Supplement

Materials and Methods

Genotype, gene expression, and methylation data

Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE) was a multicenter North American initiative to establish a cohort of asthmatic and healthy subjects for the study of integrative genomics. [1,2] The current study includes Asthma BRIDGE participants for whom genotype, methylation and gene expression data were available. Asthma BRIDGE recruited subjects from six cohorts participating in the EVE Consortium for an asthma GWAS[1,2]: the Chicago Asthma Genetics (CAG) study, the Childhood Asthma Management Program (CAMP), the Childhood Asthma Research and Education Network (CARE), the Children's Health Study (CHS), the Genomic Research on Asthma in the African Diaspora (GRAAD) study, and the Mexico City Childhood Asthma Study (MCCAS). Written informed consent was obtained from all subjects prior to enrollment in the study, in accordance with the local ethics committees at each institution.

All blood samples were processed using a standardized protocol by all of the centers involved in the study. Whole blood (WB) or $CD4^+$ T cells were collected depending on the capabilities of the study center at which the subjects were enrolled. For WB, samples collected in EDTA tubes had DNA for methylation extracted using the PureGene Blood Kit (Qiagen, Germantown, Maryland) and samples collected in RNA PaxGene tubes had RNA isolated for gene expression studies using the PreAnalytix PAXgene Blood RNA Kit (Qiagen). For $CD4^+$ T cells, blood was collected in BD Vacutainer CPT tubes with

sodium heparin (FisherSci, Hampton, New Hampshire). Peripheral blood mononuclear cells were stimulated with 25ul of Phytohaemagglutinin (PHA, at 5ug/ml, Sigma-Aldrich, St. Louis, Missouri) for 24 hours at 37 degrees Celsius. $CD4^+$ T cells were isolated using CD4 MicroBeads (Miltenyi Biotech, Cambridge, Massachusetts), followed by DNA and RNA extraction using the AllPrep DNA/RNA Mini Kit (Qiagen) to be used for methylation and gene expression studies. We analyzed the WB and $CD4^+$ T cell samples as independent cohorts as there is no subject overlap between the groups. Demographic and clinical characteristics are summarized in Table E1.

Genome-wide single nucleotide polymorphism (SNP) genotype data was already available for all Asthma BRIDGE participants.[1] A common set of SNP genotypes was obtained by imputation in each cohort using MaCH (version 1.0)[3] and the 1000 Genomes Project EUR reference phased haplotypes based on Phase 1 low coverage data (20101123 release). SNPs with minor allele frequency (MAF) < 1%, p-value for Hardy-Weinberg equilibrium test < 0.001, and/or an imputation quality score < 0.3, were excluded. Methylation and gene expression profiling was attempted on all subjects without regard to phenotype or endotype and were collected simultaneously. Poorly performing samples were removed from all analyses. Genome-wide gene-expression data generated at the Channing Division of Network Medicine was obtained from peripheral $CD4^+$ lymphocytes (n=471) and WB samples (n=335) using the Illumina Human HT-12 v4 array (Illumina, San Diego, California). Genome-wide CpG methylation signatures for 450,000 sites were generated at the University of Southern California Epigenome Center using the Illumina HumanMethylation450 arrays in subjects from peripheral $CD4^+$

lymphocytes (n=297) and WB (n=565). Subjects in CAMP were excluded from the analysis as their methylation data was generated using an earlier Illumina methylation BeadChip. Genotype, methylation and gene expression data were available for 293 and 264 subjects in the WB and $CD4^+$ T-cell cohorts, respectively.

