Multi-omics co-localization with genome-wide association studies reveals context-specific mechanisms of asthma risk variants

Marcus M. Soliai[1,2*], Atsushi Kato[3], Catherine T. Stanhope[2], James E. Norton[3], Katherine A. Naughton[2], Aiko I. Klinger[3], Robert C. Kern[4], Bruce K. Tan[4], Robert P. Schleimer[3], Dan L. Nicolae[1,2,5,6], Jayant M. Pinto[7], Carole Ober[1,2*]

[1]Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL, United States of America

[2]Departments of Human Genetics, University of Chicago, Chicago, IL, United States of America

[3]Departments of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States of America

[4]Department of Otolaryngology-Head and Neck Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL, United States of America

[5]Department of Medicine, University of Chicago, Chicago, IL, United States of America

[6]Department of Statistics, University of Chicago, Chicago, IL, United States of America

[7]Department of Surgery, University of Chicago, Chicago, IL, United States of America

*Corresponding authors

E-mail: msoliai@uchicago.edu (MS); c-ober@bsd.uchicago.edu (CO)

Abstract

Genome-wide association studies (GWASs) have identified thousands of variants associated with asthma and other complex diseases. However, the functional effects of most of these variants are unknown. Moreover, GWASs do not provide context-specific information on cell types or environmental factors that affect specific disease risks and outcomes. To address these limitations, we used cultured upper airway (sinonasal) epithelial cell models to assess transcriptional and epigenetic responses to a virus (rhinovirus [RV]) and a bacterium (*Staphylococcus aureus* [SA]) and provide context-specific functional annotations to variants discovered in GWASs of asthma. Using genome-wide genetic, gene expression, and DNA methylation data in RV-, SA- and vehicle-treated cells from 115 individuals, we mapped *cis* expression and methylation quantitative trait loci (*cis*-eQTLs and *cis*-meQTLs, respectively) in each condition. Co-localization analyses of these airway epithelial cell molecular QTLs with asthma GWAS variants revealed potential molecular disease mechanisms of asthma for GWAS variants, including QTLs at the *TLSP* locus that were common both to exposure conditions and childhood onset and adult onset asthma and at the 17q12-21 asthma locus that were specific to both RV exposure and childhood onset asthma, consistent with clinical and epidemiological studies of these loci. Overall, our study provides information on functional effects of asthma risk variants in airway epithelial cells and provides insight into disease-relevant microbial exposures that modulate genetic effects on transcriptional and epigenetic responses in cells and on risk for asthma in GWAS.

Author Summary

Both genetic and environmental factors influence asthma. Genome-wide association studies have identified thousands of genetic variants associated with asthma but do not provide information on their functional effects, tissue specificity, or environmental context. To address these limitations, we used an upper airway epithelial cell model to study responses to microbes that potentially influence airway disease inception and/or symptoms, and to understand the functional relevance of asthma risk variants. To this end, we mapped genetic variation associated with gene expression and DNA methylation in cells exposed to a virus (rhinovirus) or a bacterium (*Staphylococcus aureus*) compared to vehicle controls and tested for co-localization of these molecular traits with variants associated with adult onset and childhood onset asthma in GWAS. We report putative disease mechanisms of asthma and associated genes and DNA methylation sites in airway epithelial cells exposed to disease-promoting risk factors.

Introduction

Over the past decade, genome-wide association studies (GWASs) have identified many thousands of variants at hundreds of loci containing susceptibility genes for asthma and other complex diseases [1]. Notably, over 90% of the single nucleotide polymorphisms (SNPs) identified in GWAS reside in non-coding regions of the genome that are enriched for chromatin signatures suggestive of enhancers [2] and for expression quantitative trait loci (eQTLs) [2-4]. These features of SNPs associated in GWAS indicate that they most likely affect underlying disease pathophysiology through their effects on gene regulation. However, identifying the causal variants and their target genes for associated loci has been challenging, and the functions of most associated SNPs remain unknown. Furthermore, the significance threshold ($p<5x10^{-8}$) required to control the false discovery rate in GWASs likely excludes many true associations that do not reach this stringent threshold. We and others have suggested that SNPs with small p-

values that do not meet genome-wide significant thresholds, i.e., the mid-hanging fruit [5], may be environment- or context-specific associations that are missed in GWAS that typically do not control for either [6, 7]. Additionally, databases such as GTEx, ENCODE, and ROADMAP are used to annotate GWAS SNPs and predict molecular mechanisms through which risk variants affect disease phenotypes [4, 8, 9]. Although these resources have provided important insights into the interpretation of GWAS results, they do not include all cell types relevant to all diseases or information on environmental exposures that influence disease outcomes. As a result, annotations of asthma GWAS variants have been largely limited to studies in cell lines, blood (immune) cells, and whole lung tissue [10-12].

In vitro cell models provide an opportunity to address these limitations by identifying and characterizing genetic and molecular responses to environmental exposures in cells from disease-relevant tissues, and identifying genotypes that modify response to environmental risk [13, 14]. Joint analysis of datasets (e.g. eQTLs and GWAS) identifies variants associated with both disease risk and molecular traits as candidate causal variants that contribute to mechanisms of disease pathophysiology. A multi-trait co-localization (*moloc*) method [15] was recently developed to integrate summary data from GWAS and multiple molecular QTL datasets to identify regulatory drivers of complex phenotypes and to provide a more comprehensive analysis in multi-trait studies.

Here, we report the results of a multi-omics co-localization study to identify condition-specific regulatory effects of asthma risk variants in an epithelial cell model of microbial response. Because airway epithelium plays a crucial role in response to inhaled exposures, we used an *in vitro* upper airway (sinonasal) epithelial cell model of transcriptional and epigenetic responses to two microbial exposures that are associated with asthma, RV and SA. RV

respiratory infections are a major cause of asthma inception in young children [16] and asthma exacerbations in older children and adults [17, 18], and colonization of the nose with SA is more common among asthmatics and associated with increased asthma symptoms and exacerbations in children and young adults [19], underscoring the importance of these microbes as contextual influences in asthma pathophysiology. Using co-localizations of airway epithelial cell molecular QTLs with asthma GWAS risk variants, we characterized the effects of regulatory variation on gene expression and DNA methylation levels in each treatment condition and used this information to annotate and assign condition-specific regulatory effects at asthma GWAS risk loci. Our integrative multi-omics approach revealed potential environment-specific mechanisms of asthma pathogenesis, and further support a key role of airway epithelium in the pathogenesis of childhood onset asthma.

Results

Genome-wide identification of *cis*-eQTLs in cultured airway epithelial cells

To identify genetic variation influencing gene expression under different conditions, we performed eQTL mapping in cultured airway epithelial cells treated with two common microbial exposures (RV and SA), and their corresponding vehicles from 115 individuals (S1 Fig). Because the vehicles for RV and SA differed (bronchial epithelial basal medium [BEBM] for RVveh and Dulbecco's phosphate buffered saline [dPBS] for SAveh) and cells were cultured for different lengths of time (48 hours for RV and 24 hours for SA), we considered each vehicle as a separate treatment condition and refer to them as RVveh and SAveh, respectively. We defined eQTLs within a *cis*-window of 1 Mb from either side of the transcriptional start site (TSS) of each autosomal gene and used a false discovery rate (FDR) of 10%. Analyses were performed

123 separately for each of the four conditions, testing for associations with 6,665,553 imputed SNPs
124 and 11,231 and 11,421 autosomal genes (from RNA-seq) for the RV and SA experiments,
125 respectively (see Materials and Methods). The numbers of SNPs associated with gene expression
126 (eQTLs), SNPs that are eQTLs for at least one gene (eSNPs), and genes with at least one eQTL
127 (eGenes) in each condition are summarized in Fig 1.
128
129 **Genome-wide identification of *cis*-meQTLs in cultured airway epithelial cells**
130 In parallel to eQTL mapping, we performed methylation quantitative trait loci (meQTL)
131 mapping in the same cells used for gene expression studies. We defined meQTLs within a *cis*-
132 window of 10 Kb from either side of each CpG site on the Illumina EPIC array, using an FDR of
133 10%. We performed this analysis separately for each condition, testing for associations with the
134 same imputed SNP set as that used for eQTL mapping and interrogated 792,392 and 749,125
135 autosomal CpGs for the RV and SA experiments, respectively. A summary of the number of
136 SNPs associated with methylation levels at one or more CpG sites (meQTLs), SNPs that are
137 meQTLs for at least one CpG (meSNPs), and CpG sites with at least one meQTL (meCpGs) are
138 shown in Fig 1.
139
140 **Estimating shared and condition-specific molecular QTL effects**
141 We first explored the impact of culture conditions (RV+RVveh vs SA+SAveh) and treatments
142 (RV and/or SA vs RVveh and/or SAveh, respectively) on eQTLs and meQTL effects using an
143 empirical Bayes method, **m**ultivariate **a**daptive **sh**rinkage (*mash*) [20]. This is accomplished in
144 two general steps; we first identified existing patterns within the observed dataset, including
145 correlations among effects, sparsity, and sharing, and then we used these learned patterns to

146 make improved effect estimates and significance measurements for a given set of data from
147 multiple conditions. Compared to direct comparisons between conditions, *mash* increases power,
148 improves effect-size estimates, and provides better quantitative assessments of effect size
149 heterogeneity of molecular QTLs, thereby allowing for greater confidence in effect sharing and
150 estimates of condition-specificity. Additionally, as a confidence measurement of the direction of
151 each effect (molecular QTL), *mash* provides a 'local false sign rate' (lfsr) that is the probability
152 that the estimated effect has the incorrect sign [21], rather than the expected proportion of Type I
153 errors as would be assessed using FDR thresholds.

154 To identify condition-specific eQTLs, we analyzed the effect estimates for 337,699
155 eQTLs and observed broad sharing of eQTLs across these treatment conditions (see Materials
156 and Methods). Pairwise comparisons showed that between 63-89% of eQTLs were shared in at
157 least two conditions (lfsr < 0.05; Fig 1A and S2 Fig). The RV and RVveh cells showed the most,
158 and SA and SAveh cells the second most, eQTL effect sharing (89% and 72%, respectively),
159 while lower and similar amounts of sharing were observed between the other pairwise
160 comparisons (63-71%). In contrast, only 0.98 to 9.20% of all eQTLs were specific to one
161 condition, with the largest number detected in the SA-treated cells and the least in the RV-treated
162 cells (Fig 1A). The RV and SA culture conditions (RV+RVveh vs SA+SAveh) had relatively
163 large impacts on eQTL effects, with 10.6% and 20.6% of the eQTLs specific to culture
164 conditions, respectively. Another 10.6% of the eQTLs (35,874) were specific to the microbial-
165 treated cells only (RV+SA vs RVveh+SAveh), potentially representing genetic variants that
166 modify responses to microbes in airway epithelial cells. Examples of treatment-specific eQTLs
167 are shown in Fig 2B.

Condition-specific meQTLs were identified among the 1,669,925 meQTLs. A pair-wise analysis revealed that between 73-79% of meQTLs were shared in two or more of the four conditions (lfsr < 0.05; Fig 2C and S2 Fig), with 73% of these shared among culture conditions (RV vs RVveh or SA vs SAveh), many more than those observed for eQTLs. In contrast, only 3.0 and 5.2% of meQTLs were treatment-specific (RV+SA or RVveh+SAveh), respectively (Fig 2D), many fewer than observed for eQTLs. Examples of treatment-specific meQTLs are shown in Fig 2D.

