Multi-omics co-localization with genome-wide association studies reveals a context-specific genetic mechanism at a childhood onset asthma risk locus

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

Background

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 an upper airway (sinonasal) epithelial cell culture model to assess transcriptional and epigenetic responses to an asthma-promoting pathogen, rhinovirus (RV), and provide context-specific functional annotations to variants discovered in GWASs of asthma.

Methods

Using genome-wide genetic, gene expression and DNA methylation data in vehicle- and RV-treated airway epithelial cells (AECs) from 104 individuals, we mapped *cis* expression and methylation quantitative trait loci (*cis*-eQTLs and *cis*-meQTLs, respectively) in each condition. A Bayesian test for co-localization between AEC molecular QTLs and adult onset and childhood onset GWAS variants was used to assign function to variants associated with asthma. Mendelian randomization was applied to demonstrate DNA methylation effects on gene expression at asthma colocalized loci.

Results

Co-localization analyses of airway epithelial cell molecular QTLs with asthma GWAS variants revealed potential molecular disease mechanisms of asthma, including QTLs at the *TSLP* locus that were common to both exposure conditions and to both childhood and adult onset asthma, as

53 well as QTLs at the 17q12-21 asthma locus that were specific to RV exposure and childhood
54 onset asthma, consistent with clinical and epidemiological studies of these loci.

55 **Conclusion**

56 This study provides information on functional effects of asthma risk variants in airway epithelial
57 cells and insight into a disease-relevant viral exposure that modulates genetic effects on
58 transcriptional and epigenetic responses in cells and on risk for asthma in GWASs.

59

60

61

62

63

64

65

66

67

68

69

70

71 ## Background

72 Over the past decade, genome-wide association studies (GWASs) have identified over 60
73 asthma susceptibility loci at genome-wide levels of significance ($p<5x10^{-8}$) [1], with a locus at
74 17q12-21 being the most replicated and most significant asthma susceptibility locus in childhood
75 onset asthma (reviewed in [2]). Although asthma is typically diagnosed based on clinical
76 symptoms, such as wheeze, cough, and shortness of breath, it is actually comprised of many
77 overlapping phenotypes and distinct endotypes with shared as well as unique genetic and
78 environmental risk factors. For example, individuals with asthma differ with respect to age of
79 onset, environmental triggers of exacerbations, response to medications, obesity, and co-
80 occurrence with allergic diseases and other conditions. Recently, Pividori et al. reported 61
81 independent asthma loci, 23 of which were specific to childhood onset asthma, one that was
82 specific to adult onset asthma, and 37 that were associated with risk for both childhood onset and
83 adult onset asthma [1]. Gene and tissue enrichment patterns at these risk loci suggested that
84 epithelial cells (skin) and lung as primary etiological drivers of childhood onset and adult onset
85 asthma, respectively, while blood (immune) cell gene expression enrichments were shared by
86 both. However, GWASs do not generally consider tissue- or other environment-specific effects,
87 or gene by environment interactions. Moreover, epigenetic patterning may mediate the effects of
88 exposures on gene expression and disease risk, yet such studies have only rarely been integrated
89 with GWAS of asthma [3].

90 A challenge in interpreting GWAS results is that over 90% of disease-associated variants
91 are located in non-protein-coding regions of the genome [4], which are enriched for chromatin
92 signatures suggestive of enhancers [4] and for expression quantitative trait loci (eQTLs) [4-6].
93 SNPs associated in GWASs that also have functional annotations are more likely to be causal

94 variants, underlying disease pathophysiology through their effects on gene regulation. However,
95 identifying causal variants and their target genes at associated loci has been challenging, and the
96 functions of most SNPs associated with diseases in GWASs remain unknown. Databases such as
97 GTEx, ENCODE, and ROADMAP have been used to annotate GWAS SNPs and predict
98 molecular mechanisms through which risk variants affect disease phenotypes [3, 6-8]. But while
99 these resources have provided important insights into the interpretation of GWAS results, they
100 do not include all cell types relevant to all diseases or information on environmental exposures
101 that influence disease outcomes. As a result, annotations of asthma GWAS variants have been
102 largely limited to studies in transformed B cells lines, blood (immune) cells, and whole lung
103 tissue [1, 9, 10].

104 *In vitro* cell models provide an opportunity to address these limitations by characterizing
105 genetic and molecular responses to environmental exposures in cells from disease-relevant
106 tissues, and identifying genotypes that modify these responses [11, 12]. Joint analysis of datasets
107 (e.g. eQTLs and GWASs) can identify variants associated with both disease risk and molecular
108 traits as candidate causal variants that contribute to mechanisms of disease pathophysiology. A
109 multi-trait co-localization method (*moloc*) [13] was developed to integrate summary data from
110 GWAS and multiple molecular QTL datasets and identify candidate regulatory drivers of
111 complex phenotypes.

112 Here, we report the results of a multi-omics co-localization study to identify condition-
113 specific regulatory effects of asthma risk variants using an epithelial cell model of viral response.
114 Because airway epithelium forms a barrier to inhaled exposures, we used an *in vitro* upper
115 airway (sinonasal) epithelial cell model of transcriptional and epigenetic responses to rhinovirus
116 (RV). Primary infection of RV occurs in the nasal epithelium, and RV is a major contributor to

117 asthma inception in young children [14] and asthma exacerbations throughout life [15, 16],
118 underscoring its importance as a contextual promoter of asthma pathophysiology. We
119 demonstrate a specific enrichment of childhood onset asthma GWAS SNPs among airway
120 epithelial molecular QTLs, consistent with the important role that the epithelial barrier plays in
121 the inception of asthma in childhood [1, 17, 18]. Our integrative multi-omics approach suggests
122 an environment-specific mechanism of asthma pathogenesis at the 17q12-21 asthma locus in
123 childhood onset asthma, and a molecular mechanism shared between childhood onset and adult
124 onset GWASs in the *TSLP* gene at chromosome 5q22, highlighting complementary roles of the
125 airway epithelium in the pathogenesis of asthma.
126

127 ## Methods

128 **Ethics statement**

129 Study participants were recruited between March 2012 and August 2015, and nasal specimens
130 were collected as part of routine endoscopic sinonasal surgeries at Northwestern University
131 Feinberg School of Medicine. Informed written consent was obtained from each study
132 participant and randomly generated ID codes were assigned to all samples thereby preserving the
133 participant's anonymity and privacy. This study was approved by the institutional review boards
134 at Northwestern University Feinberg School of Medicine and the University of Chicago.
135

136 **Sample collection and composition**

137 Sinonasal epithelial cells were obtained by brushing the uncinate process collected at elective
138 surgery at Northwestern University from 56 males, 39 females, ages 18 – 73 years old (mean age
139 44), self-reported ethnicities as Caucasian (64%), Black (17%), Hispanic (13%), and "other"

140 (6%), and 43 asthmatics (current or prior status) and 61 non-asthmatics. Blood samples for
141 genotyping were collected from study participants. A summary of the study design is shown in
142 Fig. S1.
143
144 **Upper airway epithelial cell culture and RV treatment**
145 After isolation, nasal airway epithelial cells were cultured in bronchial epithelial cell growth
146 medium (Lonza, BEGM BulletKit, catalog number CC-3170) to near confluence, then frozen at -
147 80°C and stored in Liquid Nitrogen. Cells were subsequently thawed and cultured in collagen-
148 coated (PureCol, INAMED BioMaterials, catalog number 5,409, 3 mg/mL, 1:15 dilution) tissue
149 culture plates (6 wells of 2x 12 well plates) using BEGM overnight at 37°C and 5% CO_2. In
150 preparation for rhinovirus (HRV-16; RV) infection/stimulation, plates at 50-60% confluency
151 were incubated overnight in BEGM without hydrocortisone (HC) followed by a two-hour RV
152 infection at a multiplicity of infection (MOI) of 2 and vehicle treatment (Bronchial epithelial cell
153 basal medium (BEBM) + Gentamicin/Amphotericin) at 33°C (low speed rocking, ~15 RPM).
154 RV- and vehicle-treated cells were washed and then were cultured at 33°C for 46 hours (48 hours
155 total) in BEGM without HC.
156
157 **Genotyping and imputation**
158 DNA was extracted from whole blood or sinus tissue (if no blood was available) with the
159 Macherey-Nagel NucleoSpin Blood L or NucleoSpin Tissue L Extraction kits, respectively, and
160 quantified with the NanoDrop ND1000. Genotyping of all study participants was performed
161 using the Illumina Infinium HumanCore Exome+Custom Array (550,224 SNPs). After quality
162 control (QC) (excluding SNPs with HWE < 0.0001 by race/ethnicity, call rate < 0.95, MAF <

163 0.05 and individuals with genotype call rates < 0.05), 529,993 markers for 104 individuals were
164 available for analysis. Ancestry principal component analysis (PCA) was performed using 676
165 ancestry informative markers included on the array that overlap with the HapMap release 3 (Fig.
166 S2).

167 Phasing and imputation were performed using the ShapeIt2 [19] and Impute2 [20]
168 software packages, respectively. Variants were imputed in 5 Mb windows across the genome
169 against the 1000 Genomes Phase 3 haplotypes (Build 37; October 2014). Individuals were
170 categorized into two groups based on the k-means clustering of ancestry PCs, using the kmeans()
171 function in R; individuals were grouped as European or African American based on how they
172 related to the HapMap reference panel and means clustering of their ancestry PCs (Fig. S2).
173 After imputation, both groups were merged and QC was performed with gtool [21]. X and Y
174 chromosome-linked SNPs and SNPs that did not meet the QC criteria (info score < 0.8, MAF <
175 0.05, missingness > 0.05 and a probability score < 0.9) were excluded from analyses. Probability
176 scores were converted to dosages for 6,665,552 of the remaining sites used in downstream
177 analyses.

178

179 **RNA extraction and sequencing**

180 Following RV and vehicle treatments, RNA from cells underwent extraction and purification
181 using the QIAGEN AllPrep DNA/RNA Kit. RNA quality and quantity were measured at the
182 University of Chicago Functional Genomics Core using the Agilent RNA 6000 Pico assay and
183 the Agilent 2100 Bioanalyzer. RNA integrity numbers (RIN) were greater than 7.7 for all
184 samples. cDNA libraries were constructed using the Illumina TruSeq RNA Library Prep Kit v2
185 and sequenced on the Illumina HiSeq 2500 System (50 bp, single-end); RNA sequencing was

186 completed at the University of Chicago Genomics Core. Subsequently, we checked for potential
187 sample contamination and sample swaps using the publicly available software VerifyBamID
188 (http://genome.sph.umich.edu/wiki/VerifyBamID) [22] for cells from all 104 individuals
189 included in each treatment condition. We did not detect any cross-contamination between
190 samples but we did identify one sample swap between individuals, which we subsequently
191 corrected.

192 Sequences were mapped to the human reference genome (hg19) and reads per gene were
193 quantified using the Spliced Transcripts Alignment to a Reference (STAR) [23] software. X,Y,
194 and mitochondrial chromosome genes, and low count data (genes < 1CPM) were removed prior
195 to normalization via the trimmed mean of M-values method (TMM) and variance modeling
196 (voom) [24]; samples contain > 8M mapped reads. Principle components analysis (PCA)
197 identified biological and technical sources of variation in the voom-normalized RNA-seq reads.
198 We identified contributors to batch and other technical effects (days in liquid nitrogen,
199 experimental culture days, cell culture batches, RNA concentration, RNA fragment length,
200 technician, sequencing pool). Additionally, unknown sources of variation were predicted with
201 the Surrogate Variable Analysis (SVA) [25] package in R where 15 surrogate variables (SVs)
202 were estimated for the samples that were included in the experiment. Voom-normalized RNA-
203 seq data were then adjusted for technical effects, SVs, sex, and ancestry PCs (1-3) using the
204 function removeBatchEffect() from the R package limma [26]. Treatment responses in epithelial
205 cells were detected in the combined sample with 6,650 differentially expressed genes identified
206 at a FDR≤0.01 (Fig. S3).