For the 17q21 gene locus, we restricted our analysis to a region within 50 kb of the gene boundaries of *ZPBP2*, *GSDMB*, and *ORMDL3* as mapped to the hg19 build (chr17:37974455-38130456). From the imputed SNPs, we limited the analysis to the 356 SNPs in this region with a MAF > 5% in Asthma BRIDGE. Seven gene expression probes were included, three for *ZPBP2* (ILMN_2383638, ILMN_1788040 and ILMN_1666357), three for *GSDMB* (ILMN_2347193, ILMN_2260756 and ILMN_1666206) and one for *ORMDL3* (ILMN_1662174). 45 CpG sites were included in the analysis that did not have a known SNP underlying probe sequences and were not located within repeat regions (cg03293732, cg13878456, cg16293631, cg27613455, cg15637191, cg00716716, cg15636887, cg14585353, cg05330360, cg16810031, cg09639931, cg11212589, cg21499348, cg27435903, cg21300187, cg10057218, cg16822095, cg05725940, cg13212159, cg05842113, cg12655416, cg22144450, cg18711369, cg10909506, cg02305874, cg13476672, cg04145193, cg16638648, cg01482279, cg05271360, cg14647739, cg09155575, cg04308185, cg10444806, cg08932654, cg02965290, cg02642223, cg09772097, cg19422138, cg24092444, cg06532257, cg04756853, cg24910161, cg2616229 and cg08995595).

Association tests

Linear regression models, as implemented with the R package "*limma*", were used for all analyses, adjusting for age, gender, ethnicity and study site as covariates. For analyses involving gene expression and/or methylation, the respective technical processing batch was also included as a covariate. To account for unknown confounders, we generated gene expression and methylation principal components (PCs) for each cohort separately, using all subjects with available data. Gene expression PCs were generated using all 47,009 available probes. Methylation PCs were generated using 470,870 autosomal CpG probes. PCs accounting for at least 1% of the variance were included in all analyses. Only additive genetic models were considered in all analyses that included genotype.

For expression quantitative trait locus (eQTL) analysis, we considered SNPs within 50 kb of the target gene boundaries (for *ZPBP2* [3 probes], *GSDMB* [3], and *ORMDL3* [1]), resulting in 1638 SNP-probe pairs tested. For methylation QTL (mQTL) analysis, all SNP-CpG site pairs within 50 kb of each other (9,207 total) were analyzed. CpG sites within 50 kb of the gene boundaries of *ZPBP2*, *GSDMB*, and *ORMDL3* were analyzed for an association between methylation and expression of the gene, resulting in 265 CpG-probe pairs tests. To correct for multiple comparisons testing, we generated false discovery rates (FDR), and declared statistical significance at an FDR < 0.05.

Causal inference test

The causal inference test (CIT) was implemented using the R package "*cit*".[4] The CIT consists of four component tests that evaluate: (i) the association between gene

expression and genotype; (ii) the association between methylation and gene expression adjusted for genotype; (iii) the association between genotype and methylation adjusted for gene expression; and (iv) the independence of genotype from gene expression adjusted for methylation. All four conditions must be met to provide statistical support for mediation. Gene expression probes that were replicated in both cohorts for associations with both genotype and methylation were included. For a given probe, we tested all possible combinations of replicated SNPs and CpG sites with which it was associated. To account for potential confounders, we used as inputs for the CIT the residuals for gene expression and methylation after adjusting for the covariates and PCs. The FDR was generated using a permutation-based approach and results were considered statistically significant at an FDR < 0.05. All analyses were conducted with R version 3.3.

References

E1. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet 2011;43:887–92.

E2. Croteau-Chonka DC, Qiu W, Martinez FD, Strunk RC, Lemanske RF, Liu AH, et al. Gene Expression Profiling in Blood Provides Reproducible Molecular Insights into Asthma Control. Am J Respir Crit Care Med 2017;195:179–88.

E3. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34:816–34.

E4. Millstein J, Chen GK, Breton CV. cit: hypothesis testing software for mediation

analysis in genomic applications. Bioinformatics 2016;32:2364–5.

Figure E1. Summary of 17q21 univariate eQTL, mQTL, and methylation-expression analyses.

