TITLE: An Integrative Transcriptomic and Metabolomic Study of Lung Function in Asthmatic Children

AUTHORS: Rachel S. Kelly PhD [a], Bo L. Chawes MD PhD [a,b], Kevin Blighe PhD [a], Yamini V. Virkud MD [a, c], Damien C. Croteau-Chonka PhD [a], Michael J. McGeachie PhD [a], Clary B. Clish[d], Kevin Bullock[d], Juan C. Celedón MD Dr PH [e], Scott T. Weiss MD [a], Jessica A. Lasky-Su ScD [a]

AFFILIATIONS:

a) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston MA 02115, USA

b) COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

c) Division of Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh of the University of Pittsburgh Medical Center, University of Pittsburgh, Pittsburgh, PA 15260, USA

d) Broad Institute of MIT and Harvard, Cambridge, MA.02142, USA

e) Division of Allergy and Immunology, Department of Pediatrics, Massachusetts General Hospital, Boston MA 02114, USA

Corresponding Author: Rachel S Kelly: hprke@channing.harvard.edu, Channing Division of Network Medicine, 181 Longwood Ave, Boston MA 01239. Tel: +16175250065

Abbreviations:

BDR; Bronchodilator response

CAMP – Childhood Asthma Management Program

FEV_1; Forced expiratory volume in one second

FVC; Forced vital capacity

GPI; Glycosylphosphatidylinositol

HMDB; Human Metabolome Database

IMPaLA: Integrated Molecular Pathway Level Analysis

LC-MS; Liquid chromatography-tandem mass spectrometry

m/z; Mass to Charge Ratio

PD20; Airway responsiveness to methacholine, determined as the provocative dose of methacholine resulting in a 20% drop in FEV_1 from baseline

SPT; Serine palmitoyltransferase

RT; Retention Time

WGCNA; Weighted gene co-expression network analysis

Running title: Integrative Omics and asthmatic lung function

Declaration of interests: Jessica Lasky-Su is a consultant to Metabolon, Inc. The other authors have no relevant conflicts of interest to disclose.

F**unding**

The Genetics of Asthma in Costa Rica Study was supported by grants HL066289 and HL04370 from the U.S. NIH. The metabolomics work was supported by 1R01HL123915-01. Dr. Celedón's contribution was supported by grants HL117191 and HL119952 from the U.S. NIH.

ABSTRACT

Background: Single-omic analyses have provided some insight into the basis of lung function in asthmatic children, but the underlying biological pathways are still poorly understood.

Methods: Weighted gene co-expression network analysis (WGCNA) was used to identify modules of co-regulated gene transcripts and metabolites in blood, among 325 children with asthma from the 'Genetic Epidemiology of Asthma in Costa Rica' study. The biology of modules associated with lung function; as measured by FEV_1, FEV_1/FVC-ratio, bronchodilator response, and airway responsiveness to methacholine was explored. Significantly correlated gene-metabolite module pairs were then identified and the constituent features submitted for integrated pathway analysis using IMPaLA.

Results: WGCNA clustered 25,060 gene probes and 8,185 metabolite features into eight gene modules and eight metabolite modules, where four and six, respectively, were associated with lung function ($p \leq 0.05$). The gene modules were enriched for immune, mitotic and metabolic processes and asthma associated microRNA targets. The metabolite modules were enriched for lipid and amino acid metabolism. Integration of correlated gene-metabolite modules expanded the single-omic findings, linking FEV_1/FVC-ratio with *ORMDL3* and dysregulated lipid metabolism. This finding was replicated in an independent population.

Conclusions: The results of this hypothesis-generating study suggest a mechanistic basis for multiple asthma genes, including *ORMDL3*, and a role for lipid metabolism. They demonstrate that integrating multiple omic technologies may provide a more informative picture of asthmatic lung function biology than single-omic analyses.

INTRODUCTION

Asthma; a disorder characterized by variable and reversible airway obstruction, hyper-responsiveness and inflammation, represents one of the most common chronic conditions among children and adults worldwide [1,2]. Asthmatic lung function abnormalities are present early in life [3,4], track through childhood and adulthood [5,6], and are strong determinants of disease exacerbations and severity [7,8].