207

208 **DNA extraction and methylation profiling**

209 Following RV and vehicle treatments, DNA was extracted from cells as described above. DNA
210 methylation profiles for cells from each treatment were measured on the Illumina Infinium
211 MethylationEPIC BeadChip at the University of Chicago Functional Genomics Core.
212 Methylation data were preprocessed using the minfi package [27]. Probes located on sex
213 chromosomes and with detection p-values greater than 0.01 in more than 10% of samples were
214 removed from the analysis; samples with more than 5% missing probes were also removed. A
215 preprocessing control normalization function was applied to correct for raw probe values or
216 background and a Subset-quantile Within Array Normalization (SWAN) [28] was used to correct
217 for technical differences between the Infinium type I and type II probes. Additionally, we
218 removed cross-reactive probes and probes within two nucleotides of a SNP with an MAF greater
219 than 0.05 using the function rmSNPandCH() from the R package DMRcate [29].
220 PCA identified technical and biological sources of variation in the normalized DNA
221 methylation datasets. We identified contributors to batch and technical effects including array,
222 and cell harvest date. Sex, age, and smoking were significant variables in the PCA. Unknown
223 sources of variation were predicted with the SVA package where we estimated 37 SVs. SWAN
224 and quantile-normalized M-values were then adjusted for batch and technical effects, SVs, sex,
225 age, and smoking using the function removeBatchEffect() in R. Treatment effects were detected
226 in the combined sample with 1,710 differentially methylated CpGs at a FDR<0.10 (Fig. S4).
227
228 **eQTL and meQTL analyses**
229 Prior to e/meQTL analysis, voom-transformed gene expression values and normalized
230 methylation M-values were adjusted for technical (array, cell harvest date), biological variables
231 (sex, age, ancestry PCs), as well as smoking and surrogate variables as described above. Linear

232 regression between the permuted genotypes (MAF>0.05) and molecular phenotypes (gene
233 expression and methylation residuals) from each treatment condition was performed with the
234 FastQTL [30] software package within *cis*-window sizes of 1 Mb and 10 kb for eQTL and
235 meQTL analyses, respectively. Nominal passes were conducted for each eQTL and meQTL
236 analysis within FastQTL, and an FDR threshold of 0.10 was applied to adjust for multiple testing
237 within each experimental dataset with the p.adjust() function in R.
238 A conditional analysis was performed with the QTLtools [31] software package to
239 identify molecular QTLs with independent effects on gene expression and DNA methylation.
240 This was accomplished in two-steps. First, a permutation analysis was performed within a *cis*-
241 window sizes of 1 Mb and 10 kb for eQTL and meQTL analyses, respectively, to derive nominal
242 p-value thresholds per molecular phenotype. Second, a forward-backward stepwise regression is
243 applied to ultimately assign significant variants to independent signals.
244
245 **Multivariate adaptive shrinkage analysis (mash)**
246 An Empirical Bayes method of multivariate adaptive shrinkage was applied separately to the
247 eQTL and meQTL data sets as implemented in the R statistical package, mashr
248 (https://github.com/stephenslab/mashr) [32], to produce improved estimates of QTL effects and
249 corresponding significance values in each treatment condition. Mashr implements this in two
250 general steps: 1) identification of pattern sharing, sparsity, and correlation among QTL effects,
251 and 2) integration of these learned patterns to produce improved effects estimates and measures
252 of significance for eQTLs or meQTLs in each treatment condition. To fit the mash model, we
253 first estimated the correlation structure in the null test from a random dataset in which 235,851
254 and 3,959,482 phenotype-SNP pairs were chosen for eQTLs and meQTLs, respectively, from the

255 FastQTL nominal pass; because mashr is computationally intensive, the number of randomly
256 chosen gene/CpG-SNP pairs were determined based on R's memory capabilities. The data-
257 driven covariances were then estimated using the 'top' mQTL in each gene or CpG results from
258 FastQTL. Posterior summaries were then computed for the 'top' eQTL and meQTL results (see
259 [32]). The instructions found in the *mashr* eQTL analysis outline vignette were followed to run
260 mash.
261

262 **Enrichment analysis**

263 The R package, GWAS analysis of regulatory or functional information enrichment with LD
264 correction (GARFIELD) [33], was used to quantify enrichment and assess significance of
265 GWAS SNPs among eQTLs and meQTLs. GARFIELD leverages GWAS results with molecular
266 data to identify features relevant to a phenotype of interest, while accounting for LD and
267 matching for genotyped variants, by applying a logistic regression method to derive statistical
268 significance for enrichment. For this study, molecular QTLs were tested for GWAS variant
269 enrichment, estimated as odds ratios and enrichment P-values derived at four GWAS P-value
270 thresholds: 10^{-5}, 10^{-6}, 10^{-7}, and 10^{-8}. To demonstrate disease-specificity of our results, we
271 selected summary statistics from four GWASs performed in UK Biobank subjects (Alzheimer's
272 disease [34], atrial fibrillation [35], height [36], neuroticism [37], in addition to one each for
273 adult onset and childhood onset asthma [1]). Summary statistics from these six GWASs were
274 used for enrichment analyses of the 755,441 molecular QTLs combined from each treatment
275 condition. These non-asthma GWASs were chosen based on similar population backgrounds
276 (European), availability of summary statistics (as of 05/18), and not known or expected to have

277 overlapping genetics with asthma (i.e., excluding diseases with known allergic or autoimmune
278 etiologies).
279 To assess tissue-specificity of our results, we examined eQTLs from the adrenal gland,
280 frontal cortex, hypothalamus, ovary, and testis from the GTEx database version 7
281 (http://gtexportal.org) [6], and tested for enrichment of adult onset and childhood onset asthma
282 GWAS SNPs among the epithelial eQTLs from our study combined across treatment conditions.
283 GTEx data were matched with respect to sample size and number of eQTLs with those of the
284 epithelium, with the exception of testis, which was included to show the consistency of the
285 enrichment results despite it being an outlier in regards to both sample size, which was smaller,
286 and number of eQTLs, which was larger. An FDR threshold of 5% and 10% was applied to
287 eQTLs from GTEx and from our study, respectively, for a balanced, unbiased assessment of
288 enrichment. An OR > 1 and a Benjamini-Hochberg (BH) corrected p-value threshold of < 0.05
289 was used as the significance threshold for enrichment; BH adjusted p-values were calculated
290 using the p.adjust() function in R where 'n' was determined by the number of tests in each
291 respective enrichment analysis.
292
293 **Co-localization analysis**
294 To estimate the posterior probability association (PPA) that a SNP contributed to the association
295 signal in the GWAS as well as to the eQTL and/or meQTL, we applied a Bayesian statistical
296 framework implemented in the R package multiple-trait-coloc (moloc) [13]. Summary data from
297 adult onset and childhood onset asthma GWASs from [1], along with eQTL and meQTL
298 summary data from cells within each treatment condition (described above), were included in the
299 *moloc* analysis. Each co-localization analysis included summary data from a GWAS and

300 epithelial cell eQTLs and meQTLs from corresponding treatment conditions. Because a genome-
301 wide co-localization analysis was computationally untenable, genomic regions for co-
302 localization were defined using GARFIELD. First, we analyzed the enrichment pattern of
303 e/meSNPs from each treatment condition in adult onset and childhood onset GWASs using the
304 default package settings. Second, we extracted variants driving the enrichment signals at a
305 GWAS p-value threshold of $1x10^{-4}$. Regions were defined as 2 Mb windows centered around
306 these variants. Only regions with at least 10 SNPs in common between all three datasets or
307 'traits' (GWAS, eQTL, and meQTL) were assessed by moloc and 15 'configurations' of possible
308 variant sharing was computed across these three traits (see [13] for more details). PPAs ≥ 0.70
309 were considered as evidence for co-localization. Prior probabilities of $1x10^{-4}$, $1x10^{-6}$, and $1x10^{-7}$
310 were chosen for the association of one, two, or three traits, respectively, as recommended by the
311 authors of moloc.
312
313 **Mendelian randomization**
314 Mendelian Randomization was performed using the ivreg2 function in R (https://www.r-
315 bloggers.com/an-ivreg2-function-for-r/) which applies a 2-stage least squares regression (2SLS),
316 as implemented in [38]. We used the only co-localized triplet (eQTL-meQTL-GWAS) to assess
317 the causal effects of DNA methylation (cg17401724) on gene expression (*ERBB2*), using the
318 genotype at rs66826786 as the instrumental variable.
319
320 **Results**
321 **Genome-wide *cis*-eQTLs and *cis*-meQTLs mapping in cultured airway epithelial cells**

322 To identify genetic variation influencing gene expression under different conditions, we
323 performed eQTL mapping in cultured AECs exposed to RV, and its corresponding vehicle
324 control from 104 individuals (43 with doctor diagnosed asthma; 61 without a doctor's diagnosis
325 of asthma; Fig. S1). Analyses were performed separately for each treatment condition, testing for
326 associations with 6,665,552 imputed SNPs (MAF>0.05) and 11,231 autosomal genes (see
327 Methods; Additional file 1 and 2). The numbers of SNPs associated with gene expression for at
328 least one gene (eQTLs) and genes with at least one eQTL (eGenes), in any treatment, are
329 summarized in Fig. S5.
330 In parallel, we performed meQTL mapping in the same cells used for gene expression
331 studies. We performed this analysis separately for each treatment condition, testing for
332 associations with the same imputed SNP set that was used for eQTL mapping and interrogated
333 791,765 autosomal CpGs (Additional file 3 and 4). A summary of the number of SNPs
334 associated with methylation levels at one or more CpG sites (meQTLs) and CpG sites with at
335 least one meQTL (meCpGs), in any treatment, are shown in Fig. S5.
336 Each gene/CpG-variant pair was tested for a linear regression slope that significantly
337 deviated from 0. Therefore, the estimated effects for the molecular QTLs reflects both the single-
338 SNP effects of each molecular QTL as well as those that are in linkage disequilibrium (LD).
339 Accordingly, these analyses do not differentiate between causal molecular QTLs from those in
340 LD with the QTL. However, these variants are still informative in prioritizing genes and CpG
341 sites that contribute to the etiology of asthma.
342
343 **Estimating shared and condition-specific molecular QTL effects**

344 After identifying molecular QTLs in each treatment condition, we first explored the impact of
345 RV exposure on eQTLs and meQTLs by comparing RV-treated to vehicle-treated results. For
346 this analysis, we used an empirical Bayes method, multivariate adaptive shrinkage (*mash*; see the
347 "Methods" section) [32]. Compared to direct comparisons between conditions, *mash* increases
348 power, improves effect-size estimates, and provides better quantitative assessments of effect size
349 heterogeneity of molecular QTLs, thereby allowing for greater confidence in effect sharing and
350 estimates of condition-specificity. Additionally, as a confidence measurement of the direction of
351 QTL effects, *mash* provides a 'local false sign rate' (lfsr) that is the probability that the estimated
352 effect has the incorrect sign [39], rather than the expected proportion of Type I errors as would
353 be assessed using FDR thresholds.