The number of CpG methylation marks, SNPs and gene expression probes used in the association analyses are indicated below their respective label. The connecting arrows indicate the analysis type show the pairs of elements (SNPs, CpGs, expression probes) tested. For each analysis, counts and proportions of significant tests in the whole blood and $CD4^+$ T cell cohorts, and of results common to both cohorts, are provided lateral to the connecting arrows. Counts of unique CpG sites (blue boxes), SNPs (red boxes), and gene expression probes (green boxes) that replicated in both cohorts are provided medial to the connecting arrows. Replicated elements - SNPs, CpG marks, expression probes - that are common across analyses and are considered for downstream causal inference testing (CIT) are provided in shaded red, blue and green boxes, respectively.

Figure E2. Correlation of cg12655416 methylation and *ORMDL3* gene expression stratified by genotype.

Scatter plots of residuals, after adjusting for covariates, for *ORMDL3* gene expression and cg12655416 methylation in whole blood (left) and $CD4^+$ T cell (right) cohorts. Linear regression fits shown for rs12936231 genotype specified in box at upper right hand corner.

Table E1. Population characteristics

	Whole blood (n=293)	$CD4^+$ T cells (n=264)
Age, mean ± sd (yrs)	23.6 ± 4.9	27.7 ± 16.3
Male, n (%)	135 (46.1)	135 (51.1)
Asthma, n (%)	199 (67.9)	215 (81.4)
Allergies, n (%)	179 (61.1)	201 (76.1)
Hay Fever, n (%)	101 (34.5)	148 (56.1)
Ethnicity		
European	61	78
Hispanic/Latino	193	28
Black or African American	0	133
Multiple ethnicities	34	23
Other/Uncertain*	5	2
Sites		
Children's Health Study (CHS)	187	0
Mexico City Childhood Asthma Study (MCCAS)	106	0
Chicago Asthma Genetics (CAG)	0	25
Childhood Asthma Research and Education Network (CARE)	0	129
Genomic Research on Asthma in the African Diaspora (GRAAD)	0	110

* Also includes American Indian, South East Asian

Table E2. Results from regression analysis for selected eQTL SNPs

				Whole blood (n=293)				$CD4^+$ T cells (n=264)			
Gene	**SNP**	**SNP Position#**	**Alleles**	**Estimate**	**Std. Error**	**Pr(>|t|)**	**FDR**	**Estimate**	**Std. Error**	**Pr(>|t|)**	**FDR**
GSDMB*	rs12936231	38029120	G/C	0.216	0.018	3.12E-26	**2.37E-24**	0.056	0.022	0.012	0.065
	rs8067378	38051348	G/A	0.216	0.018	3.12E-26	**2.37E-24**^	0.075	0.022	9.06E-04	**9.16E-03**
	rs12453507	38053207	G/C	0.213	0.018	6.48E-26	**2.62E-24**	0.086	0.023	2.88E-04	**4.76E-03**^
	rs1008723	38066267	T/G	0.216	0.018	9.03E-27	**2.37E-24**^	0.054	0.024	0.023	0.111
	rs4795401	38067533	G/A	0.220	0.018	4.65E-27	**2.37E-24**^	0.073	0.026	0.006	**0.040**
	rs7216389	38069949	C/T	0.211	0.018	2.25E-25	**5.41E-24**	0.075	0.026	0.005	**0.033**
	rs4065275	38080865	A/G	0.210	0.018	1.28E-25	**3.49E-24**	0.058	0.021	0.006	**0.042**
	rs8076131	38080912	G/A	0.155	0.021	2.64E-12	**1.58E-11**	0.057	0.024	0.016	0.084
	rs56199421	38090808	C/T	0.191	0.019	3.98E-21	**7.49E-20**	0.086	0.021	7.51E-05	**1.40E-03**^
	rs56301252	38109155	T/G	0.145	0.021	7.77E-11	**3.88E-10**	0.091	0.024	1.42E-04	**2.47E-03**^
ORMDL3	rs12936231	38029120	G/C	0.142	0.018	1.46E-13	**1.01E-12**	0.162	0.025	3.26E-10	**1.33E-07**
	rs2290400	38066240	C/T	0.146	0.018	9.24E-15	**7.28E-14**^	0.168	0.025	2.11E-10	**1.51E-07**^
	rs56380902	38066372	C/T	0.146	0.018	9.24E-15	**7.28E-14**^	0.172	0.025	8.91E-11	**1.51E-07**^
	rs7216389	38069949	C/T	0.139	0.018	3.02E-13	**2.02E-12**	0.144	0.030	3.72E-06	**1.19E-04**
	rs9303279	38073968	G/C	0.156	0.021	6.22E-13	**4.01E-12**	0.167	0.025	1.63E-10	**1.15E-07**^
	rs4065275	38080865	A/G	0.144	0.018	1.62E-14	**1.26E-13**	0.130	0.024	1.79E-07	**1.95E-05**
	rs8076131	38080912	G/A	0.061	0.020	2.69E-03	**7.35E-03**	0.124	0.027	9.20E-06	**2.43E-04**
	rs12603332	38082807	T/C	0.144	0.018	1.09E-14	**8.53E-14**^	0.123	0.024	6.11E-07	**4.35E-05**