In total, using *mash* we identified between 3,295 - 30,994 eQTLs associated with 102 - 582 eGenes (lfsr<0.05) that were unique to a culture condition, and between 84,602 - 48,415 meQTLs associated with 7,636 - 12,817 meCpGs that were unique to a culture condition. This approach allowed us to assign QTL effects sharing and potential condition- and/or treatment-specific effects with greater confidence than by pairwise comparisons using FDR thresholds [20].

Molecular QTL co-localizations with adult onset and childhood onset asthma loci

Integrating molecular QTLs with GWAS data is a powerful way to identify functional variants that may ultimately influence disease risk. This approach can provide functional insights into known disease-associated variants as well to facilitate prioritizing variants with small p-values that do not reach criteria for genome-wide significance in GWAS. One such approach is through co-localization in which we directly test whether a genetic variant is underlying associations between two or more traits (e.g., gene expression and asthma), providing clues to causal disease pathways. We hypothesized that integrating molecular QTLs from microbial-exposed epithelial cells with results of GWASs for adult onset and childhood onset asthma would reveal genetic

191 variants that are associated with both disease risk and gene expression and/or DNA methylation
192 patterns. This could reveal potential genetic and epigenetic mechanisms of childhood onset
193 and/or adult onset asthma that are modulated by disease-relevant microbial exposures in airway
194 epithelial cells, an asthma-relevant tissue that is the target of these microbes.

195 To test this hypothesis, we extracted summary statistics from the largest GWASs of adult
196 onset and childhood onset asthma to date [12], and tested each for co-localization with genetic
197 variants associated with gene expression, DNA methylation, and asthma, using *moloc*, a
198 Bayesian statistical strategy that allows the integration and co-localization of more than two
199 molecular traits [15]. We performed a separate co-localization test in each of the four conditions
200 using variants from the GWASs of adult onset and childhood onset asthma separately. These
201 analyses provide three possible configurations in which a variant is shared between the GWAS
202 and QTL traits: eQTL-GWAS pairs, meQTL-GWAS pairs, eQTL-meQTL-GWAS triplets.

203 Using this approach, we found evidence for 72 unique multiple trait co-localizations (76
204 total) associated with eQTLs for 11 genes (*ACO2*, *ERBB2*, *FGFR4*, *FLG*, *FLG-AS1*, *FRK*,
205 *GSTO2*, *IRF5*, *ORMDL3*, *PMM1*, *POLI*) and meQTLs for 31 CpG sites (PPA>70%) for adult
206 onset or childhood onset asthma (Table 1; S1 Table). Among the 72 unique co-localizations, 14
207 were eQTL-meQTL-GWAS triplets associated with six genes (*ACO2*, *ERBB2*, *FRK*, *GSTO2*,
208 *IRF5*, *PMM1*) and 11 CpG sites. There were also 25 eQTL-GWAS pairs associated with seven
209 genes (*ACO2*, *FGFR4*, *FLG*, *ORMDL3*, *PMM1*, *POLI*) and 33 meQTL-GWAS co-localized
210 pairs associated with 24 CpG sites. The majority of the co-localizations (88%) were identified in
211 just one treatment condition (PPA>70%). Twenty-eight of the co-localizations were identified
212 only in the microbial-exposed cells (RV and/or SA) and 41 were identified only in the vehicle
213 exposed cells (RVveh and/or SAveh). These 69 treatment-specific colocalizations represent

potential response mechanisms to microbial infection that contribute to asthma risk. The remaining seven co-localizations were identified in combinations of microbe and vehicle exposed cells. Notably, over 72% of the SNPs associated with these co-localizations did not reach genome-wide significance (p-value=$5x10^{-8}$) in their respective GWASs ($p\text{-value}_{range}$ = $1.03x10^{-7}$ to $4.63x10^{-4}$), and all were identified only in the childhood onset asthma GWAS. These analyses therefore provided functional inferences both for variants that were significant in the GWASs at known asthma loci and for variants that did not meet strict criteria for significance in the GWASs, thereby facilitating prioritization of variants among the mid-hanging fruit.

Of the 76 co-localizations, only six were at adult onset asthma loci and all six were meQTL-GWAS pairs. The remaining 70 were at childhood onset asthma loci and 14 were eQTL-meQTL-GWAS triplets, 25 were eQTL-GWAS pairs, and 31 meQTL-GWAS pairs) (Table 1; S1 Table). Four meQTL-GWAS pairs were co-localized in both GWASs and two were specific to adult onset asthma. In contrast, 66 co-localizations were specific to childhood onset asthma, and were associated with 11 genes (*ACO2*, *ERBB2*, *FGFR4*, *FLG*, *FLG-AS1*, *FRK*, *GSTO2*, *IRF5*, *ORMDL3*, *PMM1*, *POLI*) and 27 CpG sites. The larger number of co-localizations for childhood onset asthma relative to adult onset asthma is consistent with the previous observation that genes at the childhood onset asthma loci were most highly expressed in skin, an epithelial cell type [12].

Table 1. Number of QTL-GWAS pairs or triplets with evidence of co-localization

GWAS	eqtl-meQTL-GWAS	eQTL-GWAS	meQTL-GWAS
Adult onset asthma	0	0	6
Childhood onset asthma	14	25	31
Adult or childhood onset GWAS	14	25	35

PPA >= 0.70

meCpGs associated with co-localized risk variants correlate with gene expression

Methylation at CpG dinucleotides plays critical roles in the regulation of various cellular processes [22] and can potentially mediate the effects of environmental exposures on gene expression [23]. Many complex diseases, including asthma, have been associated with DNA methylation patterns, supporting the important role of epigenetics in disease processes [24-26]. A mechanism through which DNA methylation can affect phenotypic outcomes is by influencing gene expression [27]. Because of the large number of meSNPs that co-localized with asthma risk variants, we next asked whether methylation levels at the CpG sites with co-localized meSNPs were also associated with the expression of nearby genes. To address this question, we defined nearby genes as those in which the meCpG was in the 5'UTR, gene body, 3'UTR or within 1500 bp of its transcriptional start site (TSS). Using these criteria, 21 of the 33 unique meQTL-GWAS pairs were assigned to 12 unique genes (*EEFSEC*, *FLG-AS1*, *FRK*, *GSTO2*, *IRF1*, *IRF5*, *MAP3K14*, *NEK6*, *POLR3H*, *SLC22A5*, *SMARCE1*, *TSLP*). We then tested for correlation between methylation and gene expression for the 22 gene-CpG pairs for each treatment condition using Spearman rank order correlation. Sixteen gene-CpG correlations ($\rho_{bsolute} = 0.26 – 0.68$) with five genes (*FLG-AS1*, *FRK*, *GSTO2*, *IRF5*, *SLC22A5*) and eight meCpGs were significant (FDR ≤ 0.05; Table 2; S3 Fig; S2 Table).

Table 2. Correlation of methylation levels and expression of the nearest gene for 8 meQTL-GWAS co-localized pairs (Spearman's rank order correlation). See text for details on gene assignments. Five gene-CpG pairs were correlated in multiple conditions (gray shaded rows).

Gene	CpG	Spearman p-value	FDR	Spearman ρ	Treatment
FLG-AS1	cg26320663	2.03E-03	2.03E-02	-0.32	rv
		6.46E-03	3.71E-02	-0.28	vehrv
FLG-AS1	cg23107878	6.49E-03	3.71E-02	-0.28	rv
FRK	cg20254830	2.70E-06	4.32E-05	0.45	sa
GSTO2	cg23659134	0.00E+00	0.00E+00	-0.66	sa
		0.00E+00	0.00E+00	-0.68	vehsa
		2.53E-05	3.37E-04	-0.42	vehrv
		7.81E-03	3.95E-02	-0.27	rv
GSTO2	cg07488549	2.94E-07	7.84E-06	0.49	vehsa
		5.73E-07	1.15E-05	0.48	sa
IRF5	cg26616347	4.08E-03	3.26E-02	-0.29	vehrv
		4.67E-03	3.40E-02	-0.28	vehsa
		7.90E-03	3.95E-02	-0.26	sa
SLC22A5	cg04774966	7.34E-05	8.39E-04	-0.39	vehsa
		3.08E-03	2.74E-02	-0.29	sa
SLC22A5	cg26647941	5.88E-03	3.71E-02	-0.27	sa

Correlations between DNA methylation and gene expression were identified in multiple treatment conditions and after exposure to RV or SA only. For example, significant correlations between DNA methylation and *GSTO2* on chromosome 10 occur in all treatment conditions ($\rho_{absolute}$ = 0.27 – 0.68), suggesting that the correlation signals are robust across treatments (Fig 3A). In contrast, we observed potential treatment-specific effects on correlations between an meCpG 241 bp upstream of *IRF5* and expression of this gene on chromosome 7 (Fig 3B), for which correlation was observed in all conditions except after RV exposure (ρ = 0.11). In contrast, correlation between DNA methylation levels one meCpG and *SLC22A5* on chromosome 5 were only observed after SA treatment (S3 Fig).

Taken together these data suggest that meCpGs in airway epithelial cells that co-localize with asthma GWAS variants influence the expression of nearby genes, and that these effects can be reduced or enhanced by microbial exposures. These observations suggest that environment-specific epigenetic responses at meCpGs contribute to asthma pathogenesis.

meCpGs at *TSLP* co-localize with an asthma risk variant

To more deeply characterize associations between an meQTL-GWAS pair and better understand shared disease mechanisms between adult onset and childhood onset asthma, we focused on the four meQTL-GWAS pairs that co-localized in both the adult onset and childhood onset asthma GWASs (S3 Table). These pairs all included an intergenic SNP (rs1837253) located 5.7 kb upstream from the TSS of the *TSLP* gene on 5q22, encoding an epithelial cell cytokine that plays a key role in the inflammatory response in asthma and other allergic diseases [28]. rs1837253 co-localized with four meCpGs (cg15557878, cg10931190, cg15089387, cg25328184) in both the adult onset ($p_{GWAS} = 2.77x10^{-13}$) and childhood onset ($p_{GWAS} = 2.33x10^{-27}$) asthma GWASs. The four meCpGs are located in the first (untranslated) exon (5' UTR) of the *TSLP* gene (Fig 4), a region characterized as a promoter in keratinocyte primary cells (NHEK; ROADMAP). In fact, rs1837253 was the lead SNP within previous asthma GWASs (e.g. [12, 29]). The rs1837253-C asthma risk allele is associated with decreased methylation at each of the four meCpGs (Fig 4). Because the meQTL for each of the four meCpGs were observed in each treatment condition, we suggest that the first untranslated exon of *TSLP* is epigenetically poised for genotype-specific expression in epithelial cells.

Although this is the first study to annotate the asthma-associated rs1837253 as an meQTL, previous studies have shown *TSLP* to be a methylation-sensitive gene and that

hypomethylation at its promoter is associated with atopic dermatitis (AD) and prenatal tobacco smoke exposure [30, 31]. Another study showed that increased excretion of TSLP in primary cultured nasal epithelial cells after exposure to polyI:C (a dsRNA surrogate used to simulate viral infection) was associated with the CC genotype [32]. Our co-localization studies further suggest that the C allele is associated both with hypomethylation at CpG sites in an untranslated exon of *TSLP*, and with adult onset and childhood onset asthma. Furthermore, the lack of LD of SNPs with rs1837253 (± 50 kb) in European ($r^2 < 0.12$) and African American ($r^2 < 0.12$) 1000Genomes reference panels suggests that this SNP may indeed be causal for variation in DNA methylation levels at *TSLP* and its association with asthma.