Reduced lung function in asthmatics is thought to emerge from complex gene-environment interactions [9]. Advances in high-throughput technologies allow us to explore such interactions at the level of the epigenome, genome, transcriptome, proteome, and metabolome. Combining the transcriptome; which reflects genomic activity, with the metabolome; which is sensitive to environmental influences and closely related to phenotype, may be particularly informative. While previous studies have investigated metabolomic and transcriptomic profiles of asthma separately, to date only two studies with a limited sample size have integrated the two "omes" together in humans [10,11]. This integrative approach demonstrated increased predictive ability for asthma and its subtypes, as well as a greater biological insight relative to the use of single omics technologies. Consequently, integrative-omics represents an exciting new avenue in asthma research [12].

Currently, there are no analytical standard for integrative omics. However, network medicine; a rapidly emerging field that moves away from previous reductionist

methodologies to combine systems biology and network science in the study of complex disease, represents a particularly promising approach. It provides a holistic methodology to better understand disease through the identification and investigation of nonlinear relationships and networks of interacting components. This provides insights into these conditions beyond the level of a single, gene or omic platform. Weighted Gene Correlation Network Analysis (WGCNA) is a network method for identifying clusters or 'modules' or highly correlated variables, such as genes or metabolites that are likely to be co-regulated, or working together in biologically coherent fashion. This module can then be summarized as a single unit which can be correlated with phenotypes or other modules of interest.

The aim of this study was to conduct an integrated analysis of the blood transcriptome and metabolome among asthmatic children participating in the 'Genetic Epidemiology of Asthma in Costa Rica' [13] cohortin order to identify biologically informative networks of both genes and metabolites associated with asthmatic lung function. Then to identify relationships between gene and metabolite networks which were found to be associated with lung function. The Genetic Epidemiology of Asthma in Costa Rica cohort recruited children with mild to moderate asthma from the Central Valley of Costa Rica. This area represents a Hispanic population-isolate which is genetically homogenous and has one of the highest prevalence's of asthma in the world (24% in children) [15], making it uniquely suited for the exploration of the genomic and integrative omic underpinnings of asthmatic lung function. In particular, the study focuses on FEV_1 and FEV_1/FVC ratios, which are thought to mediate the association between early life characteristics and asthma [16].

METHODS

Study Population

This integrative omic study was nested within the 'Genetic Epidemiology of Asthma in Costa Rica Study [13], which recruited children aged 6-14 years with mild to moderate asthma and their parents from the Central Valley of Costa Rica. Children were eligible if they had physician-diagnosed asthma and at least two episodes of troublesome respiratory symptoms or asthma attacks in the prior year, and a high probability of having ≥6 great-grandparents born in the Central Valley of Costa Rica [14,17]. A total of 1,165 asthmatic children were enrolled in the original study. All children completed a protocol including questionnaires, spirometry, allergy skin testing, house dust samples, and collection of blood at enrolment, when children were not exacerbated. The majority of blood samples were processed within four hours; RNA was extracted and stored in PaxGene tubes. Genome-wide SNP genotyping and RNA expression profiles were generated for a subset of the study population with suitable samples. Genotype data was obtained with Taqman real-time PCR with an ABI Prism 7900 machine (Applied Biosystems, Foster City, CA)[18] . Standard PCR conditions, as recommended by the manufacturer, were used. Children were prioritized for metabolomics profiling if they had both genome-wide genetic and genome-wide expression data, with the goal of conducting integrated omic analyses. Children with both metabolomics and transcriptomic profiling were included in the current study Written parental and participating child consent was obtained. The study was approved by the Institutional Review Boards of the Hospital Nacional de Ninos (San Jose, Costa Rica) and Brigham and Women's Hospital (Boston, Mass, USA).