* For *GSDMB*, results are shown for probe ILMN_1666206

Location based on hg19 build

^ Identifies top three hits for each gene in the specified cohort

Significant associations (FDR < 0.05) are marked in bold

Table E3. False discover rates (q-values) from causal inference test for *GSDMB* and *ORMDL3* gene expression with selected CpG sites and SNPs

			Whole blood (n=293)					CD4$^+$ T cells (n=264)				
Gene	**CpG**	**SNP**	**CIT**	**Exp assoc SNP**	**Exp assoc CpG gvn SNP**	**CpG assoc SNP gvn Exp**	**SNP ind of Exp gvn CpG**	**CIT**	**Exp assoc SNP**	**Exp assoc CpG gvn SNP**	**CpG assoc SNP gvn Exp**	**SNP ind of Exp gvn CpG**
GSDMB*	**cg10909506**	**rs4065275**	1.54E-03	3.90E-06	8.70E-04	5.32E-06	6.62E-04	1.75E-02	4.50E-03	7.21E-03	5.04E-06	5.87E-03
		rs7216389	1.59E-03	3.90E-06	8.01E-04	5.32E-06	7.82E-04	1.71E-02	6.64E-03	4.29E-03	5.04E-06	6.30E-03
	cg12655416	**rs4065275**	1.14E-04	3.90E-06	6.74E-06	5.32E-06	9.78E-05	9.82E-03	4.50E-03	1.18E-04	5.04E-06	5.22E-03
		rs7216389	1.15E-04	3.90E-06	6.74E-06	1.45E-05	8.98E-05	1.19E-02	6.64E-03	2.43E-05	5.33E-05	5.22E-03
	cg18711369	**rs4065275**	3.29E-03	3.90E-06	2.34E-03	5.32E-06	9.37E-04	7.70E-02	4.50E-03	4.78E-02	5.04E-06	2.63E-02
		rs7216389	1.28E-03	3.90E-06	7.07E-04	5.32E-06	5.67E-04	4.53E-02	6.64E-03	1.94E-02	5.04E-06	1.99E-02
	cg22144450	**rs4065275**	6.55E-04	3.90E-06	6.74E-06	2.88E-05	6.15E-04	1.69E-02	4.50E-03	6.41E-03	5.04E-06	6.12E-03
		rs7216389	6.29E-04	3.90E-06	6.74E-06	1.45E-05	6.04E-04	1.45E-02	6.64E-03	2.74E-03	5.04E-06	5.22E-03
	cg24910161	**rs4065275**	2.76E-02	3.90E-06	1.77E-02	5.32E-06	9.99E-03	3.13E-02	4.50E-03	2.03E-02	5.04E-06	6.74E-03
		rs7216389	2.75E-02	3.90E-06	1.79E-02	5.32E-06	9.76E-03	2.09E-02	6.64E-03	9.13E-03	5.04E-06	5.22E-03
	cg26162295	**rs4065275**	3.95E-01	3.90E-06	2.80E-01	5.32E-06	1.59E-01	3.04E-02	4.50E-03	1.93E-02	5.04E-06	6.80E-03
		rs7216389	2.64E-01	3.90E-06	1.75E-01	5.32E-06	1.09E-01	2.05E-02	6.64E-03	9.38E-03	5.04E-06	4.56E-03
ORMDL3	**cg10909506**	**rs12936231**	3.09E-03	3.90E-06	7.86E-04	5.32E-06	2.30E-03	5.27E-05	2.96E-05	1.22E-05	5.04E-06	5.87E-06