Multi-trait co-localizations of molecular QTLs and asthma risk at the 17q12-21 asthma locus

Identifying regulatory effects at GWAS risk loci through co-localization with QTLs can help to understand molecular mechanisms of disease, including those that are modulated by environmental exposures. To further explore this, we focused on the co-localizations of eQTLs and meQTLs with asthma GWAS loci at 17q12-21 (17q), the most significant and most replicated locus for childhood onset asthma (reviewed in [33]). This locus is characterized by high LD across a 206.5 kb region encoding at least 10 genes (including *ORMDL3* and *GSDMB*). SNPs extending both proximal (including *PGAP3* and *ERBB2*) and distal (including *GSDMA*) to the core region show less LD with those in the core region but have been implicated as potentially independent asthma risk loci. Previous studies have shown that SNPs at this extended locus are eQTLs for at least four genes (*ORMDL3*, *GSDMB*, *GSDMA*, *PGAP3*) in blood cells

and/or lung cells [33] and that genetic risk at this locus is mediated through early life wheezing illness [34], particularly RV-associated wheezing [35].

We identified six co-localizations at the extended 17q locus of molecular QTLs with childhood onset asthma GWAS SNPs (Fig 5A). As expected at this childhood onset locus [36, 37], we did not find co-localizations with variants from the adult onset asthma GWAS. Among the co-localizations, one eQTL-GWAS pair was with rs12603332 on the 17q core haplotype and expression of *ORMDL3* (PPA>0.70; S4 Fig). The co-localized SNP (rs12603332) is in LD (r^2>0.65; 1000 Genomes European reference panel) with other previously reported asthma-associated GWAS SNPs in this region , including some that were reported as eQTLs for *ORMDL3* and *GSDMB*, primarily in immune cells. However, in our model none of the SNPs were eQTLs for *GSDMB*, which is typically co-regulated with *ORMDL3*, at least in blood cells [33]. In contrast to other studies, the asthma-associated allele (rs12603332-??) was associated with deceased expression of *ORMDL3* in our cell model. Moreover, the co-localizations with rs12603332 were only significant in vehicle treated cells, suggesting that exposure to RV or SA weakens the effects of rs12603332 genotype on *ORMDL3* expression in epithelial cells.

We also detected four meQTL-GWAS pairs among the six co-localizations at 17q that were associated with three meCpGs (cg24910161, cg21230266, cg17401724) and four SNPs at the distal end of (rs3902025, rs4239225, rs3859191) and beyond (rs66826786) the extended 17q haplotype where there is a breakdown of LD with SNPs in the core region. Two of these CpGs were each located 27 bp upstream (cg24910161) and in an intron (cg21230266) of *GSDMA* in regions characterized by ROADMAP as enhancers in NHEK cells. SNPs in modest to perfect LD (r^2_{range}=0.46 - 1.00; 1000 Genomes European panel) with these co-localizations (rs3902025, rs4239225, rs3859191) were described in previous studies as independent GWAS signals for

asthma (rs3894194) or as an eQTL for *GSDMA* (rs3859192) [4, 38, 39]. The meQTL-GWAS co-localization associated with cg24910161 was specific to the SAveh treatment, while those associated with cg21230266 were present in three conditions (RV, SA, and SAveh).

The one eQTL-meQTL-GWAS triplet in this region was associated with expression of *ERBB2* (Fig 5B), more than 361 Kb proximal to the co-localized asthma risk variant in an intron of *MED2* (rs66826786) and to the co-localized meCpG (cg17401724), distal to rs66826786 and 8.6 kb downstream of *MED2*. This is beyond the extended 17q locus as previously defined [33] in a region characterized by ROADMAP as both an enhancer and transcriptional start site (TSS) in NHEKs. Although the meQTL associated with this triplet was present in all four conditions, the eQTL for *ERBB2* is observed only after exposure to RV. The asthma risk allele at rs66826786-C was associated with both decreased *ERBB2* transcript in RV-treated cells and decreased DNA methylation of cg17401724 in all conditions (Fig 5). The 8.6 kb distance between the promoter of *ERBB2* and its eSNP (rs66826786) is highly suggestive of a long-range interaction between *ERBB2* and the region harboring cg17401724 and rs66826786. The fact that the eQTL is observed only after RV infection, further suggests that infection with RV triggers this long-range interaction in airway epithelial cells, likely via chromatin looping between these loci. The fact that the meQTL for cg17401724 is observed in all conditions suggests an epigenetically poised chromatin state at the distal end of the locus that directly affects transcription of *ERBB2* at the proximal end of the locus after exposure to RV, and possibly to other viral exposures.

Mendelian randomization of multi-trait co-localized triplets

Co-localization analyses reveal genetic variants that are associated with asthma and molecular traits (gene expression and/or DNA methylation) but the question of causality between the molecular traits remains unanswered. To infer causal relationships between DNA methylation and gene expression, we performed Mendelian randomization (MR), a method in which genetic variation associated with modifiable exposure patterns (i.e. DNA methylation) can be used as an instrumental variable to estimate directionality of effects between correlated traits (i.e. DNA methylation and gene expression) [40]. Specifically, we applied a 2-stage least squares regression (2SLS) to estimate the causal effects of DNA methylation (exposure) on gene expression (outcome) using the QTL SNP as the genetic instrument for each of the 14 co-localized triplets in each of the four conditions (see Materials and Methods). In this way, we are able to estimate whether the asthma risk variant has an effect on gene expression levels, mediated by DNA methylation.

Using MR, we detected a causal relationship between methylation and gene expression for each of the 14 triplets, indicating that DNA methylation at the meCpG mediates the genotype effects (eQTL) on gene expression (FDR < 0.10; Table 3). Specifically, for 12 of the triplets, we detected a causal relationship between methylation and gene expression in all four treatment conditions. These triplets were associated with four genes (*ACO2*, *GSTO2*, *IRF5*, *PMM1*) and nine meCpGs. For one triplet, the association with *FRK* was significant in the SA and SAveh treatments only, suggesting the methylation effects on this gene is specific to the culture conditions used for the SA experiment (SA+SAveh). For the remaining triplet, meCpG effects on *ERBB2* gene expression was only detected in RV-treated cells (FDR < 0.02, $lfsr_{RV} = 1.6x10^{-4}$), suggesting a long-range interaction after exposure to RV, as discussed above.

Table 3. Mendelian randomization results for 14 co-localized eQTL-meQTL-GWAS triplets

Condition-Specificity	Gene	CpG	rsID	SNP Position	P-value				FDR			
					RV	RVveh	SA	SAveh	RV	RVveh	SA	SAveh
All Conditions	ACO2	cg19274703	rs132905	22:41799106	2.00E-03	1.00E-03	0.00E+00	4.00E-03	3.61E-03	2.07E-03	0.00E+00	6.59E-03
	ACO2	cg07830128	rs4822038	22:41958495	2.80E-02	2.00E-03	0.00E+00	2.00E-03	3.34E-02	3.61E-03	0.00E+00	3.61E-03
	ACO2	cg10386501	rs5758461	22:42162189	7.00E-02	0.00E+00	0.00E+00	7.00E-03	7.69E-02	0.00E+00	0.00E+00	1.03E-02
	GSTO2	cg23659134	rs156697	10:106039185	2.50E-02	3.00E-03	0.00E+00	0.00E+00	3.04E-02	5.25E-03	0.00E+00	0.00E+00
	GSTO2	cg07488549	rs276210	10:106046403	1.30E-02	0.00E+00	0.00E+00	0.00E+00	1.69E-02	0.00E+00	0.00E+00	0.00E+00
	IRF5	cg26616347	rs3778754	7:128575552	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	PMM1	cg04809988	rs12483998	22:41935362	6.00E-03	4.20E-02	1.00E-03	0.00E+00	9.33E-03	4.70E-02	2.07E-03	0.00E+00
	PMM1	cg02738086	rs9607812	22:41941243	5.00E-03	3.80E-02	1.00E-03	0.00E+00	8.00E-03	4.34E-02	2.07E-03	0.00E+00
	PMM1	cg07830128	rs4822038	22:41958495	1.10E-02	1.00E-02	1.00E-03	0.00E+00	1.50E-02	1.40E-02	2.07E-03	0.00E+00
	PMM1	cg07830128	rs9607819	22:41958862	7.00E-03	4.00E-03	0.00E+00	1.00E-03	1.03E-02	6.59E-03	0.00E+00	2.07E-03
	PMM1	cg12016267	rs715498	22:42148467	1.30E-02	2.00E-03	0.00E+00	0.00E+00	1.69E-02	3.61E-03	0.00E+00	0.00E+00
	PMM1	cg10386501	rs5758461	22:42162189	2.90E-02	1.50E-02	9.00E-03	0.00E+00	3.38E-02	1.91E-02	1.29E-02	0.00E+00
Treatment (microbe)-specific	ERBB2	cg17401724	rs66826786	17:38206092	0.00E+00	8.05E-01	1.44E-01	2.36E-01	0.00E+00	8.05E-01	1.52E-01	2.45E-01
Culture condition-specific	FRK	cg20254830	rs10456902	6:116333742	8.04E-01	1.43E-01	0.00E+00	2.40E-02	8.05E-01	1.52E-01	0.00E+00	2.99E-02

383 Overall, the MR results provide additional, orthogonal evidence for co-localized triplets
384 and novel evidence for causal inference with respect to the co-localized traits (DNA methylation,
385 gene expression). These results reinforce arguments for epigenetic mechanisms of disease that
386 occur not only independent of exposures but also for epigenetic mechanisms that modify gene
387 expression in response to environmental exposures.

388

389 ## Discussion

390 One of the major challenges of genetic research is to uncover molecular mechanisms of disease
391 and to understand how genetic and environmental factors interact to influence these mechanisms
392 and individuals' risk for disease. Genome-wide association studies have identified thousands of
393 SNPs associated with complex diseases; however, the functions of non-coding SNPs identified in
394 GWASs, and therefore the molecular mechanisms in which they result in disease, is difficult to
395 infer from GWAS alone. Furthermore, other important contributors to disease pathophysiology
396 are not readily informed by GWAS, including epigenetic, environmental, and tissue- or cell type-
397 specific effects. Cell models address these limitations and advance our understanding of disease
398 pathobiology through experimental testing of disease mechanisms in controlled environments. In
399 this multi-omics study, we leveraged an airway epithelial cell model of microbial response to
400 identify functional variants that may have context-specific effects on transcriptional and
401 epigenetic responses and participate in molecular mechanisms that lead to a disease with an
402 underlying airway epithelial etiology. SNPs that were molecular QTLs in these models were co-
403 localized with adult onset and childhood onset asthma GWAS SNPs to identify 72 unique co-
404 localizations in at least one treatment condition. Integrating this information using Mendelian

405 randomization provided inferences into causality and insight into the molecular basis of
406 childhood onset asthma.

