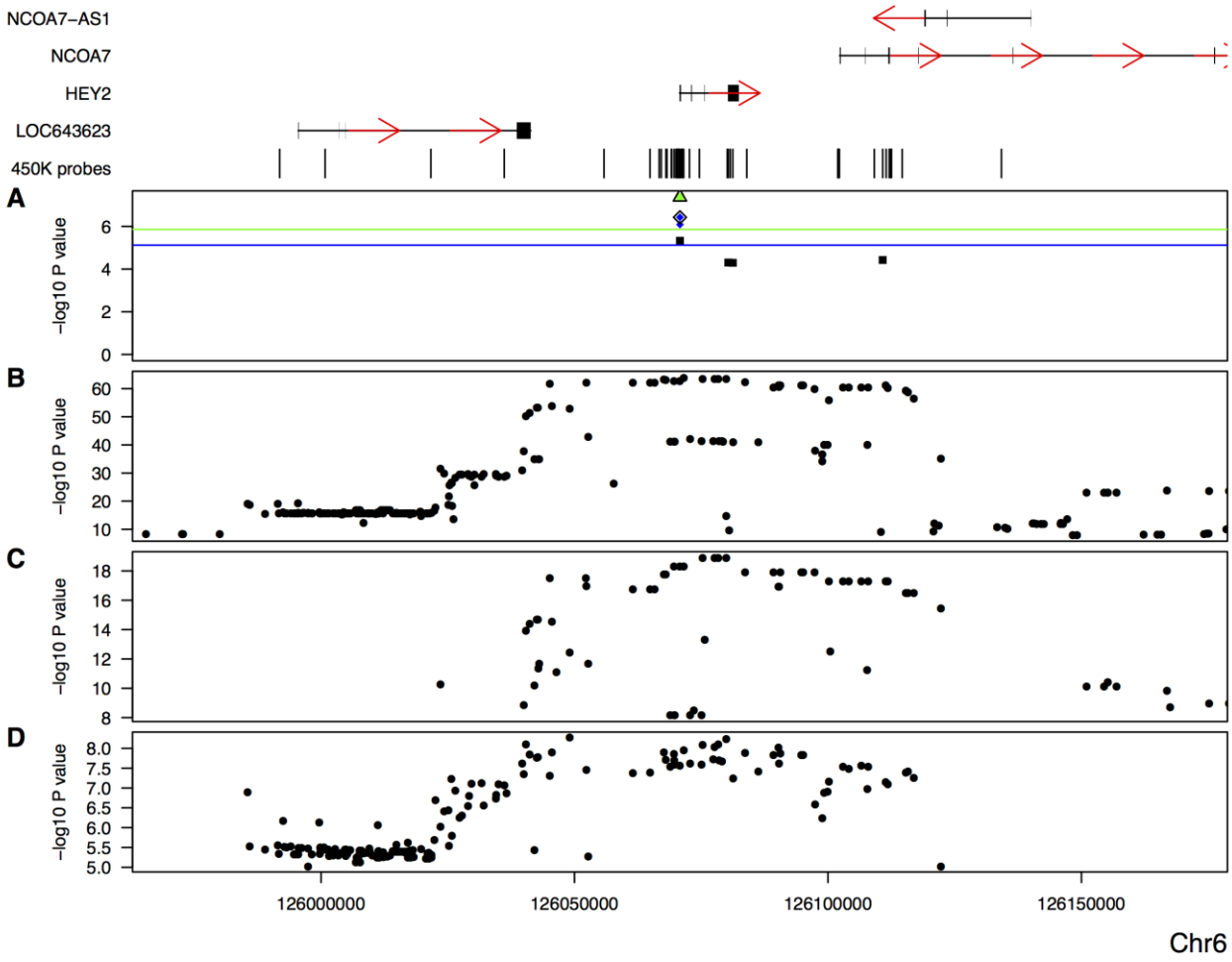
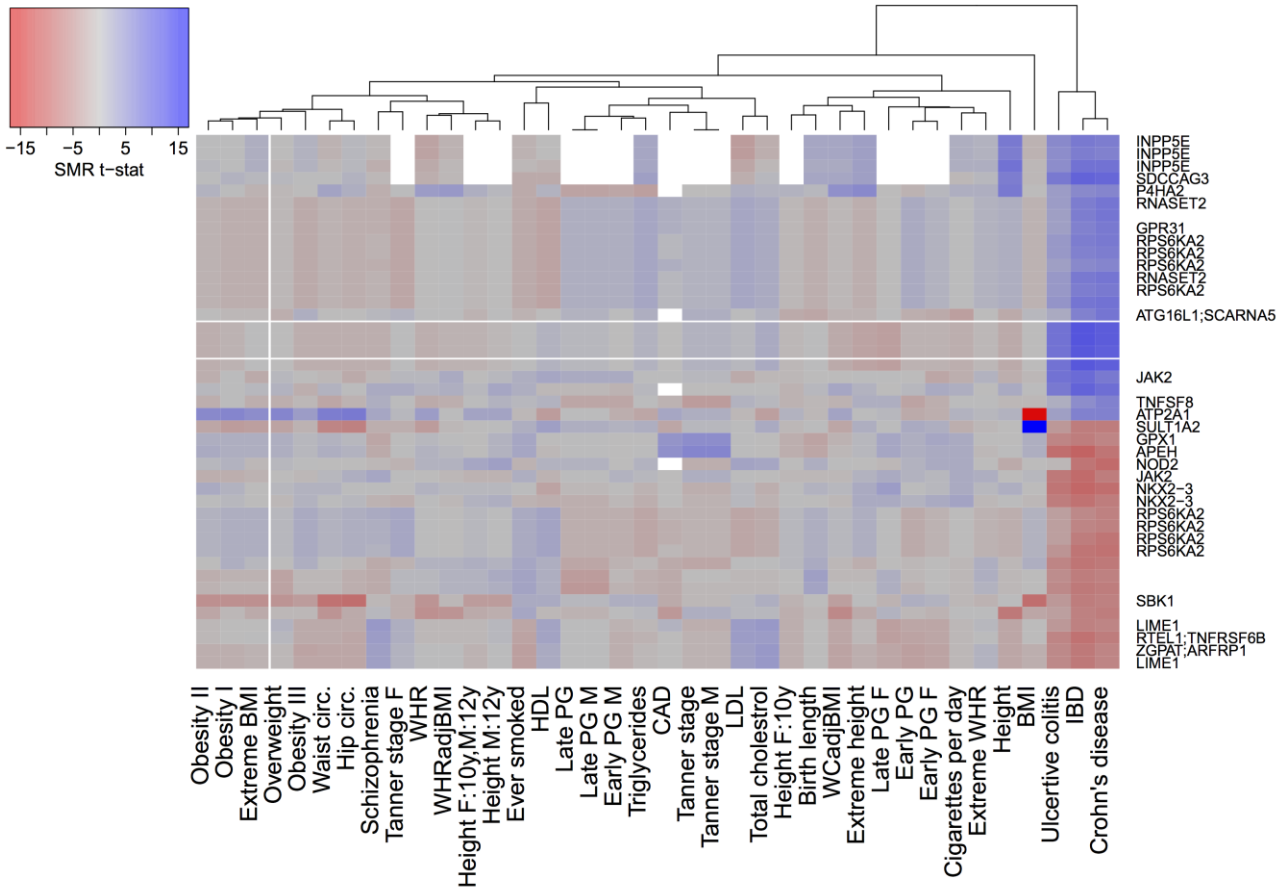