354 To identify condition-specific eQTLs, we analyzed the effect estimates of the most
355 significant eQTL for each of 11,231 genes and assessed sharing of these signals among the RV
356 and vehicle treated cells (see Methods). A pairwise comparison showed that 58.3% of eQTLs
357 were shared between RV and vehicle treatments, representing 1,564 eGenes, defined here as
358 genes with at least one eQTL at a lfsr < 0.05 (Fig. 1A; Additional file 5). We observed 660 and
359 458 condition-specific eGenes in the vehicle- and RV-treated cells, respectively. These
360 potentially represent genetic variants that modify responses to viral exposure in AECs. Examples
361 of treatment-specific eQTLs are shown in Fig. 1B. The effect estimates of the most significant
362 meQTL for each of 751,914 CpG sites were used to identify condition-specific DNA methylation
363 effects, as described above for eQTLs. A pair-wise analysis of meQTLs revealed that 89.9% of
364 meQTLs were shared between vehicle and RV treatments, representing 48,189 meCpGs, defined
365 here as CpGs with at least one meQTL at a lfsr<0.05 (Fig. 1C; Additional file 6), revealing a

366 much greater proportion shared meQTLs than those observed for eQTLs. Examples of the 5,416
367 treatment-specific meQTLs are shown in Fig. 1D.
368 In total, we identified 660 and 458 eGenes (lfsr<0.05) that were specific to vehicle or RV
369 treatment, respectively, and 5,162 and 254 meCpGs that were specific to vehicle or RV culture
370 treatment, respectively, with greater confidence than by pairwise comparisons using FDR
371 thresholds [32].
372
373 **Molecular QTLs in the airway epithelium are enriched for asthma GWAS SNPs**
374 Although the majority of variation in the human genome is non-functional [40], GWAS loci tend
375 to be most enriched for functional annotations in disease-relevant cells [4, 41, 42]. To assess
376 whether the 755,441 molecular QTLs identified in our study (i.e., the union of eQTLs and
377 meQTLs from each treatment condition at FDR<0.10) are enriched for GWAS variants and
378 whether these enrichments show tissue specificity, we first extracted summary statistics from a
379 publicly available GWAS data for childhood onset and adult onset asthma [1] and for four
380 diseases without known allergic or autoimmune etiologies (Alzheimer's disease [34], atrial
381 fibrillation [35], height [36], and neuroticism [37]). There were statistically significant
382 enrichments (OR>1 and BH-adjusted P-value<0.05; see Methods) for the childhood and adult
383 onset asthma GWAS SNPs among the molecular QTLs at each of four GWAS thresholds (Table
384 1), consistent with the strong epithelial cell involvement in asthma in general and with childhood
385 onset asthma in particular. In contrast, there were no significant enrichments for SNPs from four
386 of the other GWASs among the epithelial cell molecular QTLs. These results highlight the
387 specific enrichment of asthma GWAS SNPs among airway epithelial molecular QTLs compared
388 to SNPs from GWASs of diseases without known epithelial cells involvement.
389

Table 1. Enrichment estimates of airway epithelial cell molecular QTLs for GWAS SNPs. *P*-values that are significant after BH correction (see Methods) are shown in bolded type.

GWAS	N_{Cases}	$N_{Controls}$	N_{Total}	GWAS Threshold	OR	P
Alzheimer's Disease [34]	47,793	328,320	376,311	$1x10^{-5}$	0.78	$3.95x10^{-1}$
				$1x10^{-6}$	0.81	$5.47x10^{-1}$
				$1x10^{-7}$	0.80	$5.74x10^{-1}$
				$1x10^{-8}$	0.53	$2.03x10^{-1}$
Atrial Fibrillation [35]	60,620	970,216	1,030,836	$1x10^{-5}$	1.14	$4.73x10^{-1}$
				$1x10^{-6}$	1.07	$1.71x10^{-1}$
				$1x10^{-7}$	1.12	$6.89x10^{-1}$
				$1x10^{-8}$	1.48	$2.64x10^{-1}$
Height [36]	NA	NA	253,288	$1x10^{-5}$	1.23	$2.49x10^{-1}$
				$1x10^{-6}$	1.38	$1.43x10^{-1}$
				$1x10^{-7}$	1.50	$1.22x10^{-1}$
				$1x10^{-8}$	1.51	$1.56x10^{-1}$
Neuroticism [37]	130,664	330,470	461,134	$1x10^{-5}$	1.35	$1.10x10^{-1}$
				$1x10^{-6}$	1.72	$3.84x10^{-2}$
				$1x10^{-7}$	3.52	$3.28x10^{-3}$
				$1x10^{-8}$	9.09	$2.23x10^{-3}$
Adult Onset Asthma [1]	**21,564**	**318,237**	**339,801**	**$1x10^{-5}$**	**5.09**	**$7.07x10^{-5}$**
				$1x10^{-6}$	**5.36**	**$1.86x10^{-3}$**
				$1x10^{-7}$	8.11	$2.18x10^{-3}$
				$1x10^{-8}$	**19.3**	**$1.16x10^{-3}$**
Childhood Onset Asthma [1]	**9,433**	**318,237**	**327,670**	**$1x10^{-5}$**	**2.92**	**$2.97x10^{-6}$**
				$1x10^{-6}$	**4.77**	**$1.23x10^{-7}$**
				$1x10^{-7}$	**5.42**	**$9.44x10^{-7}$**
				$1x10^{-8}$	**4.05**	**$1.64x10^{-4}$**

To further assess the specificity of airway epithelial molecular QTLs to asthma, we compared GWAS SNP enrichments among the eQTLs in our study to those from tissues that are not known to be involved in asthma. To this end, we tested for enrichment of asthma GWAS SNPs among eQTLs (FDR<0.05) in five different tissues from the GTEx database (adrenal, frontal cortex, hypothalamus, ovary, testis) [43], and compared them to enrichments among the eQTLs from our study. We observed a significant enrichment (OR>1 and BH-adjusted P<0.05)

397 of childhood onset asthma GWAS SNPs among the epithelial cell eQTLs at all GWAS P-value
398 thresholds $\leq 1x10^{-7}$ (Table 2), while enrichments for adult onset asthma GWAS SNPs among the
399 epithelial cell eQTLs were not observed at any GWAS threshold (Table S1). Except for the
400 hypothalamus, which showed some enrichment at $P < 10^{-5}$), no other enrichments of asthma
401 GWAS SNPs were observed among eQTLs in other tissues, further supporting the specificity of
402 our model and previous studies suggesting that epithelial barrier defects underlie risk for
403 childhood onset, but not adult onset, asthma [1, 17, 18].
404

Table 2. Enrichment estimates of eQTLs for childhood onset asthma GWAS SNPs from six tissues. Significant *P*-values after BH correction are shown in bolded type. Results for adult onset asthma is shown in Table S1.

Tissue	GWAS Threshold	OR	P	N	N_{eSNP}
Adrenal	$1x10^{-5}$	1.32	$1.84x10^{-1}$	175	588,348
	$1x10^{-6}$	1.02	$9.51x10^{-1}$		
	$1x10^{-7}$	0.74	$3.91x10^{-1}$		
	$1x10^{-8}$	1.05	$8.97x10^{-1}$		
Brain - Frontal Cortex	$1x10^{-5}$	1.72	$2.02x10^{-2}$	118	367,312
	$1x10^{-6}$	1.33	$3.46x10^{-1}$		
	$1x10^{-7}$	1.19	$6.27x10^{-1}$		
	$1x10^{-8}$	1.46	$3.29x10^{-1}$		
Brain - Hypothalamus	**$1x10^{-5}$**	**2.27**	**$9.38x10^{-4}$**	**108**	**251,506**
	$1x10^{-6}$	1.72	$9.68x10^{-2}$		
	$1x10^{-7}$	1.64	$1.93x10^{-1}$		
	$1x10^{-8}$	1.64	$2.55x10^{-1}$		
Ovary	$1x10^{-5}$	1.78	$2.05x10^{-2}$	122	292,461
	$1x10^{-6}$	1.51	$1.88x10^{-1}$		
	$1x10^{-7}$	1.11	$8.02x10^{-1}$		
	$1x10^{-8}$	1.06	$8.96x10^{-1}$		
Testis	$1x10^{-5}$	0.85	$3.90x10^{-1}$	225	1,358,512
	$1x10^{-6}$	0.74	$1.82x10^{-1}$		
	$1x10^{-7}$	0.83	$4.52x10^{-1}$		
	$1x10^{-8}$	0.81	$4.62x10^{-1}$		
Airway Epithelial Cells	**$1x10^{-5}$**	**3.01**	**$2.05x10^{-4}$**	**104**	**185,407**
	$1x10^{-6}$	**2.89**	**$3.89x10^{-3}$**		

	$1x10^{-7}$	**3.05**	**$7.68x10^{-3}$**
	$1x10^{-8}$	2.41	$1.01x10^{-1}$

405

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

407 Integrating molecular QTLs with GWAS data is a powerful way to identify functional variants
408 that may ultimately influence disease risk [44, 45] and to assign function to known disease-
409 associated variants. Co-localization approaches directly test whether the same genetic variant is
410 underlying associations between two or more traits (e.g., gene expression and asthma), providing
411 clues to causal disease pathways. We hypothesized that integrating molecular QTLs from RV-
412 and vehicle-exposed epithelial cells with results of GWASs for adult onset and childhood onset
413 asthma would reveal genetic and epigenetic mechanisms that modulate risk for childhood and/or
414 adult onset asthma.

415 To test this hypothesis, we extracted summary statistics from large GWASs of adult onset
416 asthma and childhood onset asthma [1], and tested each for co-localization with genetic variants
417 associated with gene expression, DNA methylation, and asthma, using *moloc*, a Bayesian
418 statistical approach that allows integration and co-localization of more than two molecular traits
419 [13]. We performed four separate co-localization tests for each treatment conditions with each of
420 the GWASs. Each analysis provided three possible configurations in which a variant is co-
421 localized between the GWAS and QTLs: eQTL-GWAS pairs, meQTL-GWAS pairs, eQTL-
422 meQTL-GWAS triplets. Estimates of a posterior probability of association (PPA) is provided,
423 reflecting the evidence for a colocalized SNP being causal for the associations in the GWAS and
424 for the corresponding eQTL and/or meQTL.

425 Using this approach, we found evidence for a total of 19 unique multiple trait co-
426 localizations (Table 3). A single meQTL-GWAS pair was co-localized in both the adult onset

427 and childhood onset asthma GWASs. An additional 18 co-localizations were detected only in the
428 childhood onset asthma GWAS, including a single eQTL-meQTL-GWAS triplet associated with
429 the *ERBB2* gene, three eQTL-GWAS pairs associated with three genes (*FLG, FLG-AS1,*
430 *ORMDL3*), and 15 meQTL-GWAS pairs associated with 11 CpG sites (Table 3; Table S2). No
431 co-localizations were specific to the adult onset asthma GWAS. Among the co-localized eGenes,
432 based on previous studies, *FLG* was predicted to have decreased expression of in childhood
433 onset asthma [46], *ERBB2* was predicted to have decreased expression of in severe asthma [47],
434 and *ORMDL3* was predicted to have increased expression (Table S3) [48]. The larger number of
435 co-localizations for childhood onset asthma relative to adult onset asthma is consistent both with
436 the previous observation that genes at the childhood onset asthma loci were most highly
437 expressed in skin, an epithelial cell type [1] and with the enrichment of childhood onset asthma
438 GWAS SNPs among epithelial cell eQTLs described above.
439

Table 3. Number of QTL-GWAS pairs or triplets with evidence of co-localization (PPA≥0.70). The one meQTL-GWAS pair in the adult onset asthma GWAS is also among the 15 meQTL-GWAS pairs in the childhood onset GWAS.