407 It is notable that we identified only six co-localizations with adult onset asthma GWAS
408 SNPs, compared to 70 with childhood onset asthma GWAS SNPs. None of the colocalizations in
409 the adult onset GWAS included an eQTL compared to 39 childhood onset co-localizations with
410 eQTLs, and only four meQTL-GWAS pairs were shared between adult and childhood onset
411 asthma, despite the fact the participants in our cell culture studies were all adults. Moreover,
412 even though there were 2.5-times the number of loci associated with childhood onset asthma
413 compared to adult onset asthma in the GWASs [12], we identified more than 11-times more co-
414 localizations using the childhood onset compared to the adult onset GWAS results (70 vs. 6,
415 respectively). These observations likely reflect the more important role of gene regulation and
416 dysregulation in airway epithelium in the etiology of childhood onset asthma compared to adult
417 onset asthma [41]. Focusing on other tissues (e.g., lung tissue) or cell types (e.g., immune cells)
418 might yield more co-localizations with adult onset GWAS SNPs or more shared co-localizations
419 between adult onset and childhood onset asthma.

420 Our study provides mechanistic evidence for associations between GWAS SNPs and
421 asthma at two important asthma loci: the *TSLP* locus at 5q22.1 and the 17q12-21 locus. Co-
422 localizations of the asthma associated SNP rs1837253 with DNA methylation levels in the *TSLP*
423 gene suggest an epigenetic mechanism of disease that contributes to both adult and childhood
424 onset asthma. Associations of this SNP with asthma have been highly replicated in GWASs and
425 TSLP is recognized as an important regulator in asthma pathogenesis through its broad effects on
426 innate and adaptive immune cells promoting Th2 inflammation [42]. Our data further show that
427 the effect of rs1837253 genotype on risk for both adult and childhood onset asthma is mediated

through DNA methylation levels at CpG sites in the untranslated first exon of the *TSLP* gene in airway epithelial cells. Finally, the lack of LD with other SNPs in a 100 kb window suggests that rs1837253 may indeed be the causal asthma SNP at this important locus.

Mendelian randomization of the childhood onset asthma 17q locus eQTL-meQTL-GWAS triplets revealed a novel epigenetic mechanism through which a SNP at the17q locus was associated with expression of *ERBB2* only after exposure to RV. This eQTL was mediated through differential methylation that was present in all treatment conditions. Previous studies have shown that variation at the 17q core locus confers risk to asthma only among children who experience wheezing illness in early life [36, 37], particularly with RV-associated wheezing [35]. Our study further connects RV infection and genotype at this locus but implicates for the first time interaction effects between genetic and methylation variation at the distal end of the locus and the expression of *ERBB2* at the proximal end of the locus only in RV infected epithelial cells. The SNP that is the eQTL for *ERBB2* in RV infected epithelial cells was also associated with asthma in the childhood onset asthma GWAS ($p_{GWAS} = 6.43x10^{-26}$), directly connecting the eQTL for *ERBB2* in RV-treated cells to asthma risk. The asthma associated allele, rs66826786-C, which was associated with decreased expression of *ERBB2* in our study (Fig 5B), is consistent with results of a study of 155 asthma cases and controls that reported an inverse correlation between *ERBB2* expression in *ex vivo* lower airway epithelial cells and asthma severity [43]. These combined data suggest that the low expression of *ERBB2* associated with asthma severity may be modulated by RV, the most common trigger of asthma exacerbations, via epigenetic mechanisms involving DNA methylation and long-range chromatin looping between the proximal and distal ends of this important locus. Our findings further highlight the importance of

RV exposure at this prominent asthma risk locus and provide mechanistic evidence for this genotype-exposure interaction.

Many associations in GWASs have small p-values that do not reach genome-wide significance ($p<5x10^{-8}$) but may be true signals. Distinguishing true positive from false signals for variants among these mid-hanging fruit can be challenging. Cell culture models provide a way to identify functional variants that regulate gene expression and/or epigenetic marks, establishing a framework for distinguishing true from false positive associations. In our study, over 72% of the co-localizations (52 total co-localizations; 22 eQTL-GWAS pairs; 17 meQTL-GWAS pairs; 13 triplets) were with a GWAS SNP that did not meet genome-wide significance in the GWAS for childhood onset asthma (GWAS p-value range $1.0x10^{-7}$ - $4.6x10^{-4}$; S1 Table). These co-localizations were associated with eight eGenes (*ACO2*, *FGFR4*, *FRK*, *GSTO2*, *IRF5*, *PMM1*, *POLI*) and 20 meCpGs. Notably, the majority of the co-localized triplets (13 of 14) were associated with SNPs that did not reach genome-wide significance, perhaps because the variants have exposure-specific or endotype-specific effects that are heterogeneous among subjects included in asthma GWASs. Annotating SNPs among the mid-hanging fruit for functionality provides a more complete picture of the genetic architecture of asthma and a paradigm for selecting loci for further studies.

Our study has several limitations. First, the sample sizes for the eQTL and meQTL studies were smaller than the minimum recommended by moloc ($n_{min}=300$) [15]. In such cases, *moloc* can miss true co-localizations in QTL datasets. For example, an eQTL-GWAS pair with supporting evidence may, in reality, be an eQTL-meQTL-GWAS triplet. As a result, the eQTL-GWAS and meQTL-GWAS pairs that we identified could be eQTL-meQTL-GWAS triplets that we were not powered to detect, or we may have missed other co-localizations entirely.

Nonetheless, the 72 unique co-localization detected in our study are likely to be real although replication studies in larger samples will increase the confidence in our findings. Second, we focused our studies on one cell type (upper airway sinonasal epithelium), four exposures (RV, SA, RVveh, SAveh), and one epigenetic mark (DNA methylation). It is likely that many of the co-localizations are not specific to airway epithelium or to these four conditions, and that additional epigenetic marks, such as those associated with chromatin accessibility, are involved in these or other co-localizations. Studies in other cell types and evaluation of additional exposures and epigenetic marks in larger sample sizes will be necessary to validate the cell- and condition-specific colocalizations identified here and extend these studies to additional conditions and molecular traits. Finally, characterizing chromatin conformational changes in airway epithelial cells before and after exposure to RV will allow a direct assessment of the chromatin looping at 17q that may occur in response to viral infection and potentially identify other context-specific interactions.

In summary, we identified *cis*-eQTLs and *cis*-meQTLs in an airway epithelial cell model of microbial response to RV and SA and integrated these data with asthma GWASs to assign potential molecular mechanisms for variants associated with asthma in two large GWASs. By combining co-localization analysis with Mendelian randomization, we provide robust statistical evidence of epigenetic mechanisms that contribute to childhood onset asthma, at least one of which is modulated by exposure to RV. We demonstrate that a multi-omics approach using disease-relevant cell types and exposures allows prioritization of disease-associated variants and provides insight into potential epigenetic mechanisms of asthma pathogenesis.

Materials and Methods

Ethics statement

Study participants were recruited between March 2012 and August 2015. Nasal specimens were collected as part of routine endoscopic endonasal surgeries. Informed written consent was obtained from each study participant and randomly generated ID codes were assigned to all samples thereby preserving the participant's anonymity and privacy. This study was approved by the institutional review boards of Northwestern University Feinberg School of Medicine and the University of Chicago.

Sample collection and composition

Sinonasal epithelial cells were obtained by brushing the uncinate process collected at elective surgery at Northwestern University from 68 males, 47 females, ages 18 – 73 years old (mean age 44), and self-reported ethnicities as Caucasian (67%), Black (16%), Hispanic (9%), and "other" (8%). Blood samples for genotyping were collected from study participants. A summary of the study design is shown in S1 Fig.

Upper airway epithelial cell culture and microbial treatments

After isolation, nasal airway epithelial cells were cultured in bronchial epithelial cell growth medium (Lonza, BEGM BulletKit, catalog number CC-3170) to near confluence, then frozen at -80°C and stored in Liquid Nitrogen. Cells were subsequently thawed and cultured in collagen-coated (PureCol, INAMED BioMaterials, catalog number 5,409, 3 mg/mL, 1:15 dilution) tissue culture plates (6 wells of 2x 12 well plates) using BEGM overnight at 37°C and 5% CO_2. In preparation for rhinovirus (HRV-16; RV) RV infection/stimulation, plates at 50-60% confluency

519 were incubated overnight in BEGM without hydrocortisone (HC) followed by a two-hour RV
520 infection at a multiplicity of infection (MOI) of 2 and vehicle treatment (Bronchial epithelial cell
521 basal medium (BEBM) + Gentamicin/Amphotericin) at 33°C (low speed rocking, ~15 RPM).
522 RV- and vehicle-treated cells were washed and then were cultured at 33°C for 46 hours (48 hours
523 total) in BEGM without HC. Prior to heat-killed *Staphylococcus aureus* (SA, Life Technologies,
524 catalog number S-2859)-stimulation, cells were cultured to near 100% confluency in BEGM and
525 were further incubated without HC overnight at 37°C. Cells were then stimulated with SA ($5e^8$
526 particles/mL) and vehicle control (sterile 1x dPBS) for 24 hours at 37 °C in 5% CO_2.
527
528 **Genotyping and imputation**
529 DNA was extracted from whole blood or sinus tissue (if no blood was available) with the
530 Macherey-Nagel NucleoSpin Blood L or NucleoSpin Tissue L Extraction kits, respectively, and
531 quantified with the NanoDrop ND1000. Genotyping of all study participants was performed
532 using the Illumina Infinium HumanCore Exome+Custom Array (550,224 SNPs). After quality
533 control (QC) (excluding SNPs with HWE < 0.0001 by race/ethnicity, call rate < 0.95, MAF <
534 0.05 and individuals with genotype call rates < 0.05), 529,993 markers for 115 individuals were
535 available for analysis. Ancestral principal component analysis (PCA) was performed using 676
536 ancestral informative markers included on the array that overlap with the HapMap release 3 (S5
537 Fig).
538 Phasing and imputation were performed using the ShapeIt2 [44] and Impute2 [45]
539 software packages, respectively. Variants were imputed in 5 Mb windows across the genome
540 against the 1000 Genomes Phase 3 haplotypes (Build 37; October 2014). Individuals were
541 categorized into two groups based on the k-means clustering of ancestral PCs, using the

kmeans() function in R; individuals were grouped as European or African American based on how they related to the HapMap reference panel and means clustering of their ancestral PCs (S5 Fig). After imputation, both groups were merged and QC was performed with gtool [46]. X and Y chromosome-linked SNPs and SNPs that did not meet the QC criteria (info score < 0.8, MAF < 0.05, missingness > 0.05 and a probability score < 0.9) were excluded from analyses. Probability scores were converted to dosages for 6,665,552 of the remaining sites used in downstream analyses.

RNA extraction and sequencing

Following RV and SA treatments, RNA from cells underwent extraction and purification using the QIAGEN AllPrep DNA/RNA Kit. RNA quality and quantity were measured at the University of Chicago Functional Genomics Core using the Agilent RNA 6000 Pico assay and the Agilent 2100 Bioanalyzer. RNA integrity numbers (RIN) were greater than 7.7 for all samples. cDNA libraries were constructed using the Illumina TruSeq RNA Library Prep Kit v2 and sequenced on the Illumina HiSeq 2500 System (50 bp, single-end); RNA sequencing was completed at the University of Chicago Genomics Core. Subsequently, we checked for potential sample contamination and sample swaps using the publicly available software VerifyBamID (http://genome.sph.umich.edu/wiki/VerifyBamID) [47] for cells from all 115 individuals included in each treatment condition. We did not detect any cross-contamination between samples but we did identify one sample swap between individuals, which we subsequently corrected.