		rs4065275	5.85E-03	3.90E-06	1.72E-03	5.32E-06	4.13E-03	4.82E-05	2.96E-05	1.22E-05	5.04E-06	1.36E-06
		rs7216389	3.55E-03	3.90E-06	1.03E-03	5.32E-06	2.52E-03	1.02E-04	8.44E-05	1.22E-05	5.04E-06	3.56E-08
		rs8076131	1.54E-03	1.53E-03	6.74E-06	5.32E-06	6.11E-06	4.69E-05	2.96E-05	1.22E-05	5.04E-06	5.21E-08
	cg12655416	**rs12936231**	4.11E-03	3.90E-06	1.41E-03	5.32E-06	2.69E-03	1.15E-04	2.96E-05	1.22E-05	5.04E-06	6.86E-05
		rs4065275	1.27E-02	3.90E-06	5.75E-03	5.32E-06	6.94E-03	7.86E-05	2.96E-05	1.22E-05	5.04E-06	3.17E-05
		rs7216389	6.86E-03	3.90E-06	3.30E-03	5.32E-06	3.56E-03	9.82E-04	8.44E-05	1.22E-05	8.63E-04	2.23E-05
		rs8076131	1.55E-03	1.53E-03	6.74E-06	5.32E-06	1.59E-05	6.32E-05	2.96E-05	1.22E-05	5.04E-06	1.63E-05
	cg18711369	**rs12936231**	1.14E-04	3.90E-06	6.74E-06	5.32E-06	9.77E-05	4.72E-04	2.96E-05	1.22E-05	5.04E-06	4.25E-04
		rs4065275	2.74E-04	3.90E-06	6.74E-06	5.32E-06	2.58E-04	7.59E-05	2.96E-05	1.22E-05	5.04E-06	2.90E-05
		rs7216389	1.16E-04	3.90E-06	6.74E-06	5.32E-06	1.01E-04	1.12E-04	8.44E-05	1.22E-05	5.04E-06	9.95E-06
		rs8076131	1.54E-03	1.53E-03	6.74E-06	5.32E-06	2.65E-07	5.42E-05	2.96E-05	1.22E-05	5.04E-06	7.30E-06
	cg22144450	**rs12936231**	6.22E-04	3.90E-06	3.77E-05	5.32E-06	5.75E-04	5.61E-04	2.96E-05	1.22E-05	5.04E-06	5.14E-04
		rs4065275	7.78E-04	3.90E-06	4.40E-05	5.32E-06	7.25E-04	1.70E-04	2.96E-05	1.22E-05	5.04E-06	1.23E-04
		rs7216389	6.03E-04	3.90E-06	1.90E-05	5.32E-06	5.75E-04	2.10E-04	8.44E-05	1.22E-05	5.71E-05	5.67E-05
		rs8076131	1.57E-03	1.53E-03	6.74E-06	3.78E-05	3.30E-06	1.06E-04	2.96E-05	1.22E-05	5.04E-06	5.96E-05
	cg26162295	**rs12936231**	3.14E-01	3.90E-06	2.23E-01	5.32E-06	1.18E-01	5.68E-01	2.96E-05	4.46E-01	5.04E-06	2.20E-01
		rs4065275	5.38E-01	3.90E-06	4.02E-01	5.32E-06	2.26E-01	3.58E-01	2.96E-05	2.52E-01	5.04E-06	1.41E-01
		rs7216389	3.24E-01	3.90E-06	2.26E-01	5.32E-06	1.26E-01	8.76E-02	8.44E-05	5.36E-02	5.04E-06	3.58E-02
		rs8076131	1.98E-03	1.53E-03	3.01E-04	5.32E-06	1.52E-04	1.49E-01	2.96E-05	9.31E-02	5.04E-06	6.11E-02