GWAS	eQTL-meQTL-GWAS	eQTL-GWAS	meQTL-GWAS
Adult onset asthma	0	0	1
Childhood onset asthma	1	3	15

PPA >= 0.70

440

441 The significance threshold ($p<5x10^{-8}$) required to control the false discovery rate in
442 GWASs likely excludes many true associations that do not reach this stringent cutoff. We and
443 others have suggested that these SNPs, i.e., the mid hanging fruit [49], may be environment- or
444 context-specific associations that are missed in GWASs that typically do not control for either
445 [50, 51]. Notably, 10 of the 19 SNPs associated with co-localizations in the childhood onset

446 asthma GWAS did not reach genome-wide significance in the GWAS. These results therefore
447 provided functional inferences both for variants that were significant in a GWAS at known
448 asthma loci and for variants that did not meet strict criteria for significance in the GWAS,
449 thereby facilitating prioritization of variants among the mid-hanging fruit [49]. Two examples of
450 co-localizations with prominent asthma-associated loci are described in the following sections.
451

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

453 To more deeply characterize the co-localizations, we first focused on the only meQTL-GWAS
454 pair in both the adult onset and childhood onset asthma GWASs. This pair included an intergenic
455 SNP (rs1837253) located 5.7 kb upstream from the transcriptional start site (TSS) of the *TSLP*
456 gene on chromosome 5q22, encoding an epithelial cell cytokine that plays a key role in the
457 inflammatory response in asthma and other allergic diseases [52]. rs1837253 co-localized with a
458 single meQTL (cg15557878) in both the adult onset ($p_{GWAS} = 2.77x10^{-13}$) and childhood onset
459 ($p_{GWAS} = 2.33x10^{-27}$) asthma GWASs. The meCpG is located in the first (untranslated) exon (5'
460 UTR) of the *TSLP* gene (Fig. 2), a region characterized as a promoter in normal human
461 epidermal keratinocyte cells (NHEK; ROADMAP). In fact, rs1837253 was the sentinel SNP at
462 this locus in GWASs of asthma (e.g. [1, 53]) and of moderate-to-severe asthma [54]. In our
463 study, the rs1837253-C asthma risk allele was associated with hypermethylation in primary
464 cultured AECs at cg15557878 (Fig. 2), but was not associated with the expression of *TSLP* in
465 either treatment condition (not shown).

466 Previous studies have shown *TSLP* to be a methylation-sensitive gene and that
467 hypomethylation at its promoter is associated with atopic dermatitis (AD) and prenatal tobacco
468 smoke exposure [55, 56]. Another study showed that the rs1837253-CC genotype was associated

469 with increased excretion of TSLP in cultured AECs after exposure to polyI:C (a dsRNA
470 surrogate of viral stimulation) [57]. Neither finding could be addressed in our study. Moreover,
471 we were unable to identify any SNPs in LD with rs1837253 (± 50 kb) in either European or
472 African American ($r^2 < 0.12$) 1000Genomes reference panels, implying that this SNP may
473 indeed be the causal SNP at this locus. Our results further suggest that DNA methylation levels
474 in AECs may underlie this effect.
475
476 **Multi-trait co-localizations of molecular QTLs and asthma risk at the 17q12-21 asthma**
477 **locus**
478 To further explore the possibility that some mechanisms of asthma risk are exposure-specific, we
479 focused on the co-localizations of eQTLs and meQTLs with asthma-associated SNPs at the
480 17q12-21 (17q) locus, the most replicated locus for childhood onset asthma (reviewed in [2]).
481 This locus is characterized by high LD across a core region of 150 kb, encoding at least 4 genes
482 (including *ORMDL3* and *GSDMB*). SNPs extending both proximal (including *PGAP3* and
483 *ERBB2*) and distal (including *GSDMA*) to the core region show less LD with those in the core
484 region and have been implicated as potentially independent asthma risk loci. Previous studies
485 have shown that SNPs at this extended locus are eQTLs for at least four genes (*ORMDL3*,
486 *GSDMB*, *GSDMA*, *PGAP3*) in blood and/or lung cells [2] and that genetic variants at this
487 childhood onset asthma locus are also strongly associated with early life wheezing illness [58,
488 59], particularly RV-associated wheezing illness [60].
489 We identified six co-localizations at the extended 17q locus of molecular QTLs that were
490 specific to childhood onset asthma GWAS SNPs. Among these co-localizations, one eQTL-
491 GWAS pair with rs12603332 and expression of *ORMDL3* was only in vehicle-treated cells

492 (PPA≥0.70; Fig. 3A-B). The co-localized SNP (rs12603332) is in LD (r^2>0.74 in 1000 Genomes
493 European reference panel) with other previously reported asthma-associated GWAS SNPs in this
494 region, including some that were reported as eQTLs for *ORMDL3* and *GSDMB*, primarily in
495 blood immune cells. However, in contrast to studies in *ex vivo* upper AECs [61], none of the
496 SNPs were eQTLs for *GSDMB* in our *in vitro* culture model. That the co-localization with
497 rs12603332 and *ORMDL3* expression was only significant in vehicle treated cells reflects the
498 blunting of the eQTL effects (Fig. 3B), and possibly the overall decreased expression of
499 *ORMDL3* (Fig. 3C), in RV-treated cells.

500 We also detected three meQTL-GWAS pairs among the six co-localizations at the 17q
501 locus that were associated with two meCpGs (cg21230266, cg17401724) and three SNPs at the
502 distal end of (rs4239225, rs3859191) and beyond (rs66826786) the extended locus near *GSDMA*,
503 where there is some reduction of LD with SNPs in the core region (Fig. 3D-F). One of these
504 CpGs was located in an intron (cg21230266) of *GSDMA* in regions characterized by ROADMAP
505 as enhancers in NHEK cells. SNPs in modest to perfect LD (r^2_{range}=0.46 - 1.00; 1000 Genomes
506 European panel) with these co-localizations (rs4239225, rs3859191) were described in previous
507 studies as an independent GWAS signal for asthma (rs3894194) or an eQTL for *GSDMA*
508 (rs3859192) [6, 62, 63]. These three meQTL-GWAS co-localizations were detected only in the
509 RV-treated cells, although the meQTL signal for each of the three co-localizations was also
510 detected in the vehicle treatment, likely due to decreased power to co-localize these meQTLs
511 from the vehicle-treated cells. Additionally, there were no statistically significant differences in
512 DNA methylation levels observed between the vehicle and RV treatments (Fig. 3F).

513 The one eQTL-meQTL-GWAS triplet detected in our study at the 17q locus (Fig. 4A,
514 upper panel. The co-localization included an eQTL for *ERBB2*, at the proximal end of the locus

515 and more than 361 kb from the co-localized asthma risk variant in an intron of *MED24*
516 (rs66826786) and the co-localized meCpG (cg17401724) at the distal end of the locus (Fig. 3A,
517 middle panel), . *MED24* is beyond the extended 17q12-21 locus as previously defined [2] in a
518 region characterized by ROADMAP as both an enhancer and TSS in NHEKs. The eQTL for
519 *ERBB2* is observed only after exposure to RV (Fig. 3A middle and lower panels), though the
520 meQTL associated with this triplet was present in both vehicle and RV treatment conditions (Fig.
521 3B upper and lower panels, respectively). The asthma risk allele, rs66826786-T, was associated
522 with decreased DNA methylation of cg17401724 in both conditions but with decreased *ERBB2*
523 expression only in RV-treated cells. Overall, *ERBB2* expression decreased in response to RV
524 exposure in AECs (Fig. 3D). The 361 kb distance between the promoter of *ERBB2* and its eSNP
525 (rs66826786) suggests long-range interaction between *ERBB2* and the region harboring
526 cg17401724 and rs66826786. The fact that the eQTL is observed only after RV infection, further
527 suggests either that infection with RV triggers this long-range interaction in AECs via chromatin
528 looping between these loci, or that RV infection results in the recruitment of negative
529 transcription factors in this region that is already epigenetically poised. In fact, the observation
530 that the meQTL for cg17401724 is observed in both conditions indeed suggests an epigenetically
531 poised chromatin state at the distal end that directly affects transcription of *ERBB2* at the
532 proximal end of the locus after exposure to RV, and possibly to other viruses.

533

534 **Mendelian randomization of multi-trait co-localized triplets**

535 Co-localization analyses reveal genetic variants that are associated with both asthma and
536 molecular traits (gene expression and/or DNA methylation) but the question of causality between
537 the molecular traits remains unanswered. To infer causal relationships between DNA

538 methylation and gene expression on asthma risk, we performed Mendelian randomization (MR),
539 a method in which genetic variation associated with modifiable exposure patterns (i.e. DNA
540 methylation) can be used as an instrumental variable to estimate the causal influence of an
541 exposure on an outcome (i.e. DNA methylation on gene expression) [64]. Specifically, we
542 applied a two-stage least squares regression (2SLS; see Methods) regression to estimate the
543 effects of DNA methylation (exposure) on gene expression (outcome) in each treatment
544 condition, and used the QTL SNP (rs66826786) in the co-localized triplets at 17q as the genetic
545 instrument (see Methods). In this way, we are able to estimate whether the effect of the asthma
546 risk variant on gene expression levels is mediated by DNA methylation.

547 MR suggested a causal relationship between methylation and gene expression in RV-
548 treated cells for the co-localized triplet, indicating that the genotype effect at rs66826786 on
549 expression of *ERBB2* is mediated by methylation at the meCpG (Table 4). The contribution of
550 the meCpG on *ERBB2* gene expression was only detected in RV-treated cells (P-value<$1x10^{-10}$)
551 while no evidence was detected in vehicle-treated cells (P-value=0.81), further suggesting a gene
552 regulatory mechanism that is triggered after exposure to RV. The MR result provides orthogonal
553 evidence for the co-localization of this triplet and novel evidence for causal inference with
554 respect to the co-localized traits (DNA methylation, gene expression). These data also reinforce
555 arguments for epigenetic mechanisms modifying gene expression, and potentially disease risk, in
556 response to environmental exposures [65, 66].

557

558 **Discussion**

559 One of the major challenges of complex disease genetics is to uncover molecular mechanisms of
560 pathogenesis and to understand how genetic and environmental factors interact to influence risks
561 for disease. While GWASs have identified thousands of SNPs associated with disease
562 phenotypes, interpretation and downstream follow-up studies of GWAS results have been
563 limited. Cell models can advance our understanding of disease pathobiology through
564 experimental testing of disease mechanisms in a controlled environment. In this multi-omics
565 study, we leveraged an airway epithelial cell model of microbial response to identify potentially
566 functional variants, some of which have context-specific effects on transcriptional and epigenetic
567 responses, and molecular mechanisms of disease. We show that asthma GWAS SNPs were
568 specifically enriched among molecular QTLs in airway epithelial cells compared to SNPs from
569 other GWASs, and among AEC eQTLs compared to eQTLs from other tissues. Finally, SNPs
570 that were molecular QTLs in our study co-localized with asthma GWAS SNPs, identifying 18
571 unique co-localizations that included both known asthma loci (e.g., 17q12-21 and *TSLP*) and loci
572 that did not meet stringent criteria for genome-wide significance in the GWASs (Table S2).