Sequences were mapped to the human reference genome (hg19) and reads per gene were quantified using the Spliced Transcripts Alignment to a Reference (STAR) [48] software. X,Y,

and mitochondrial chromosome genes, and low count data (genes < 1CPM) were removed prior to normalization via the trimmed mean of M-values method (TMM) and variance modeling (voom) [49]; samples contain > 8M mapped reads. Principle components analysis (PCA) identified biological and technical sources of variation in the voom-normalized RNA-seq reads. We identified contributors to batch and other technical effects in the RV experiment (days in liquid nitrogen, experimental culture days, cell culture batches, RNA concentration, RNA fragment length, technician, sequencing pool) and SA experiment (day in liquid nitrogen, experimental culture days, cell culture batches, RNA concentration, RIN score, cDNA library concentration). Sex was a significant variable in the PCA for the RV experiment. Additionally, unknown sources of variation were predicted with the Surrogate Variable Analysis (SVA) [50] package in R where 15 and 21 surrogate variables (SVs) were estimated for the RV and SA experiments, respectively. Voom-normalized RNA-seq data were then adjusted for technical effects, SVs, sex, and ancestral PCs (1-3) using the function removeBatchEffect() from the R package limma [51].

DNA extraction and methylation profiling

Following RV and SA treatments, DNA was extracted from cells and purified using the QIAGEN AllPrep DNA/RNA Kit. DNA methylation profiles for cells from each treatment were measured on the Illumina Infinium MethylationEPIC BeadChip at the University of Chicago Functional Genomics Core. Methylation data were preprocessed using the minfi package [52]. Probes located on sex chromosomes and with detection p-values greater than 0.01 in more than 10% of samples were removed from the analysis; samples with more than 5% missing probes were also removed. A preprocessing control normalization function was applied to correct for

raw probe values or background and a Subset-quantile Within Array Normalization (SWAN) [53] was used to correct for technical differences between the Infinium type I and type II probes. Additionally, we removed cross-reactive probes and probes within two nucleotides of a SNP with an MAF greater than 0.05 using the function rmSNPandCH() from the R package DMRcate [54].

PCA identified technical and biological sources of variation in the normalized DNA methylation datasets. We identified contributors to batch and technical effects in the RV experiment (array, cell harvest date) and SA experiment (day in liquid nitrogen, array, cell harvest date, DNA concentration). Sex, age, and smoking were significant variables in the PCA across each treatment condition. Unknown sources of variation were predicted with the SVA package where 37 SVs were estimated for both the RV and SA experiments. SWAN and quantile-normalized M-values were then adjusted for batch and technical effects, SVs, sex, age, smoking and steroid use (for the SA dataset) using the function removeBatchEffect() in R.

eQTL and meQTL analyses

Prior to e/meQTL analysis, voom-transformed gene expression values and normalized methylation M-values were adjusted for potential batch, technical, and biological variables as described above. Linear regression between the permuted genotypes and molecular phenotypes (gene expression and methylation residuals) from each treatment condition was performed with the FastQTL [55] software package within *cis*-window sizes of 1 Mb and 10 Kb for eQTL and meQTL analyses, respectively. Nominal passes were conducted for each eQTL and meQTL analysis within FastQTL, and an FDR threshold of 0.10 was applied to adjust for multiple testing within each experimental dataset with the p.adjust() function in R.

611

612 **Multivariate adaptive shrinkage analysis (mash)**

613 An Empirical Bayes method of multivariate adaptive shrinkage was applied separately to the
614 eQTL and meQTL data sets as implemented in the R statistical package, mashr
615 (https://github.com/stephenslab/mashr) [20], to produce improved estimates of QTL effects and
616 corresponding significance values in each treatment condition. Mashr implements this in two
617 general steps: 1) identification of pattern sharing, sparsity, and correlation among QTL effects,
618 and 2) integration of these learned patterns to produce improved effects estimates and measures
619 of significance for eQTLs or meQTLs in each treatment condition. To fit the mash model, we
620 identified eQTLs and meQTLs at an FDR < 20% in at least one treatment condition to generate a
621 list of covariance matrices, constructed to represent patterns of effects in the data which included
622 both 'data-driven' and 'canonical' estimates (see [20]). The instructions found in the *mashr* data-
623 driven vignette were followed to run mash.

624

625 **Co-localization analysis**

626 To estimate the posterior probability that a variant was contributing to the signal of a genetic
627 variant was also associated with asthma, gene expression, and/or DNA methylation, we applied a
628 Bayesian statistical framework implemented in the R package multiple-trait-coloc (moloc) [15].
629 Summary data from adult onset and childhood onset asthma GWAS from [12], along with eQTL
630 and meQTL summary data from cells within each treatment condition (described above), were
631 included in the *moloc* analysis. Each co-localization analysis included summary data from a
632 GWAS and epithelial cell eQTLs and meQTLs from corresponding treatment conditions.
633 Because a genome-wide co-localization analysis was computationally untenable, genomic

regions for co-localization were defined using the GWAS Analysis of Regulatory of Functional Information Enrichment with LD correction (GARFIELD) package implemented in R [56] . First, we analyzed the enrichment pattern of e/meSNPs from all four treatment conditions in adult onset and childhood onset GWASs using the default package settings. Second, we extracted variants driving the enrichment signals at a GWAS p-value threshold of $1x10^{-4}$. Regions were defined as 2 Mb windows centered around these variants. Only regions with at least 10 SNPs in common between all three datasets or 'traits' (GWAS, eQTL, and meQTL) were assessed by moloc and 15 'configurations' of possible variant sharing was computed across these three traits (see [15] for more details). PPAs $\geqq$ 70% were considered as evidence for co-localization. Prior probabilities of $1x10^{-4}$, $1x10^{-6}$, and $1x10^{-7}$ were chosen for the association of one, two, or three traits, respectively, as recommended by the authors of moloc.

Mendelian randomization

Mendelian Randomization was performed using the ivreg2 function in R (https://www.r-bloggers.com/an-ivreg2-function-for-r/) which applies 2SLS regression, as implemented in [24]. We co-localized triplets (eQTL-meQTL-GWAS) to assess the causal effects of DNA methylation on gene expression, using genotypes as the instrumental variable. P-values were adjusted using the FDR method in the p.adjust() function in R. An FDR of less than 0.05 was considered to be significant.

Acknowledgments

The authors acknowledge Christine Billstrand and Raluca Nicolae for sample processing and library preparation, and study subjects for their participation.

657

658 # References

659 1. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI
660 Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids
661 research. 2017;45(D1). doi: 10.1093/nar/gkw1133.

662 2. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic
663 localization of common disease-associated variation in regulatory DNA. Science.
664 2012:1222794. doi: 10.1126/science.1222794.

665 3. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are
666 more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet.
667 2010;6(4):e1000888. Epub 2010/04/07. doi: 10.1371/journal.pgen.1000888. PubMed PMID:
668 20369019; PubMed Central PMCID: PMCPMC2848547.

669 4. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
670 multitissue gene regulation in humans. Science. 2015;348(6235):648-60. Epub 2015/05/09.
671 doi: 10.1126/science.1262110. PubMed PMID: 25954001; PubMed Central PMCID:
672 PMCPMC4547484.

673 5. Ober C. Asthma Genetics in the Post-GWAS Era. Annals of the American Thoracic Society.
674 2016;13 Suppl 1(Supplement 1):90. doi: 10.1513/AnnalsATS.201507-459MG.

675 6. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a
676 genetic journey. Human molecular genetics. 2008;17(R2):65. doi: 10.1093/hmg/ddn289.

677 7. Bonnelykke K, Ober C. Leveraging gene-environment interactions and endotypes for
678 asthma gene discovery. J Allergy Clin Immunol. 2016;137(3):667-79. Epub 2016/03/08.
679 doi: 10.1016/j.jaci.2016.01.006. PubMed PMID: 26947980; PubMed Central PMCID:
680 PMCPMC5004762.

681 8. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science.
682 2004;306(5696):636-40. Epub 2004/10/23. doi: 10.1126/science.1105136. PubMed PMID:
683 15499007.

684 9. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al.
685 Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317-30.
686 Epub 2015/02/20. doi: 10.1038/nature14248. PubMed PMID: 25693563; PubMed Central
687 PMCID: PMCPMC4530010.

10. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017;49(12):1752-7. Epub 2017/10/31. doi: 10.1038/ng.3985. PubMed PMID: 29083406; PubMed Central PMCID: PMCPMC5989923.

11. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet. 2018;50(1):42-53. Epub 2017/12/24. doi: 10.1038/s41588-017-0014-7. PubMed PMID: 29273806; PubMed Central PMCID: PMCPMC5901974.

12. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood onset and adult onset asthma: Genome- and Transcriptome-wide Studies. The Lancet Respiratory Medicine, in press. 2018. https://www.biorxiv.org/content/10.1101/427427v1.

13. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc Natl Acad Sci U S A. 2012;109(4):1204-9. Epub 2012/01/12. doi: 10.1073/pnas.1115761109. PubMed PMID: 22233810; PubMed Central PMCID: PMC3268270.

14. Kim-Hellmuth S, Bechheim M, Putz B, Mohammadi P, Nedelec Y, Giangreco N, et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. Nat Commun. 2017;8(1):266. Epub 2017/08/18. doi: 10.1038/s41467-017-00366-1. PubMed PMID: 28814792; PubMed Central PMCID: PMCPMC5559603.

15. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, et al. A Bayesian Framework for Multiple Trait Colo-calization from Summary Association Statistics. Bioinformatics. 2018. Epub 2018/03/27. doi: 10.1093/bioinformatics/bty147. PubMed PMID: 29579179.

16. Jackson DJ, Gangnon RE, Evans MD, Roberg KA, Anderson EL, Pappas TE, et al. Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. Am J Respir Crit Care Med. 2008;178(7):667-72. Epub 2008/06/21. doi: 10.1164/rccm.200802-309OC. PubMed PMID: 18565953; PubMed Central PMCID: PMCPMC2556448.

17. Johnston SL, Pattemore PK, Sanderson G, Smith S, Lampe F, Josephs L, et al. Community study of role of viral infections in exacerbations of asthma in 9-11 year old children. BMJ. 1995;310(6989):1225-9. Epub 1995/05/13. PubMed PMID: 7767192; PubMed Central PMCID: PMCPMC2549614.

18. Busse WW, Lemanske RF, Jr., Gern JE. Role of viral respiratory infections in asthma and asthma exacerbations. Lancet. 2010;376(9743):826-34. Epub 2010/09/08. doi:

10.1016/S0140-6736(10)61380-3. PubMed PMID: 20816549; PubMed Central PMCID: PMCPMC2972660.

19. Davis MF, Peng RD, McCormack MC, Matsui EC. Staphylococcus aureus colonization is associated with wheeze and asthma among US children and young adults. The Journal of allergy and clinical immunology. 2015;135(3):811-3.e5. Epub 2014/12/20. doi: 10.1016/j.jaci.2014.10.052. PubMed PMID: 25533526.