* For *GSDMB*, results shown for probe ILMN_1666206

Results from causal inference test (CIT)

CIT - Overall permutation-based FDR for CIT

Exp assoc SNP - Permutation-based FDR from CIT for association between gene expression (Exp) and SNP

Exp assoc CpG gvn SNP - Permutation-based FDR from CIT for association between gene expression and CpG given (gvn) SNP

CpG assoc SNP gvn Exp - Permutation-based FDR from CIT for association between CpG and SNP given gene expression

SNP ind of Exp gvn CpG - Permutation-based FDR from CIT for equivalence test of SNP being independent (ind) of gene expression given CpG

Figure E1.

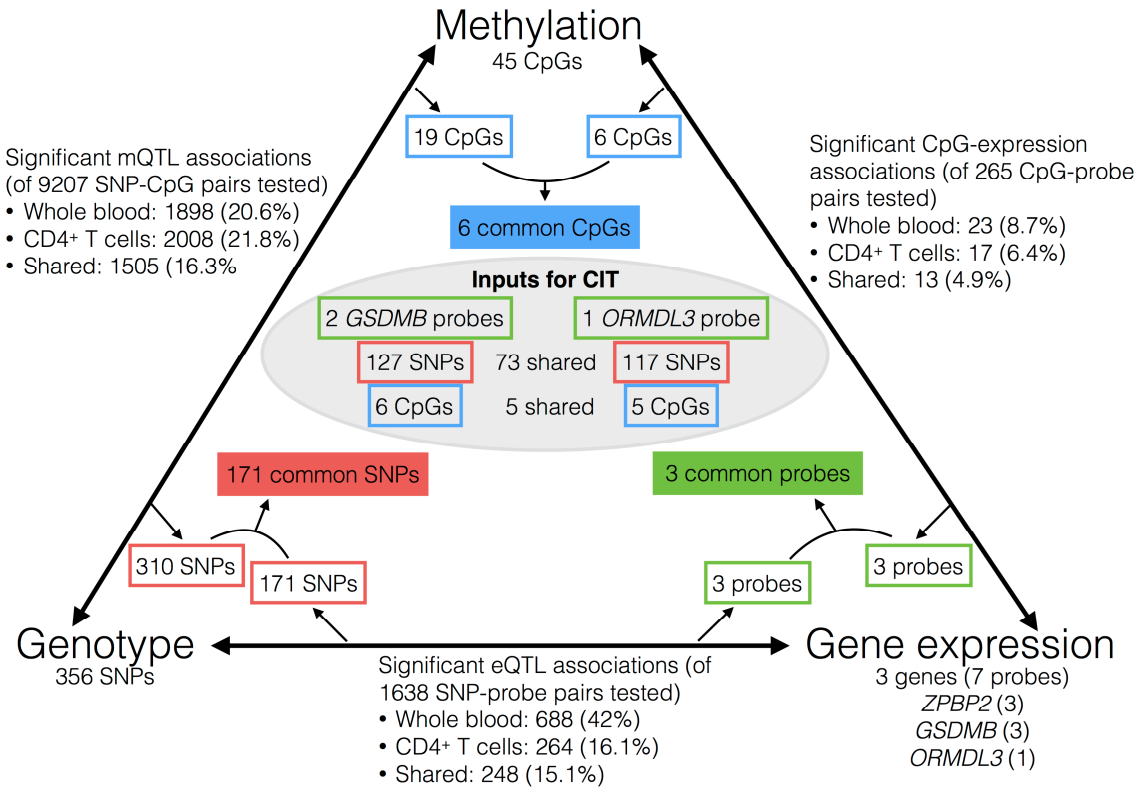
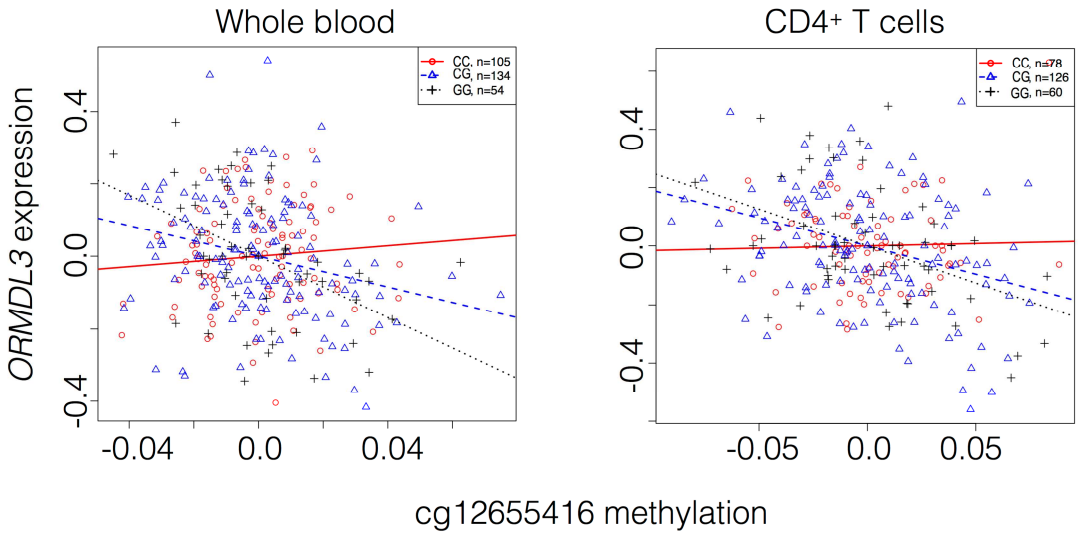