573 The results of enrichment analyses further highlighted the important role of airway
574 epithelium in asthma GWAS discoveries. The enrichment of childhood onset asthma GWAS
575 SNPs among epithelial eQTLs is particularly noteworthy, as it not only supports the tissue
576 specificity of our model but also identified genomic loci with molecular mechanisms that have
577 not been described prior to our study. These results are also consistent with previous studies
578 suggesting that functional variants from disease-relevant tissues are more enriched among
579 GWAS loci for those diseases [4, 41, 42]. The more modest enrichment of adult onset asthma
580 GWAS SNPs among epithelial eQTLs may be due to the overall smaller effect sizes of SNPs at
581 adult onset asthma loci compared to childhood onset asthma loci, to the less important role of

582 epithelial cells in the pathophysiology of adult onset asthma, or to the greater heterogeneity and
583 lesser heritability of adult onset asthma [1]. Other differences between the adult onset and
584 childhood onset asthma GWASs were observed. For example, only a single co-localization was
585 detected with adult onset asthma GWAS SNPs, compared to 19 with childhood onset asthma
586 GWAS SNPs. None of the co-localizations in the adult onset GWAS included an eQTL
587 compared to four childhood onset co-localizations with eQTLs, and the one meQTL-GWAS pair
588 in the adult onset asthma GWAS was also present in the childhood onset asthma. These
589 differences were additionally surprising because although there were 2.5-times the number of
590 loci associated with childhood onset asthma compared to adult onset asthma in the GWASs [1],
591 there were nearly 20-times more co-localizations in the childhood onset compared to the adult
592 onset GWAS (19 vs. 1, respectively). These observations likely reflect the more important role
593 of gene regulation and dysregulation in airway epithelium in the etiology of childhood onset
594 asthma compared to adult onset asthma [17, 18]. Focusing on other asthma relevant tissues (e.g.,
595 lung tissue) or cells (e.g., immune cells) might reveal additional novel molecular mechanisms
596 and differences between childhood onset and adult onset asthma.

597 Our study provides mechanistic evidence for associations between GWAS SNPs and
598 asthma at two important loci: the *TSLP* and 17q12-21 loci. Co-localizations of the asthma
599 associated SNP rs1837253 with DNA methylation levels in the *TSLP* gene suggest an epigenetic
600 mechanism of disease that contributes to both adult and childhood onset asthma, and is robust to
601 RV versus vehicle treatment. Associations of this SNP with asthma have been highly replicated
602 in GWASs, and TSLP is recognized as having an important role in asthma pathogenesis through
603 its broad effects on innate and adaptive immune cells promoting Th2 inflammation [67]. Our
604 data further show that the effect of rs1837253 genotype on risk for asthma may be mediated

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

608 Since its discovery over a decade ago, the 17q12-21 locus has been an important focus of
609 asthma research. Several studies have revealed the complex nature of this locus including the
610 differences in LD structure across populations, and contrasting gene expression patterns and
611 eQTLs at this locus in asthma-relevant cell types (reviewed in [2, 61]). In our study, using an
612 airway epithelium cell model of RV infection, additional dimensions of complexity at this locus
613 were revealed. For example, genes in the core region have been considered the most likely
614 candidate genes mediating effects of genetic variation on risk of childhood onset asthma [61].
615 However, our study further shows that genes at both the proximal and distal ends of this locus,
616 *ERBB2* and *GSDMA*, respectively, may contribute to asthma risk in the presence of RV
617 infection. Mendelian randomization revealed a novel epigenetic mechanism through which a
618 SNP at the distal boundary of the locus was associated with expression of *ERBB2* at the proximal
619 boundary of the locus, only after exposure to RV. The eQTL effect on *ERBB2* expression in RV-
620 treated cells was mediated through differential methylation of a CpG site at the distal locus,
621 which was present in both treatment conditions. Previous studies have shown that variation at the
622 17q core locus confers risk to asthma only among children with wheezing illness in early life
623 [68], particularly with RV-associated wheezing [59, 60]. Our study further connected RV
624 infection and genotype at this locus to the *ERBB2* gene for the first time, as well as to an
625 interaction between genetic and methylation variation at the distal end of the locus with the
626 expression of *ERBB2* at the proximal end of the locus in RV infected epithelial cells. The SNP
627 that is the eQTL for *ERBB2* in RV infected epithelial cells was associated with childhood onset

628 asthma ($p_{GWAS} = 6.43x10^{-26}$ [1]), directly connecting the eQTL for *ERBB2* in RV-treated cells to
629 asthma risk. The asthma associated allele, rs66826786-T, was associated with decreased
630 expression of *ERBB2* after RV infection in our study (Fig. 3A), consistent with results of a study
631 of 155 asthma cases and controls reporting an inverse correlation between *ERBB2* expression in
632 *ex vivo* lower AECs and asthma severity [47]. These combined data suggest that decreased
633 expression of *ERBB2* associated with asthma severity may be modulated by RV, the most
634 common trigger of asthma exacerbations, via epigenetic mechanisms involving DNA
635 methylation and long-range chromatin looping between the proximal and distal ends of this
636 important locus. In addition, meQTLs in *GSDMA*, at the proximal end of the locus, co-localized
637 with GWAS SNPs in RV-treated cells only. Together, these findings further highlight the
638 importance of RV exposure at this prominent asthma risk locus and provide mechanistic
639 evidence for a genotype by exposure interaction, and raise the possibility that SNPs in the core
640 region primarily confer risk for inception of early onset asthma whereas SNPs in the proximal
641 and distal ends of the locus primarily modulate gene-environment interactions.

642 Many of the associations in GWASs that do not reach stringent criteria for genome-wide
643 significance ($p<5x10^{-8}$) may be true signals. Distinguishing true from false positive signals for
644 variants among the mid-hanging fruit (e.g., p-values between 10^{-5} and $> 10^{-8}$) can be challenging.
645 In our study, over 57% of the co-localizations were with a GWAS SNP that did not meet
646 genome-wide significance (childhood onset asthma GWAS p-value range $6.1x10^{-7} – 1.4x10^{-5}$;
647 Table S2). One possibility for this is because the variants have exposure-specific, tissue-specific,
648 or endotype-specific effects, which are heterogeneous among subjects included in GWASs.
649 Therefore, annotating SNPs among the mid-hanging fruit for functionality provides more

650 confidence to these findings, a more complete picture of the genetic architecture of asthma, and a
651 model for prioritizing these loci for further studies.
652 Our study has several limitations. First, the sample sizes for the eQTL and meQTL
653 studies were smaller than the most reliable sample size recommended by moloc (n_{min}=300) [13].
654 In such cases, *moloc* can miss true co-localizations in QTL datasets. For example, an eQTL-
655 GWAS pair with supporting evidence may, in reality, be an eQTL-meQTL-GWAS triplet. As a
656 result, the eQTL-GWAS and meQTL-GWAS pairs that we identified could be eQTL-meQTL-
657 GWAS triplets that we were not powered to detect, or we may have missed other co-localizations
658 entirely. For example, although only a single meQTL co-localized with a GWAS SNP at the
659 *TSLP* locus, the same SNP, rs1837253, was an meQTL for three additional CpGs (Fig. S6),
660 representing additional potential contributors to asthma disease mechanisms. Nonetheless, the 19
661 unique co-localizations detected in our study are likely to be real, although future studies in
662 larger samples will increase confidence in our findings. Second, we focused our studies on one
663 cell type (upper airway sinonasal epithelium), two exposures (vehicle and RV), and one
664 epigenetic mark (DNA methylation). It is possible that other asthma-relevant co-localizations are
665 specific other tissues or cell types or to other exposures or culture conditions, and that additional
666 epigenetic marks, such as those associated with chromatin accessibility, would be additionally
667 informative. These extended studies will be necessary to validate the specificity and provide a
668 more complete catalog of asthma-relevant co-localizations. Finally, characterizing chromatin
669 conformational changes in AECs before and after exposure to RV will allow a direct assessment
670 of the chromatin looping at the extended 17q12-21 locus that may occur in response to viral
671 infection and potentially identify other context-specific interactions.

672 In summary, we identified *cis*-eQTLs and *cis*-meQTLs in an airway epithelial cell model
673 of host cell response to RV and integrated those data with asthma GWASs to assign potential
674 molecular mechanisms for variants associated with asthma in two large GWASs. By combining
675 enrichment studies, co-localization analysis, and Mendelian randomization, we provide robust
676 statistical evidence of epigenetic mechanisms in upper airway cells contributing to childhood
677 onset asthma. We demonstrate that a multi-omics approach using a disease-relevant cell type and
678 disease-relevant exposure allows prioritization of disease-associated variants and provides
679 insight into potential epigenetic mechanisms of asthma pathogenesis.
680

681 ## Acknowledgments

682 The authors acknowledge Christine Billstrand and Raluca Nicolae for sample processing
683 and library preparation, and study subjects for their participation. This work was supported by
684 NIH grants U19 AI106683 and R01 HL129735. M.M.S. was supported in part by T32
685 GM007197.

686

687 ## References

688 1. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK: **Shared and distinct genetic risk**
689 **factors for childhood-onset and adult-onset asthma: genome-wide and**
690 **transcriptome-wide studies**. *Lancet Respir Med* 2019, **7**(6):509-522.

691 2. Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, Igartua
692 C, Morin A, Washington C, 3rd, Nicolae D *et al*: **A decade of research on the 17q12-21**
693 **asthma locus: Piecing together the puzzle**. *J Allergy Clin Immunol* 2018.

694 3. Gerasimova A, Chavez L, Li B, Seumois G, Greenbaum J, Rao A, Vijayanand P, Peters
695 B: **Predicting cell types and genetic variations contributing to disease by combining**
696 **GWAS and epigenetic data**. *PLoS One* 2013, **8**(1):e54359.

697 4. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP,
698 Sandstrom R, Qu H, Brody J: **Systematic localization of common disease-associated**
699 **variation in regulatory DNA**. *Science* 2012:1222794.

700 5. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ: **Trait-associated SNPs**
701 **are more likely to be eQTLs: annotation to enhance discovery from GWAS**. *PLoS*
702 *Genet* 2010, **6**(4):e1000888.

703 6. Consortium GT: **Human genomics. The Genotype-Tissue Expression (GTEx) pilot**
704 **analysis: multitissue gene regulation in humans**. *Science* 2015, **348**(6235):648-660.

705 7. Consortium EP: **The ENCODE (ENCyclopedia Of DNA Elements) Project**. *Science*
706 2004, **306**(5696):636-640.

707 8. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-
708 Moussavi A, Kheradpour P, Zhang Z, Wang J *et al*: **Integrative analysis of 111**
709 **reference human epigenomes**. *Nature* 2015, **518**(7539):317-330.

710 9. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, Helmer Q,
711 Tillander A, Ullemar V, van Dongen J *et al*: **Shared genetic origin of asthma, hay fever**
712 **and eczema elucidates allergic disease biology**. *Nat Genet* 2017, **49**(12):1752-1757.

713 10. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, Barr
714 RG, Beaty TH, Becker AB, Beilby J *et al*: **Multiancestry association study identifies**
715 **new asthma risk loci that colocalize with immune-cell enhancer marks**. *Nat Genet*
716 2018, **50**(1):42-53.

717 11. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y: **Deciphering the**
718 **genetic architecture of variation in the immune response to Mycobacterium**
719 **tuberculosis infection**. *Proc Natl Acad Sci U S A* 2012, **109**(4):1204-1209.

720 12. Kim-Hellmuth S, Bechheim M, Putz B, Mohammadi P, Nedelec Y, Giangreco N, Becker
721 J, Kaiser V, Fricker N, Beier E *et al*: **Genetic regulatory effects modified by immune**
722 **activation contribute to autoimmune disease associations**. *Nat Commun* 2017,
723 **8**(1):266.

724 13. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, Pickrell J,
725 Jaffe AE, CommonMind C, Pasaniuc B *et al*: **A Bayesian Framework for Multiple**
726 **Trait Colo-calization from Summary Association Statistics**. *Bioinformatics* 2018.

727 14. Jackson DJ, Gangnon RE, Evans MD, Roberg KA, Anderson EL, Pappas TE, Printz MC,
728 Lee WM, Shult PA, Reisdorf E *et al*: **Wheezing rhinovirus illnesses in early life**

729 **predict asthma development in high-risk children**. *Am J Respir Crit Care Med* 2008,
730 **178**(7):667-672.