20. Urbut SM, Wang G, Carbonetto P, Stephens M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat Genet. 2019;51(1):187-95. Epub 2018/11/28. doi: 10.1038/s41588-018-0268-8. PubMed PMID: 30478440; PubMed Central PMCID: PMCPMC6309609.

21. Stephens M. False discovery rates: a new deal. Biostatistics. 2017;18(2):275-94. Epub 2016/10/21. doi: 10.1093/biostatistics/kxw041. PubMed PMID: 27756721; PubMed Central PMCID: PMCPMC5379932.

22. Schubeler D. Function and information content of DNA methylation. Nature. 2015;517(7534):321-6. Epub 2015/01/17. doi: 10.1038/nature14192. PubMed PMID: 25592537.

23. Martin EM, Fry RC. Environmental Influences on the Epigenome: Exposure- Associated DNA Methylation in Human Populations. Annu Rev Public Health. 2018;39:309-33. Epub 2018/01/13. doi: 10.1146/annurev-publhealth-040617-014629. PubMed PMID: 29328878.

24. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight. 2016;1(20):e90151. Epub 2016/12/13. doi: 10.1172/jci.insight.90151. PubMed PMID: 27942592; PubMed Central PMCID: PMCPMC5139904.

25. DeVries A, Wlasiuk G, Miller SJ, Bosco A, Stern DA, Lohman IC, et al. Epigenome-wide analysis links SMAD3 methylation at birth to asthma in children of asthmatic mothers. J Allergy Clin Immunol. 2017;140(2):534-42. Epub 2016/12/25. doi: 10.1016/j.jaci.2016.10.041. PubMed PMID: 28011059; PubMed Central PMCID: PMCPMC5479765.

26. Yang IV, Pedersen BS, Liu A, O'Connor GT, Teach SJ, Kattan M, et al. DNA methylation and childhood asthma in the inner city. J Allergy Clin Immunol. 2015;136(1):69-80. Epub 2015/03/15. doi: 10.1016/j.jaci.2015.01.025. PubMed PMID: 25769910; PubMed Central PMCID: PMCPMC4494877.

27. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465-76. Epub 2008/05/09. doi: 10.1038/nrg2341. PubMed PMID: 18463664.

28. Ying S, O'Connor B, Ratoff J, Meng Q, Mallett K, Cousins D, et al. Thymic stromal lymphopoietin expression is increased in asthmatic airways and correlates with expression of Th2-attracting chemokines and disease severity. J Immunol. 2005;174(12):8183-90. Epub 2005/06/10. PubMed PMID: 15944327.

29. 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(9):887-92. Epub 2011/08/02. doi: 10.1038/ng.888. PubMed PMID: 21804549; PubMed Central PMCID: PMCPMC3445408.

30. Wang IJ, Chen SL, Lu TP, Chuang EY, Chen PC. Prenatal smoke exposure, DNA methylation, and childhood atopic dermatitis. Clin Exp Allergy. 2013;43(5):535-43. Epub 2013/04/23. doi: 10.1111/cea.12108. PubMed PMID: 23600544.

31. Luo Y, Zhou B, Zhao M, Tang J, Lu Q. Promoter demethylation contributes to TSLP overexpression in skin lesions of patients with atopic dermatitis. Clin Exp Dermatol. 2014;39(1):48-53. Epub 2013/12/18. doi: 10.1111/ced.12206. PubMed PMID: 24341479.

32. Hui CC, Yu A, Heroux D, Akhabir L, Sandford AJ, Neighbour H, et al. Thymic stromal lymphopoietin (TSLP) secretion from human nasal epithelium is a function of TSLP genotype. Mucosal Immunol. 2015;8(5):993-9. Epub 2014/12/18. doi: 10.1038/mi.2014.126. PubMed PMID: 25515628.

33. Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, et al. A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle. J Allergy Clin Immunol. 2018. Epub 2018/01/09. doi: 10.1016/j.jaci.2017.12.974. PubMed PMID: 29307657.

34. Smit LA, Bouzigon E, Pin I, Siroux V, Monier F, Aschard H, et al. 17q21 variants modify the association between early respiratory infections and asthma. Eur Respir J. 2010;36(1):57-64. Epub 2009/12/25. doi: 10.1183/09031936.00154509. PubMed PMID: 20032010.

35. Caliskan M, Bochkov YA, Kreiner-Moller E, Bonnelykke K, Stein MM, Du G, et al. Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. N Engl J Med. 2013;368(15):1398-407. Epub 2013/03/29. doi: 10.1056/NEJMoa1211592. PubMed PMID: 23534543; PubMed Central PMCID: PMCPMC3755952.

36. Bouzigon E, Corda E, journal … A-H. Effect of 17q21 variants and smoking exposure in early-onset asthma. New England journal …. 2008. doi: 10.1056/NEJMoa0806604.

37. Loss GJ, Depner M, Hose AJ, Genuneit J, Karvonen AM, Hyvarinen A, et al. The Early Development of Wheeze. Environmental Determinants and Genetic Susceptibility at 17q21.

793 Am J Respir Crit Care Med. 2016;193(8):889-97. Epub 2015/11/18. doi:
794 10.1164/rccm.201507-1493OC. PubMed PMID: 26575599.

795 38. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale,
796 consortium-based genomewide association study of asthma. N Engl J Med.
797 2010;363(13):1211-21. Epub 2010/09/24. doi: 10.1056/NEJMoa0906312. PubMed PMID:
798 20860503; PubMed Central PMCID: PMCPMC4260321.

799 39. Marinho S, Custovic A, Marsden P, Smith JA, Simpson A. 17q12-21 variants are associated
800 with asthma and interact with active smoking in an adult population from the United
801 Kingdom. Ann Allergy Asthma Immunol. 2012;108(6):402-11 e9. Epub 2012/05/26. doi:
802 10.1016/j.anai.2012.03.002. PubMed PMID: 22626592.

803 40. Smith GD. Mendelian Randomization for Strengthening Causal Inference in Observational
804 Studies: Application to Gene x Environment Interactions. Perspect Psychol Sci.
805 2010;5(5):527-45. Epub 2010/09/01. doi: 10.1177/1745691610383505. PubMed PMID:
806 26162196.

807 41. Busse WW. The atopic march: Fact or folklore? Ann Allergy Asthma Immunol.
808 2018;120(2):116-8. Epub 2018/02/08. doi: 10.1016/j.anai.2017.10.029. PubMed PMID:
809 29413332.

810 42. West EE, Kashyap M, Leonard WJ. TSLP: A Key Regulator of Asthma Pathogenesis. Drug
811 Discov Today Dis Mech. 2012;9(3-4). Epub 2013/12/19. doi: 10.1016/j.ddmec.2012.09.003.
812 PubMed PMID: 24348685; PubMed Central PMCID: PMCPMC3859144.

813 43. Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, et al. Gene
814 expression correlated with severe asthma characteristics reveals heterogeneous mechanisms
815 of severe disease. American journal of respiratory and critical care medicine.
816 2017;195(11):1449-63. doi: 10.1164/rccm.201607-1407OC.

817 44. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and
818 population genetic studies. Nat Methods. 2013;10(1):5-6. Epub 2012/12/28. doi:
819 10.1038/nmeth.2307. PubMed PMID: 23269371.

820 45. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3
821 (Bethesda). 2011;1(6):457-70. Epub 2012/03/03. doi: 10.1534/g3.111.001198. PubMed
822 PMID: 22384356; PubMed Central PMCID: PMCPMC3276165.

823 46. Freeman C, Marchini J. GTOOL: A program for transforming sets of genotype data for use
824 with the programs SNPTEST and IMPUTE, Oxford, UK. GTOOL: A program for
825 transforming sets of genotype data for use with the programs SNPTEST and IMPUTE,
826 Oxford, UK.

47. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet. 2012;91(5):839-48. Epub 2012/10/30. doi: 10.1016/j.ajhg.2012.09.004. PubMed PMID: 23103226; PubMed Central PMCID: PMCPMC3487130.

48. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. Curr Protoc Bioinformatics. 2015;51:11 4 1-9. Epub 2015/09/04. doi: 10.1002/0471250953.bi1114s51. PubMed PMID: 26334920; PubMed Central PMCID: PMCPMC4631051.

49. Law CW, Chen Y, Genome … SW. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. 2014.

50. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3. Epub 2012/01/20. doi: 10.1093/bioinformatics/bts034. PubMed PMID: 22257669; PubMed Central PMCID: PMCPMC3307112.

51. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. Epub 2015/01/22. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510.

52. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9. Epub 2014/01/31. doi: 10.1093/bioinformatics/btu049. PubMed PMID: 24478339; PubMed Central PMCID: PMCPMC4016708.

53. Maksimovic J, Gordon L. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. 2012.

54. Peters TJ, Buckley MJ. De novo identification of differentially methylated regions in the human genome. De novo identification of differentially methylated regions in the human genome. 2015.

55. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 2016;32(10):1479-85. Epub 2015/12/29. doi: 10.1093/bioinformatics/btv722. PubMed PMID: 26708335; PubMed Central PMCID: PMCPMC4866519.

56. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nat Genet. 2019;51(2):343-53. Epub 2019/01/30. doi: 10.1038/s41588-018-0322-6. PubMed PMID: 30692680.

866 **Fig 1. Summary of eQTL and meQTL mapping results.** Venn diagrams of overlapping
867 eQTLs (A) and meQTLs (B) in each condition (FDR<0.10). (C) Summary of eQTL and meQTL
868 mapping results for each treatment condition.
869
870 **Fig 2. Molecular QTL effects sharing across treatment conditions (lfsr<0.05).** Heatmaps
871 showing pairwise comparison of eQTL (A) and meQTL (B) pairwise sharing between treatment
872 conditions. (C) Examples of RV- (left) and SA-specific (right) eQTLs. (D) Examples of vehicle-
873 (left) and SA-specific (right) meQTLs.
874
875 **Fig 3. Spearman correlation plots showing correlations between co-localized meCpGs and**
876 **nearby genes in each treatment condition.** Scatterplots showing the correlation of DNA
877 methylation levels and gene expression in each treatment condition. (A) cg23659134 and
878 *GSTO2*. (B) cg26616347 and *IRF5* expression. In all panels, the y-axis shows methylation levels
879 and x-axis shows gene expression levels. See text for details.
880
881 **Fig 4. Co-localization of rs1837253 with DNA methylation levels of four meCpGs at *TSLP*.**
882 rs1837253 (red vertical bar) is associated with DNA methylation levels at four meCpGs (yellow
883 vertical bars). Box plots show DNA methylation levels (y-axes) for each meCpGs by rs1837253
884 genotype (x-axes) in each treatment condition.
885
886 **Fig 5. Co-localizations at the 17q asthma susceptibility locus.** (A) e/meSNPs and CpGs
887 included in each of the six co-localizations are shown by vertical colored lines. Solid lines
888 indicate the location of the SNPs; dashed lines show the location of CpGs. SNP-CpG co-