NCOA7−AS1
NCOA7
HEY2
LOC643623
450K probes
A
B
C
D
−log10 P value
126000000
126050000
126100000
126150000
Chr6

Figure 3


−15
−5
5
15
SMR t−stat
INPP5E
INPP5E
INPP5E
SDCCAG3
P4HA2
RNASET2
GPR31
RPS6KA2
RPS6KA2
RPS6KA2
RNASET2
RPS6KA2
ATG16L1;SCARNA5
JAK2
TNFSF8
ATP2A1
SULT1A2
GPX1
APEH
NOD2
JAK2
NKX2−3
NKX2−3
RPS6KA2
RPS6KA2
RPS6KA2
RPS6KA2
SBK1
LIME1
RTEL1;TNFRSF6B
ZGPAT;ARFRP1
LIME1
Obesity II
Obesity I
Extreme BMI
Overweight
Obesity III
Waist circ.
Hip circ.
Schizophrenia
Tanner stage F
WHR
WHRadjBMI
Height F:10y,M:12y
Height M:12y
Ever smoked
HDL
Late PG
Late PG M
Early PG M
Triglycerides
CAD
Tanner stage
Tanner stage M
LDL
Total cholestrol
Height F:10y
Birth length
WCadjBMI
Extreme height
Late PG F
Early PG
Early PG F
Cigarettes per day
Extreme WHR
Height
BMI
Ulcertive colitis
IBD
Crohn's disease