731 15. Johnston SL, Pattemore PK, Sanderson G, Smith S, Lampe F, Josephs L, Symington P,
732 O'Toole S, Myint SH, Tyrrell DA *et al*: **Community study of role of viral infections in**
733 **exacerbations of asthma in 9-11 year old children**. *BMJ* 1995, **310**(6989):1225-1229.

734 16. Busse WW, Lemanske RF, Jr., Gern JE: **Role of viral respiratory infections in asthma**
735 **and asthma exacerbations**. *Lancet* 2010, **376**(9743):826-834.

736 17. Busse WW: **The atopic march: Fact or folklore?** *Ann Allergy Asthma Immunol* 2018,
737 **120**(2):116-118.

738 18. Han H, Roan F, Ziegler SF: **The atopic march: current insights into skin barrier**
739 **dysfunction and epithelial cell-derived cytokines**. *Immunol Rev* 2017, **278**(1):116-130.

740 19. Delaneau O, Zagury JF, Marchini J: **Improved whole-chromosome phasing for disease**
741 **and population genetic studies**. *Nat Methods* 2013, **10**(1):5-6.

742 20. Howie B, Marchini J, Stephens M: **Genotype imputation with thousands of genomes**.
743 *G3 (Bethesda)* 2011, **1**(6):457-470.

744 21. Freeman C, Marchini J: **GTOOL: A program for transforming sets of genotype data**
745 **for use with the programs SNPTEST and IMPUTE, Oxford, UK**. *GTOOL: A*
746 *program for transforming sets of genotype data for use with the programs SNPTEST and*
747 *IMPUTE, Oxford, UK.*

748 22. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M,
749 Kang HM: **Detecting and estimating contamination of human DNA samples in**
750 **sequencing and array-based genotype data**. *Am J Hum Genet* 2012, **91**(5):839-848.

751 23. Dobin A, Gingeras TR: **Mapping RNA-seq Reads with STAR**. *Curr Protoc*
752 *Bioinformatics* 2015, **51**:11 14 11-19.

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

756 25. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing**
757 **batch effects and other unwanted variation in high-throughput experiments**.
758 *Bioinformatics* 2012, **28**(6):882-883.

759 26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers**
760 **differential expression analyses for RNA-sequencing and microarray studies**.
761 *Nucleic Acids Res* 2015, **43**(7):e47.

762 27. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD,
763 Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the**
764 **analysis of Infinium DNA methylation microarrays**. *Bioinformatics* 2014,
765 **30**(10):1363-1369.

766 28. Maksimovic J, Gordon L: **SWAN: Subset-quantile within array normalization for**
767 **illumina infinium HumanMethylation450 BeadChips**. *SWAN: Subset-quantile within*
768 *array normalization for illumina infinium HumanMethylation450 BeadChips* 2012.

769 29. Peters TJ, Buckley MJ: **De novo identification of differentially methylated regions in**
770 **the human genome**. *De novo identification of differentially methylated regions in the*
771 *human genome* 2015.

772 30. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O: **Fast and efficient QTL**
773 **mapper for thousands of molecular phenotypes**. *Bioinformatics* 2016, **32**(10):1479-
774 1485.

775 31. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET: **A complete**
776 **tool set for molecular QTL discovery and analysis**. *Nat Commun* 2017, **8**:15452.

777 32. Urbut SM, Wang G, Carbonetto P, Stephens M: **Flexible statistical methods for**
778 **estimating and testing effects in genomic studies with multiple conditions**. *Nat Genet*
779 2019, **51**(1):187-195.

780 33. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K, Timpson NJ,
781 Consortium UK, Dunham I, Birney E *et al*: **GARFIELD classifies disease-relevant**
782 **genomic features through integration of functional annotations with association**
783 **signals**. *Nat Genet* 2019, **51**(2):343-353.

784 34. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J,
785 Karlsson IK, Hagg S, Athanasiu L *et al*: **Genome-wide meta-analysis identifies new**
786 **loci and functional pathways influencing Alzheimer's disease risk**. *Nat Genet* 2019,
787 **51**(3):404-413.

788 35. Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE, Herron TJ,
789 McCarthy S, Schmidt EM, Sveinbjornsson G *et al*: **Biobank-driven genomic discovery**
790 **yields new insight into atrial fibrillation biology**. *Nat Genet* 2018, **50**(9):1234-1239.

791 36. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K,
792 Luan J, Kutalik Z *et al*: **Defining the role of common variation in the genomic and**
793 **biological architecture of adult human height**. *Nat Genet* 2014, **46**(11):1173-1186.

794 37. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, Harris SE, Marioni
795 RE, Liewald DC, Fawns-Ritchie C *et al*: **Association analysis in over 329,000**
796 **individuals identifies 116 independent variants influencing neuroticism**. *Nat Genet*
797 2018, **50**(1):6-11.

798 38. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET,
799 Sperling AI, Solway J, White SR, Nobrega MA *et al*: **DNA methylation in lung cells is**
800 **associated with asthma endotypes and genetic risk**. *JCI Insight* 2016, **1**(20):e90151.

801 39. Stephens M: **False discovery rates: a new deal**. *Biostatistics* 2017, **18**(2):275-294.

802 40. Gulko B, Hubisz MJ, Gronau I, Siepel A: **A method for calculating probabilities of**
803 **fitness consequences for point mutations across the human genome**. *Nat Genet* 2015,
804 **47**(3):276-283.

805 41. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical
806 Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri,
807 Nih/Nimh, Nih/Nida *et al*: **Genetic effects on gene expression across human tissues**.
808 *Nature* 2017, **550**(7675):204-213.

809 42. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu
810 H, Zang C, Farh K *et al*: **Partitioning heritability by functional annotation using**
811 **genome-wide association summary statistics**. *Nat Genet* 2015, **47**(11):1228-1235.

812 43. Consortium GT: **The Genotype-Tissue Expression (GTEx) project**. *Nat Genet* 2013,
813 **45**(6):580-585.

814 44. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C,
815 Plagnol V: **Bayesian test for colocalisation between pairs of genetic association**
816 **studies using summary statistics**. *PLoS Genet* 2014, **10**(5):e1004383.

817 45. Wen X, Pique-Regi R, Luca F: **Integrating molecular QTL data into genome-wide**
818 **genetic association analysis: Probabilistic assessment of enrichment and**
819 **colocalization**. *PLoS Genet* 2017, **13**(3):e1006646.

820 46. Sandilands A, Sutherland C, Irvine AD, McLean WH: **Filaggrin in the frontline: role in**
821 **skin barrier function and disease**. *J Cell Sci* 2009, **122**(Pt 9):1285-1294.

822 47. Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, Meyers
823 DA, Milosevic J, Tedrow JR, Wu W: **Gene expression correlated with severe asthma**
824 **characteristics reveals heterogeneous mechanisms of severe disease**. *American*
825 *journal of respiratory and critical care medicine* 2017, **195**(11):1449-1463.

826 48. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg
827 A, Bufe A, Rietschel E *et al*: **Genetic variants regulating ORMDL3 expression**
828 **contribute to the risk of childhood asthma**. *Nature* 2007, **448**(7152):470-473.

829 49. Ober C: **Asthma Genetics in the Post-GWAS Era**. *Annals of the American Thoracic*
830 *Society* 2016, **13 Suppl 1**(Supplement 1):90.

831 50. McCarthy MI, Hirschhorn JN: **Genome-wide association studies: potential next steps**
832 **on a genetic journey**. *Human molecular genetics* 2008, **17**(R2):65.

833 51. Bonnelykke K, Ober C: **Leveraging gene-environment interactions and endotypes for**
834 **asthma gene discovery**. *J Allergy Clin Immunol* 2016, **137**(3):667-679.

835 52. Ying S, O'Connor B, Ratoff J, Meng Q, Mallett K, Cousins D, Robinson D, Zhang G,
836 Zhao J, Lee TH *et al*: **Thymic stromal lymphopoietin expression is increased in**
837 **asthmatic airways and correlates with expression of Th2-attracting chemokines and**
838 **disease severity**. *J Immunol* 2005, **174**(12):8183-8190.

839 53. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE,
840 Himes BE, Levin AM, Mathias RA, Hancock DB *et al*: **Meta-analysis of genome-wide**
841 **association studies of asthma in ethnically diverse North American populations**. *Nat*
842 *Genet* 2011, **43**(9):887-892.

843 54. Shrine N, Portelli MA, John C, Soler Artigas M, Bennett N, Hall R, Lewis J, Henry AP,
844 Billington CK, Ahmad A *et al*: **Moderate-to-severe asthma in individuals of European**
845 **ancestry: a genome-wide association study**. *Lancet Respir Med* 2019, **7**(1):20-34.

846 55. Wang IJ, Chen SL, Lu TP, Chuang EY, Chen PC: **Prenatal smoke exposure, DNA**
847 **methylation, and childhood atopic dermatitis**. *Clin Exp Allergy* 2013, **43**(5):535-543.

848 56. Luo Y, Zhou B, Zhao M, Tang J, Lu Q: **Promoter demethylation contributes to TSLP**
849 **overexpression in skin lesions of patients with atopic dermatitis**. *Clin Exp Dermatol*
850 2014, **39**(1):48-53.

851 57. Hui CC, Yu A, Heroux D, Akhabir L, Sandford AJ, Neighbour H, Denburg JA: **Thymic**
852 **stromal lymphopoietin (TSLP) secretion from human nasal epithelium is a function**
853 **of TSLP genotype**. *Mucosal Immunol* 2015, **8**(5):993-999.

854 58. Smit LA, Bouzigon E, Pin I, Siroux V, Monier F, Aschard H, Bousquet J, Gormand F,
855 Just J, Le Moual N *et al*: **17q21 variants modify the association between early**
856 **respiratory infections and asthma**. *Eur Respir J* 2010, **36**(1):57-64.

857 59. Loss GJ, Depner M, Hose AJ, Genuneit J, Karvonen AM, Hyvarinen A, Roduit C,
858 Kabesch M, Lauener R, Pfefferle PI *et al*: **The Early Development of Wheeze.**
859 **Environmental Determinants and Genetic Susceptibility at 17q21**. *Am J Respir Crit*
860 *Care Med* 2016, **193**(8):889-897.

861 60. Caliskan M, Bochkov YA, Kreiner-Moller E, Bonnelykke K, Stein MM, Du G, Bisgaard
862 H, Jackson DJ, Gern JE, Lemanske RF, Jr. *et al*: **Rhinovirus wheezing illness and**
863 **genetic risk of childhood-onset asthma**. *N Engl J Med* 2013, **368**(15):1398-1407.

864 61. Ober C, McKennan CG, Magnaye KM, Altman MC, Washington C, 3rd, Stanhope C,
865 Naughton KA, Rosasco MG, Bacharier LB, Billheimer D *et al*: **Expression quantitative**
866 **trait locus fine mapping of the 17q12-21 asthma locus in African American children:**
867 **a genetic association and gene expression study**. *Lancet Respir Med* 2020, **8**(5):482-
868 492.

869 62. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E,
870 Farrall M, Lathrop M, Cookson W *et al*: **A large-scale, consortium-based genomewide**
871 **association study of asthma**. *N Engl J Med* 2010, **363**(13):1211-1221.

872 63. Marinho S, Custovic A, Marsden P, Smith JA, Simpson A: **17q12-21 variants are**
873 **associated with asthma and interact with active smoking in an adult population**
874 **from the United Kingdom**. *Ann Allergy Asthma Immunol* 2012, **108**(6):402-411 e409.

875 64. Smith GD: **Mendelian Randomization for Strengthening Causal Inference in**
876 **Observational Studies: Application to Gene x Environment Interactions**. *Perspect*
877 *Psychol Sci* 2010, **5**(5):527-545.