889 localizations are indicated by vertical lines of the same color. A single eQTL-GWAS pair
890 (rs12603332) for *ORDML3* is shown in orange. The three meQTL-GWAS pairs are shown in
891 green (rs66826786; cg1740124), purple (rs4239225; cg21230266), and turquoise (rs2859191;
892 cg21230266); cg21230266, shown in purple and turquoise, co-localizes with both rs4239225 and
893 rs2859191. The SNP (rs66828786) and meCpG (cg17401724) associated with the triplet
894 containing *ERBB2* is shown in red. (B) Box plots for eQTLs (top row) and meQTLs (bottom
895 row) associated with the eQTL-meQTL-GWAS triplet. The SNP in this triplet is associated with
896 *ERBB2* expression only in the RV-treated cells, but with DNA methylation levels (cg17401724)
897 in all treatments.
898
899
900
901
902
903
904

Supporting information

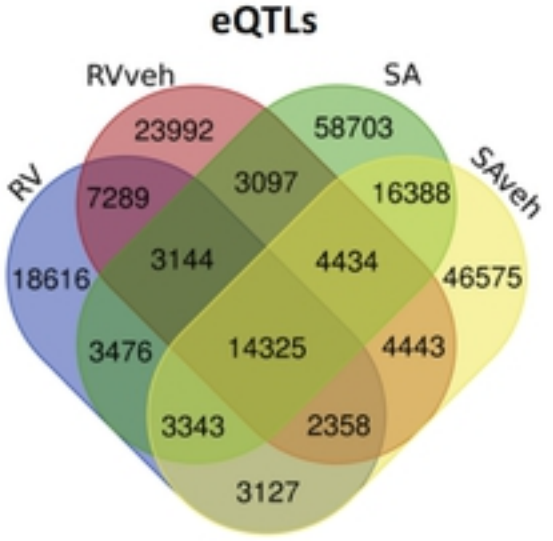
S1 Fig. Overview of the e/meQTL and colocalization studies in upper airway epithelial cells treated with RV and SA. (A) Step-wise experimental design to identify treatment-specific e/meQTLs in airway epithelial cells from 115 individuals: 1. Nasal epithelial cells collected from study participants were cultured and treated with either RV or SA for 48 and 24 hours, respectively. 2. Gene expression and DNA methylation measured in nasal epithelial cells from each treatment condition. 3. Genotype profiling to identify genetic variation influencing gene expression and DNA methylation to RV- and SA-treatment. 4. QC and analyses including e/meQTL mapping, multi-trait co-localization analysis, and Mendelian randomization. (B) Venn diagram of asthma and atopy status for study subjects. (C) Breakdown of the number of subjects for each experiment (RV and SA) and molecular QTL mapping.

S2 Fig. Summary of molecular QTL effects sharing across treatment conditions (lfsr<0.05). (A) Number of eQTLs and corresponding eGenes and eSNPs shared between treatment conditions. (B) Number of meQTLs and corresponding meCpGs and meSNPs shared between treatment conditions.

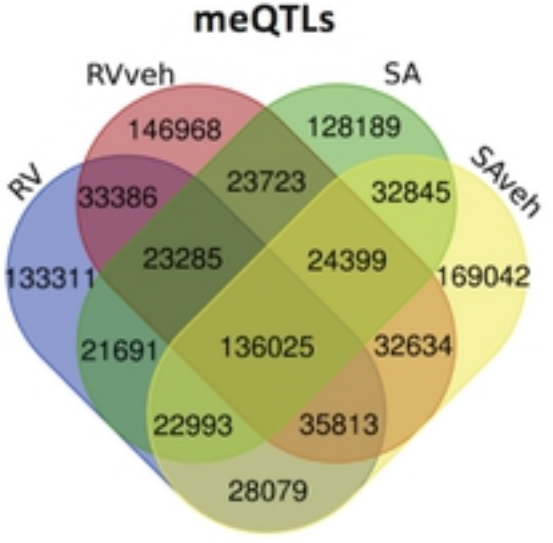
S3 Fig. Scatterplots showing the Spearman correlation of DNA methylation levels for meCpGs of co-localized meQTLs and gene expression of nearby genes for each treatment condition. Examples of treatment and culture effects on DNA methylation and gene expression correlations. (A) Correlations for DNA methylation and *FRK* gene expression suggest culture-specific effects. (B and C) Correlations for DNA methylation and *GSTO2* do not show preference to any treatment condition. Although the meCpGs (cg23659134, cg07488549) that

928 are correlated with *GSTO2* gene expression are both located intergenically and within the same
929 intron as *GSTO2*, the direction of their effects on *GSTO2* expression are in opposing directions.
930 (D-F) DNA methylation and gene expression correlations suggestive of RV-specific effects in
931 which the correlation is reduced after treatment with RV.
932
933 **S4 Fig. Association of *ORMDL3* expression and genotype.** Box plots of an eQTL for
934 *ORMDL3* and rs12603332 for RV, RVveh, SA, and SAveh treatment conditions (A-D,
935 respectively).
936
937 **S5 Fig. PCA and k-means clustering of genotypes.** (A) PCA plot of study participant's
938 genotypes (circles) projected on HapMap genotypes (squares). (B) Scree plot of k-means
939 clustering of ancestral PCs in which the within groups sum of squares (y-axis) is plotted against
940 the number of potential group clusters (x-axis); using the 'elbow criterion', it is determined that
941 two clusters are best representative of how many clusters we can group the study samples. (C)
942 PCA plot of study participants grouped into two cluster for genotype imputation, European (red),
943 and African American (Blue), according to the k-means clustering criterion.
944
945 **S1 Table.** *moloc* results indicating molecular QTL-GWAS pars and triplets.
946 **S2 Table.** Gene expression and meCpG Spearman correlations.
947 **S3 Table.** Shared meQTL-GWAS pairs between adult onset and childood onset asthma.

A



B



C

eQTL and meQTL mapping results (FDR<0.10)

	RV	RVveh	SA	SAveh
eQTLs	55,678	63,082	106,910	94,993
eSNPs	52,519	59,689	97,627	87,971
eGenes	1,637	1,873	2,599	2,406
meQTLs	434,483	456,233	413,150	481,830
meSNPs	306,850	320,174	293,276	336,901
meCpGs	40,789	42,038	38,501	44,840

Fig 1

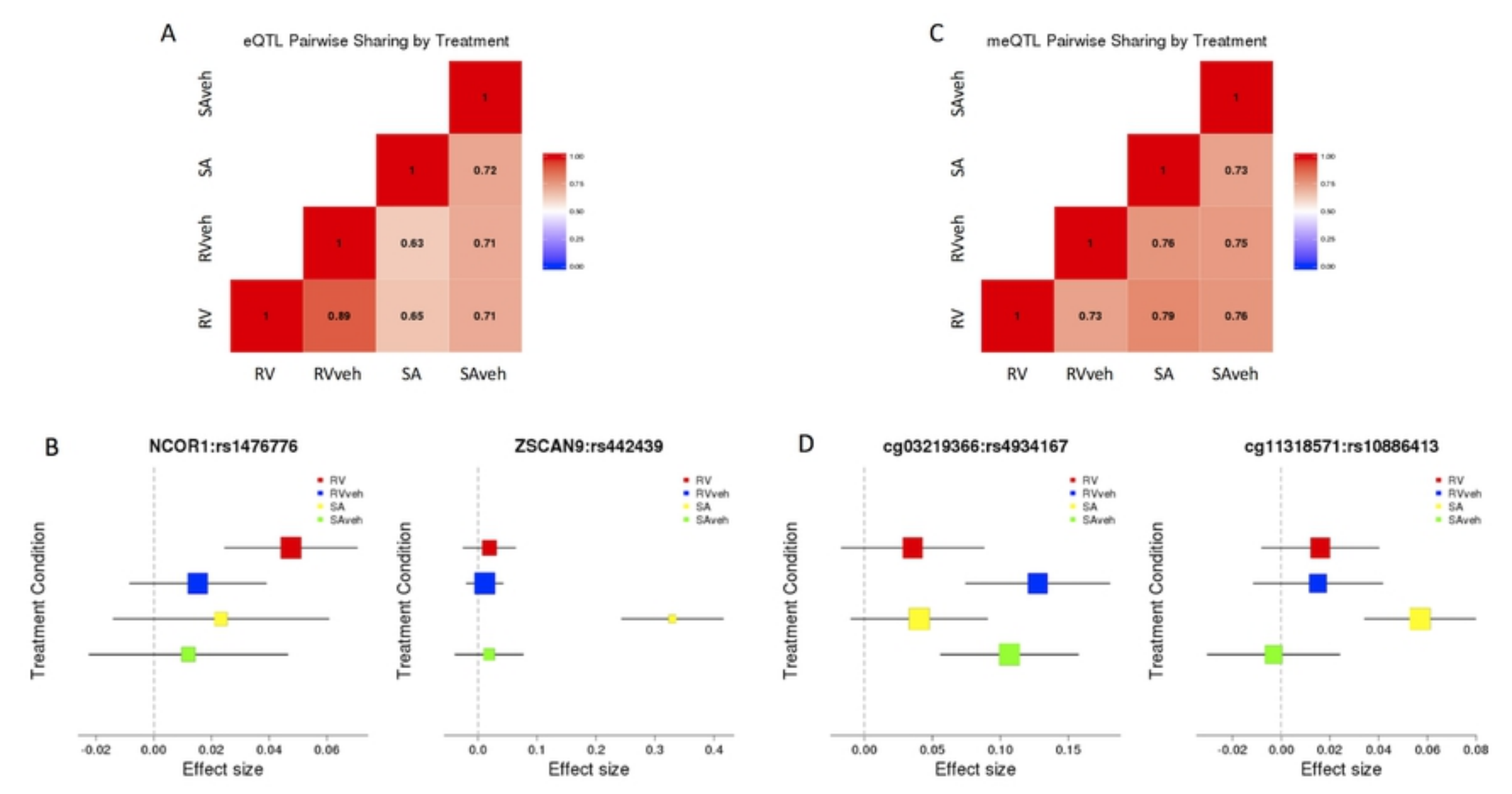

A
eQTL Pairwise Sharing by Treatment
SAveh
SA
RVveh
RV
1
0.72
0.63
0.71
0.89
0.65
0.71
RV
RVveh
SA
SAveh
C
meQTL Pairwise Sharing by Treatment
1
0.73
0.76
0.75
0.73
0.79
0.76
B
NCOR1:rs1476776
ZSCAN9:rs442439
D
cg03219366:rs4934167
cg11318571:rs10886413
Treatment Condition
RV
RVveh
SA
SAveh
Effect size
-0.02
0.00
0.02
0.04
0.06
0.0
0.1
0.2
0.3
0.4
0.05
0.10
0.15
0.08