Methylation
45 CpGs
19 CpGs
6 CpGs
6 common CpGs
Inputs for CIT
2 GSDMB probes
1 ORMDL3 probe
127 SNPs
73 shared
117 SNPs
6 CpGs
5 shared
5 CpGs
Significant mQTL associations (of 9207 SNP-CpG pairs tested)
• Whole blood: 1898 (20.6%)
• CD4+ T cells: 2008 (21.8%)
• Shared: 1505 (16.3%
Significant CpG-expression associations (of 265 CpG-probe pairs tested)
• Whole blood: 23 (8.7%)
• CD4+ T cells: 17 (6.4%)
• Shared: 13 (4.9%)
171 common SNPs
3 common probes
310 SNPs
171 SNPs
3 probes
3 probes
Genotype
356 SNPs
Significant eQTL associations (of 1638 SNP-probe pairs tested)
• Whole blood: 688 (42%)
• CD4+ T cells: 264 (16.1%)
• Shared: 248 (15.1%)
Gene expression
3 genes (7 probes)
ZPBP2 (3)
GSDMB (3)
ORMDL3 (1)

Figure E2.


Whole blood
CD4+ T cells
CC, n=105
CG, n=134
GG, n=54
CC, n=78
CG, n=126
GG, n=60
ORMDL3 expression
0.4
0.0
-0.4
-0.04
0.0
0.04
-0.05
0.0
0.05
cg12655416 methylation