878 65. Liu J, Ballaney M, Al-alem U, Quan C, Jin X, Perera F, Chen LC, Miller RL: **Combined**
879 **inhaled diesel exhaust particles and allergen exposure alter methylation of T helper**
880 **genes and IgE production in vivo**. *Toxicol Sci* 2008, **102**(1):76-81.

881 66. Clifford RL, Jones MJ, MacIsaac JL, McEwen LM, Goodman SJ, Mostafavi S, Kobor
882 MS, Carlsten C: **Inhalation of diesel exhaust and allergen alters human bronchial**
883 **epithelium DNA methylation**. *J Allergy Clin Immunol* 2017, **139**(1):112-121.

884 67. West EE, Kashyap M, Leonard WJ: **TSLP: A Key Regulator of Asthma Pathogenesis**.
885 *Drug Discov Today Dis Mech* 2012, **9**(3-4).

886 68. Bouzigon E, Corda E, journal … A-H: **Effect of 17q21 variants and smoking exposure**
887 **in early-onset asthma**. *New England journal* ... 2008.

888

889

890 **Fig. 1. Summary of molecular effects sharing across treatment conditions (lfsr<0.05).** Venn
891 diagrams of eGenes **(A)** and meCpGs **(B)** shared between vehicle- and RV-treated airway
892 epithelial cells. Forest plots showing examples of RV- (left) and Vehicle-specific (right) eQTLs
893 **(C)** and meQTLs **(D)**.

894

895 **Fig. 2. Co-localization of rs1837253 with DNA methylation levels for cg15557878 at *TSLP*.**
896 rs1837253 (red vertical bar, upper panel)) is associated with DNA methylation levels at
897 cg15557878 (orange vertical bar, upper panel). Box plots show DNA methylation levels (y-axes)
898 for each meCpGs by rs1837253 genotype (x-axes) in each treatment condition (lower panel).

899

900 **Fig. 3. Co-localization pairs at the 17q asthma susceptibility locus.** Upper panel: Box plots
901 for the co-localized *ORMDL3* eQTL in cultured airway epithelial cells treated with vehicle **(A)**
902 and RV **(B)**. The effect size of the correlation between *ORMDL3* and rs12603332 decreases after
903 treatment with RV (FDR<0.10); and *ORMDL3* gene expression decreases after treatment with
904 RV **(C)**. Box plots of the cg21230266 meQTL treated with vehicle (**D**) and RV (**E**), and
905 methylation levels at cg21230266 (**F**). The meQTLs and overall methylation levels are similar
906 in vehicle and RV treatments. Lower panel: The extended 17q12-21 locus. Co-localizations are
907 shown by the vertical colored lines; rs4239225, a SNP in LD with rs3859191 (solid turquoise
908 lines), is also significantly correlated with DNA methylation levels at cg21230266 in vehicle and

909 RV (not shown; see Table S2). See Fig 4B for box plots of the cg17401724:rs66826786 meQTL.
910 Solid lines indicate the position of the colocalized SNP. Dashed lines indicate the location of
911 meCpG pairs. Traits of the same co-localization are shown in the same color. A single eQTL-
912 GWAS pair for *ORMDL3* is shown in orange; meQTL-GWAS pairs are shown in turquoise and
913 green.
914
915 **Fig. 4. Co-localization of rs66826786 with *ERBB2* expression and DNA methylation levels**
916 **for cg17401724. (A)** LocusZoom plots of childhood onset GWAS results at the 17q locus
917 showing the *ERBB2* gene at the proximal (left) end of the locus and the co-localized eQTL
918 (rs668826786) at the distal (right) end of the locus (modified from Pividori ?). The SNP
919 (rs66826786), which colocalized with associations for childhood onset asthma, *ERBB2*
920 expression, and DNA methylation at cg17401724, is shown as a purple diamond in each of three
921 LocusZoom plot. Upper panel: childhood onset asthma GWAS (modified from Pividori et al.
922 2019). Middle panel: *ERBB2* eQTLs for vehicle-treated cultured airway epithelial cells. Lower
923 panel: *ERBB2* eQTLs for RV-treated airway epithelial cells. Boxplots for *ERBB2* gene
924 expression by rs66826786 genotype is shown within the middle and lower LocusZoom plots. **(B)**
925 Boxplots for cg17401724 meQTLs in vehicle-treated (upper panel) and RV-treated (lower panel)
926 cultured airway epithelial cells. **(C)** *ERBB2* gene expression in vehicle-treated and RV-treated
927 cells.
928
929
930
931

932 ## Supporting information

933 **Fig. S1** Overview of the e/meQTL and co-localization studies in NECs treated with RV. **(A)**
934 Step-wise experimental design to identify treatment-specific e/meQTLs in NECs from 104
935 individuals: 1. NECs collected from study participants were cultured and treated with RV and a
936 vehicle for 48 hours. 2. Gene expression and DNA methylation measured in NECs from each
937 treatment condition. 3. Genotype profiling to identify genetic variation influencing gene
938 expression and DNA methylation to RV- and SA-treatment. 4. QC and analyses including
939 e/meQTL mapping, multi-trait co-localization analysis, and Mendelian randomization. **(B)**
940 Breakdown of the number of subjects for each experiment and molecular QTL mapping.
941

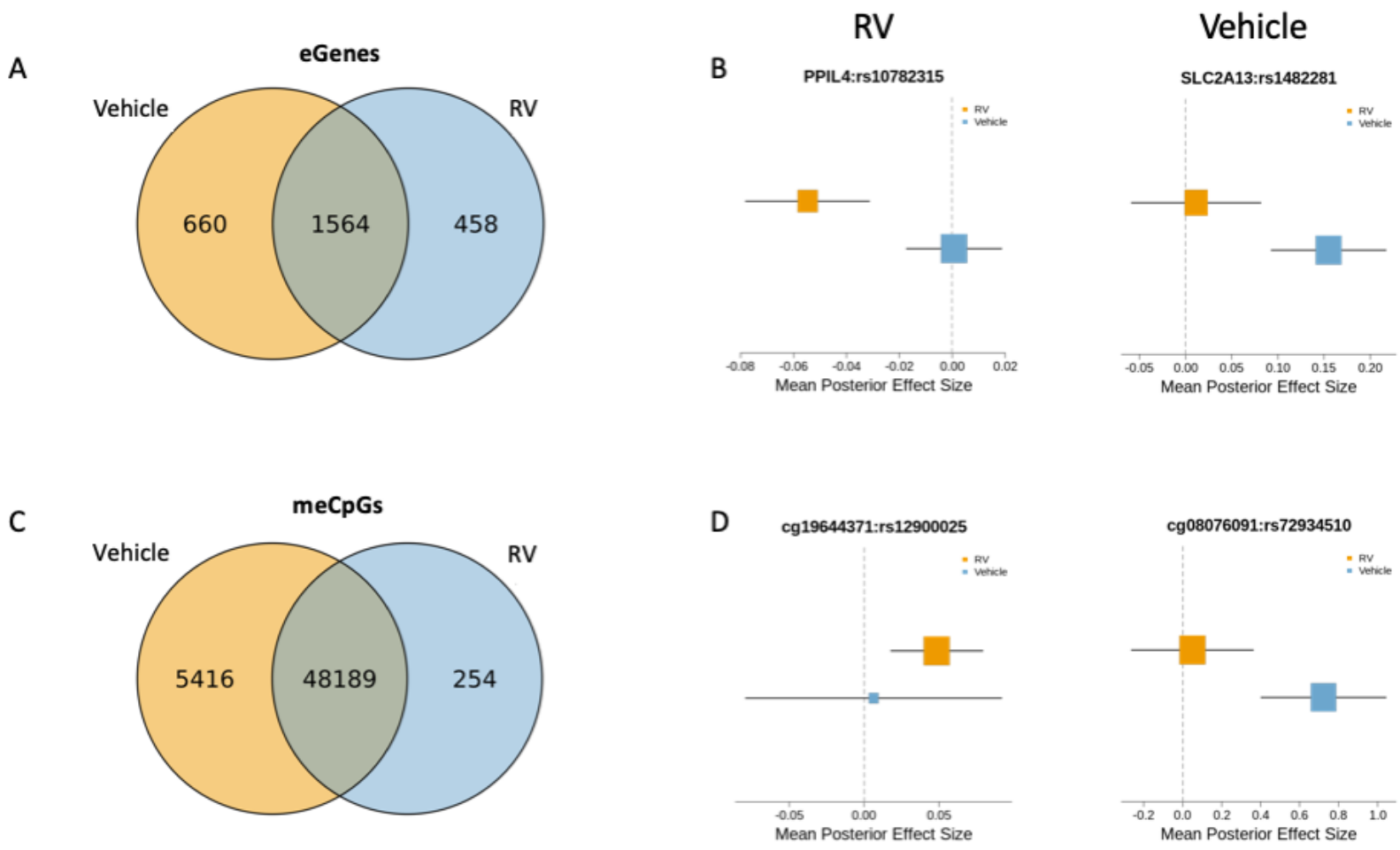
942 **Fig. S2 PCA and k-means clustering of genotypes. (A)** PCA plot of study participant's
943 genotypes (circles) projected on HapMap genotypes (squares). **(B)** Scree plot of k-means
944 clustering of ancestral PCs in which the within groups sum of squares (y-axis) is plotted against
945 the number of potential group clusters (x-axis); using the 'elbow criterion', it is determined that
946 two clusters are best representative of how many clusters study samples can be grouped into. **(C)**
947 PCA plot of study participants grouped into two cluster for genotype imputation, European (red),
948 and African American (Blue), according to the k-means clustering criterion.
949

950 **Fig. S3** PCA of gene expression in RV- and vehRV-treated epithelial cells. **(A)** PCA plot of
951 epithelial cell gene expression from 95 individuals treated with vehicle and RV before regressing
952 out covariates. **(B)** PCA plot of gene expression in vehicle- and RV-treated cells after regressing
953 out covariates. Tables showing p-values of correlation with PCs and covariates before **(C)** and

954 after **(D)** regression. **(E)** Volcano plot showing differential gene expression in response to RV
955 treatment.
956
957 **Fig. S4** PCA of DNA methylation in vehicle- and RV-treated in cultured airway epithelial cells.
958 **(A)** PCA plot of cultured airway epithelial DNA methylation from 103 individuals treated with
959 vehicle and RV before regressing out covariates. **(B)** PCA plot of DNA methylation in vehicle-
960 and RV-treated cells after regressing out covariates. Tables showing p-values of correlation with
961 PCs and covariates before **(C)** and after **(D)** regression. **(E)** Volcano plot showing differential
962 DNA methylation in response to RV treatment.
963
964 **Fig. S5** Summary results for molecular QTL mappings. Venn diagrams of eQTLs **(A)** and
965 meQTLs **(B)** in each condition (FDR<0.10). **(C)** Summary of eQTL and meQTL mapping results
966 for each treatment condition. The number of SNPs associated with the gene expression of at least
967 one gene or CpG and the number of genes or CpGs whose expression or DNA methylation levels
968 was associated with at least one SNP.
969
970 **Fig. S6** meQTLs at rs1837253 located in the first untranslated exon of the *TSLP* gene. Box plots
971 of three meQTLs that were identified in both the vehicle- (left) and RV-treated (right) AECs
972 were correlated with the asthma risk variant (rs1837253) but did not show evidence of co-
973 localization. These meQTLs may represent additional evidence of an epigenetic mechanism
974 contributing to asthma.
975

976 **Table S1** Enrichment estimates of eQTLs for adult onset asthma GWAS SNPs from six tissues.
977 P-values that are significant after BH correction are shown in bolded type.
978
979 **Table S2** *moloc* results indicating molecular QTL-GWAS pars and triplets
980
981 **Table S3** Adult onset and childhood onset asthma GWAS risk allele effects on gene expression
982 and DNA methylation
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997