Fig 2

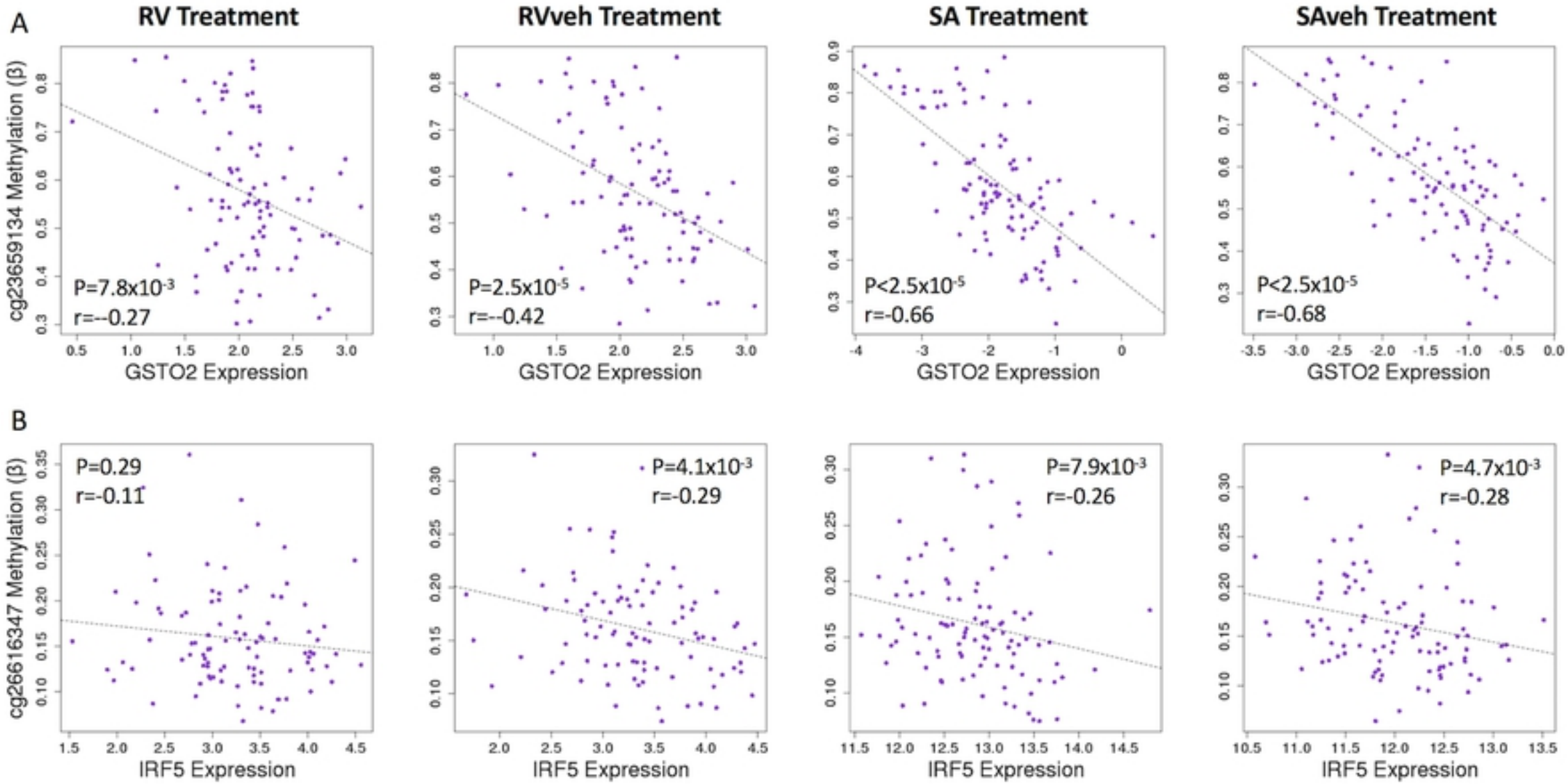

A
RV Treatment
RVveh Treatment
SA Treatment
SAveh Treatment
cg23659134 Methylation (β)
P=7.8x10-3
r=-0.27
P=2.5x10-5
r=-0.42
P<2.5x10-5
r=-0.66
P<2.5x10-5
r=-0.68
GSTO2 Expression
B
cg26616347 Methylation (β)
P=0.29
r=-0.11
P=4.1x10-3
r=-0.29
P=7.9x10-3
r=-0.26
P=4.7x10-3
r=-0.28
IRF5 Expression

Fig 3

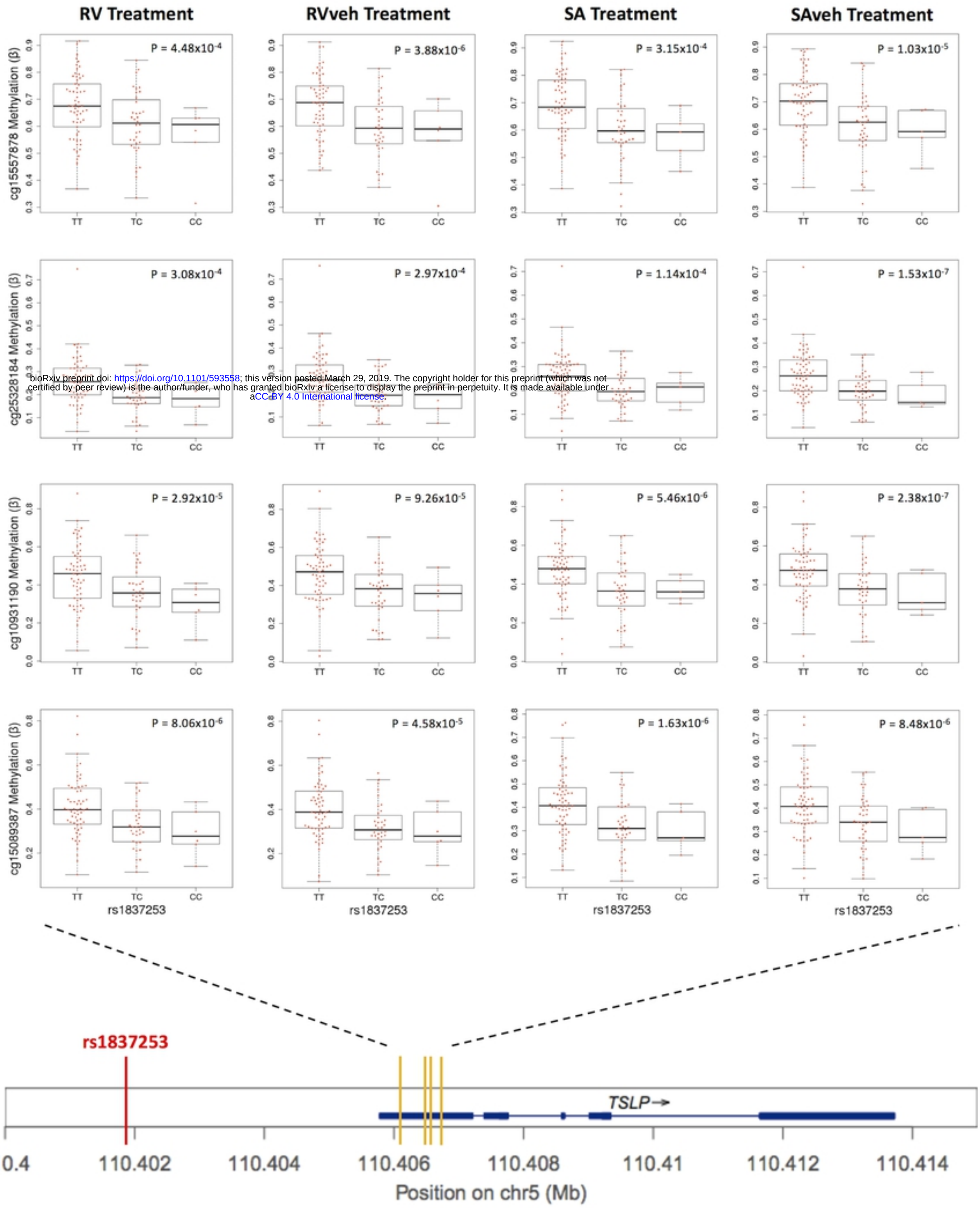


Fig 4

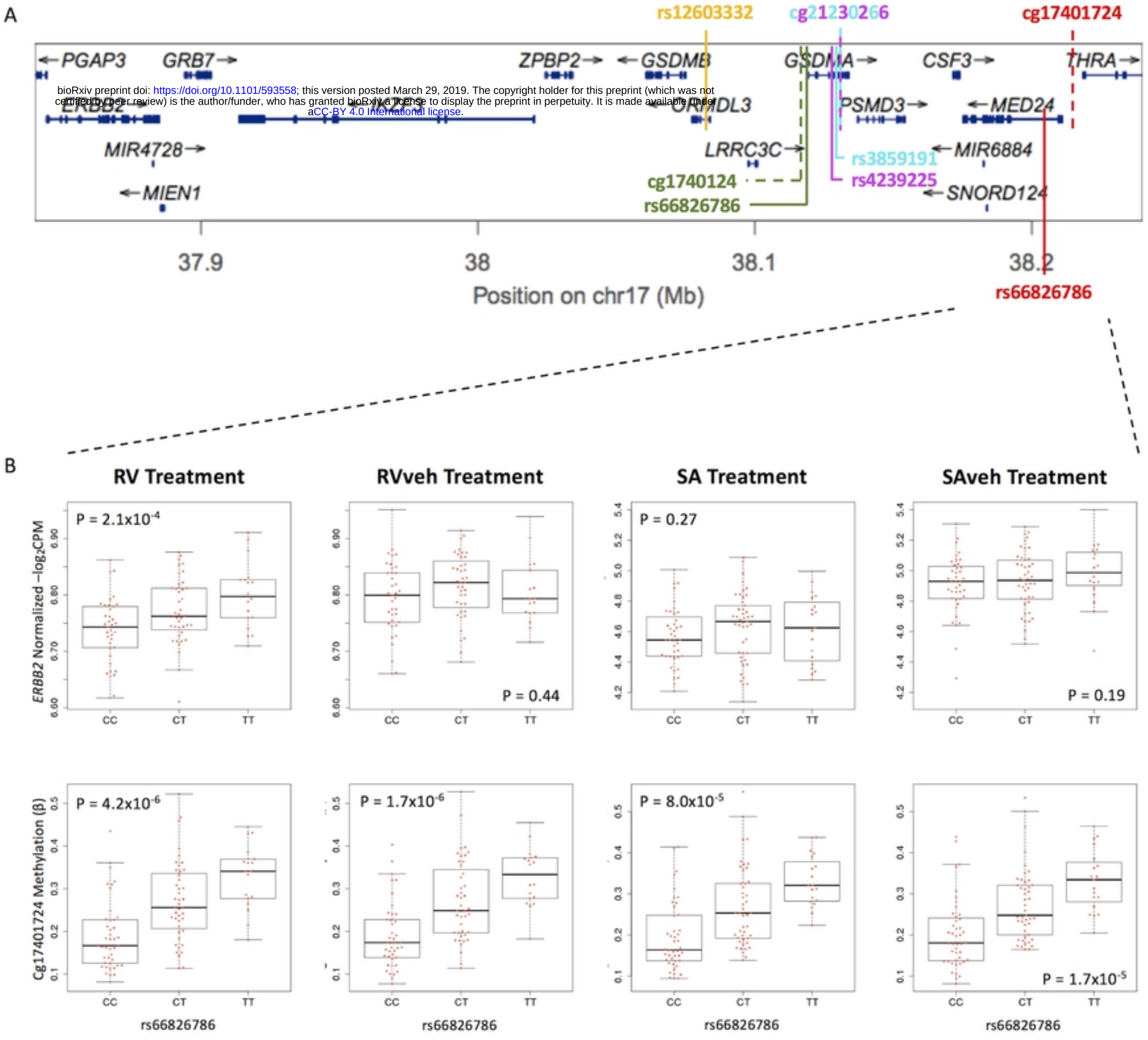


Fig 5