Figure 1

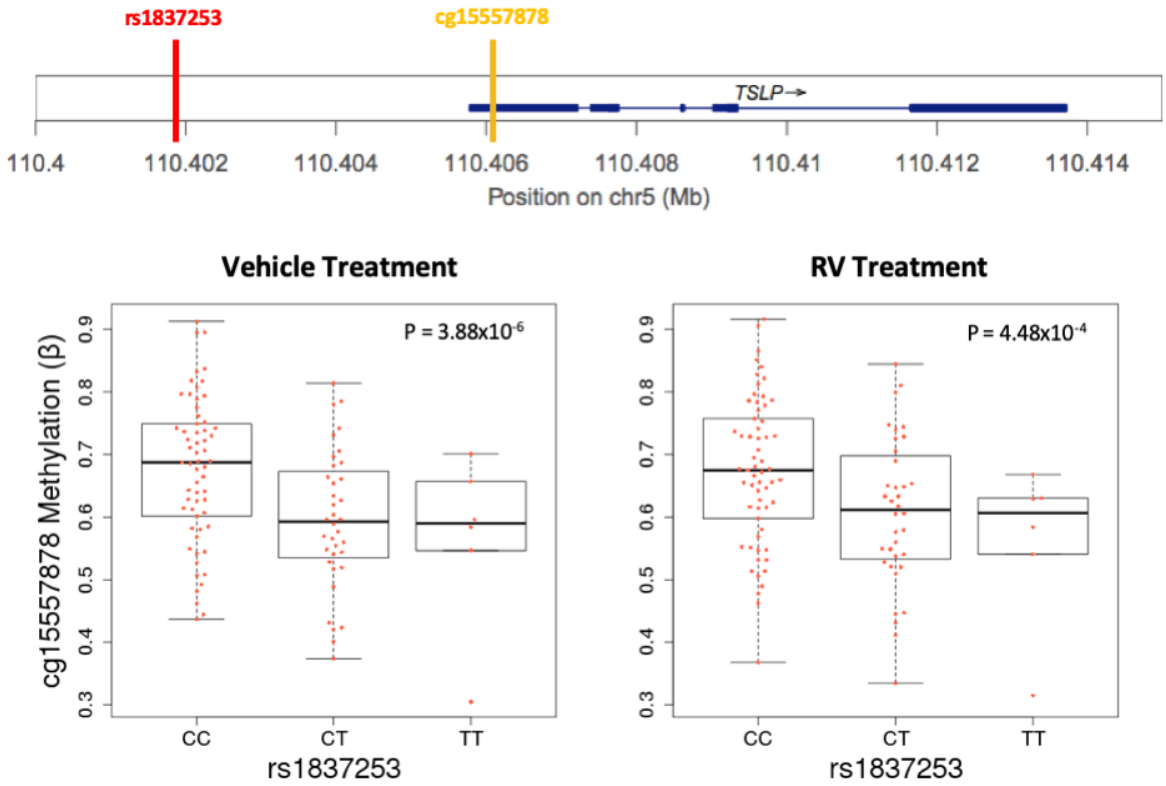

A
eGenes
Vehicle
RV
660
1564
458
B
RV
Vehicle
PPIL4:rs10782315
SLC2A13:rs1482281
RV
Vehicle
-0.08 -0.06 -0.04 -0.02 0.00 0.02
-0.05 0.00 0.05 0.10 0.15 0.20
Mean Posterior Effect Size
Mean Posterior Effect Size
C
meCpGs
Vehicle
RV
5416
48189
254
D
cg19644371:rs12900025
cg08076091:rs72934510
RV
Vehicle
-0.05 0.00 0.05
-0.2 0.0 0.2 0.4 0.6 0.8 1.0
Mean Posterior Effect Size
Mean Posterior Effect Size

1001 **Figure 2**

1002

1003

1004



1005

1006

1007

1008

1009

1010

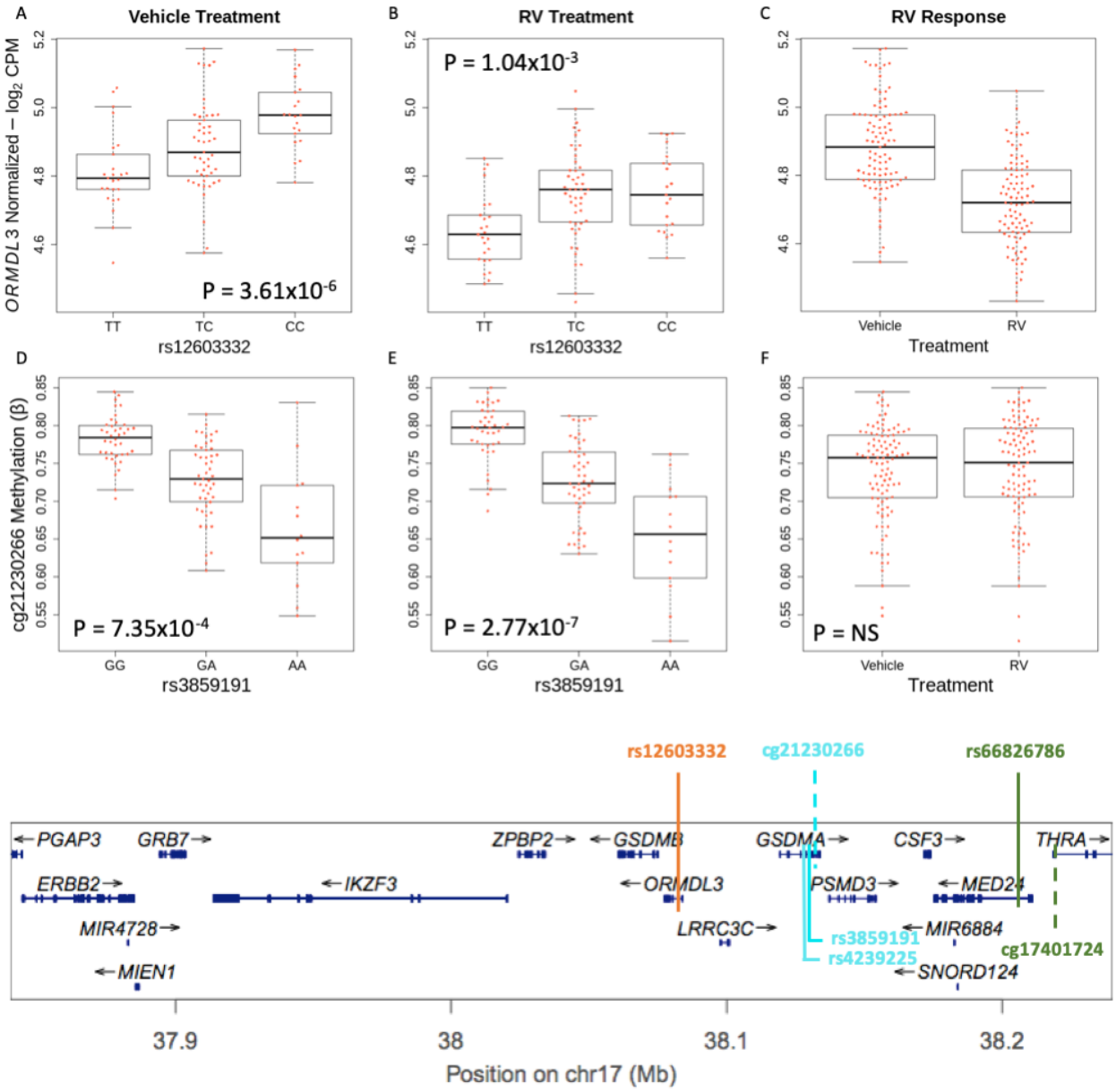
1011

1012

1013 **Figure 3**

1014

1015



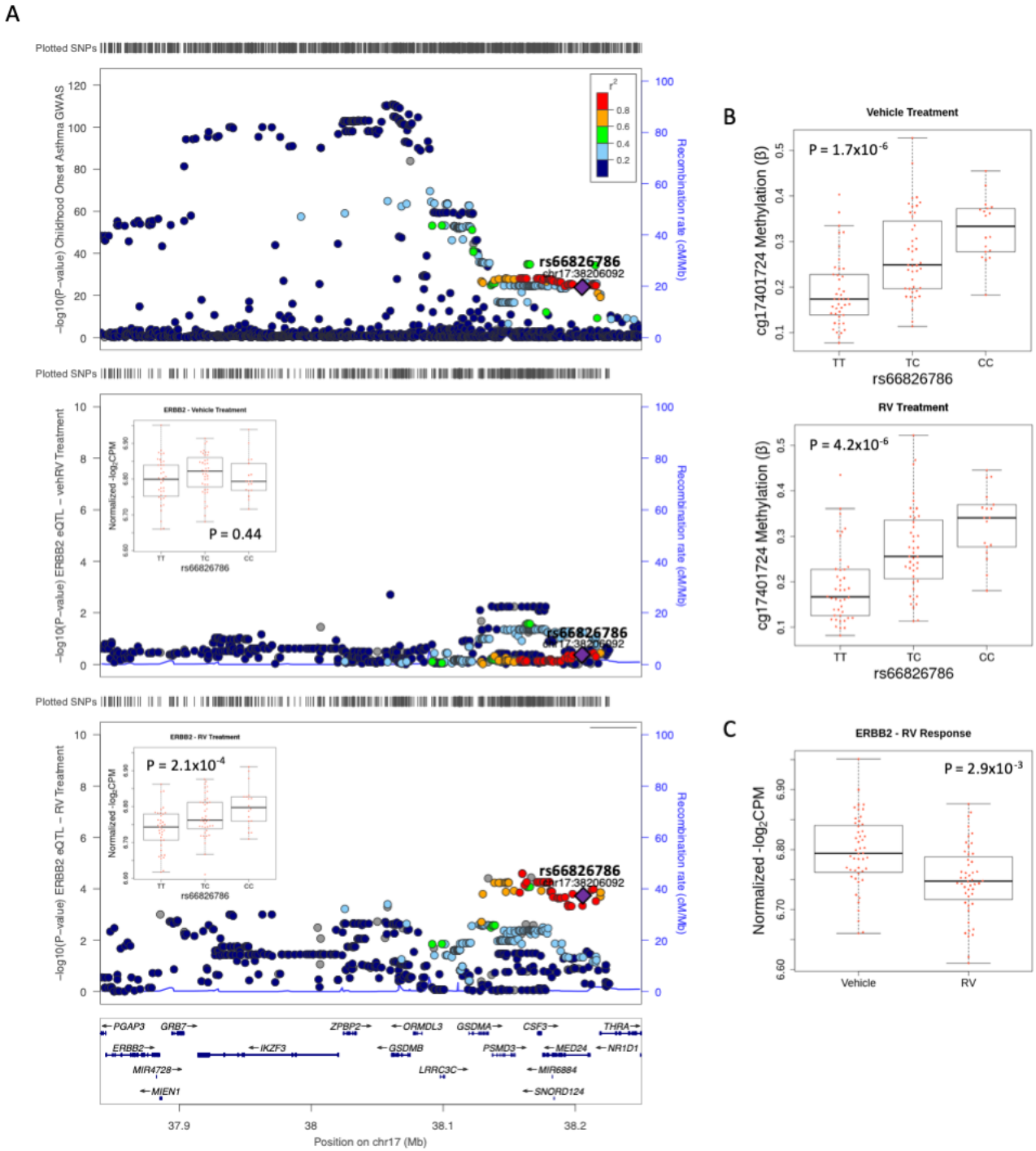
1016

1017

1018

1019

1020 **Figure 4**



1021