Lung Function

At enrollment, baseline lung function was investigated by spirometry (forced expiratory volume in one second (FEV_1), and the ratio of forced expiratory volume to forced vital capacity (FEV_1/FVC ratio)), bronchodilator response (BDR, percentage difference in FEV_1 from baseline after inhaled albuterol), and airway responsiveness to methacholine (PD20; determined as the provocative dose of methacholine resulting in a 20% drop in FEV_1 from baseline) (see **e-Methods** for details).

Transcriptomic Profiling

Whole-blood gene expression profiles were generated with 47,009 probes from the Illumina HumanHT-12 v4 Expression BeadChip (Illumina, Inc., San Diego, CA) that passed stringent and commonly used quality control (QC) metrics. Expression data were then log2-transformed and quantile-normalized as a single batch using the "lumiT" and "lumiN" functions, respectively, from the R package "lumi" (version 2.22). Principal components (PCs) of gene expression were generated specifically for each of the four cohorts using the "getPCAFunc" function from the R package "iCheck" (version 0.6), which provides several functions for visualizing QC metrics of Illumina gene expression data and for filtering based on those outputs. Phenotypes and gene expression were measured concurrently. Prior to downstream statistical analyses, we applied a standard non-specific variance filter to the expression data using the "nsFilter" function from the R package Croteau-Chonka et al. E5 "genefilter" (version 1.52). Probes not annotated with a valid Entrez gene identifier or Human Genome Organization (HUGO) gene symbol and probes with interquartile ranges (IQR) of expression variance below the 50th percentile

were removed to select only the most informative probes [19]. Data were then collapsed to a single probe per gene based on the largest IQR of expression variance [20].

Metabolomic Profiling

Plasma samples were shipped from the sample repository to the Broad Institute (Cambridge, MA, USA) on dry ice for metabolomic profiling. Samples were thawed on ice for sub-aliquoting for each of the metabolomic methods and then re-frozen on dry ice and stored at -80C until work up for LC-MS analyses. Four liquid chromatography-tandem mass spectrometry (LC-MS) platforms measured complementary sets of metabolite classes: (i) HILIC-positive platform, amines and polar metabolites that ionize in the positive ion mode using hydrophilic interaction liquid chromatography (HILIC) and MS analyses; (ii) HILIC-negative platform, central metabolites and polar metabolites that ionize in the negative ion mode using HILIC chromatography with an amine column and targeted MS; (iii) C8 platform, polar and non-polar lipids using reverse phase chromatography and full scan MS; and (iv) C18 platform, free fatty acids, bile acids, and metabolites of intermediate polarity using reverse chromatography with a T3 UPLC column (C18 chromatography) and MS analyses in the negative ion mode.

Quality control was performed using previously described methods [21]: Metabolite features with a signal-to-noise ratio <10 were considered unquantifiable and excluded, as were features with undetectable/missing levels for >10% of the samples. All remaining missing values were imputed with the median peak intensity for that feature across the whole population. Features with a coefficient of variance in the quality-control samples

greater than 25% were excluded to ensure good technical reproducibility. The mean and interquartile range percentage coefficient of variance of the remaining features was 13.1% (8.7%, 17.1%). Features were indexed by their mass-to-charge ratio and retention time, and metabolite identities were confirmed using known standards. Metabolite features were analyzed as measured LC-MS peak areas, and were log-transformed and *pareto* scaled prior to analysis. This data processing pipeline has previously been utilized for a number of peer-reviewed publications [17,21-23].

Of 18,064 measured metabolite features, 8,185 passed quality control and data processing procedures (**e-Methods**). Of these, 574 could be reliably annotated to known metabolites by matching measured mass to charge ratios (m/z) and measured retention times (rt) with authentic reference standards. Reference compounds were spiked into biological samples to mitigate any matrix effects. MS/MS data were not acquired during the profiling analyses in order to facilitate acquisition of sufficient numbers of data points across peaks for precise quantitation. To confirm IDs during analytical runs, synthetic mixtures of standards as well as pooled QC samples were analyzed to confirm RT and m/z matches.

To reflect that fact that only a small proportion of metabolites could be annotated, both known and unknown metabolite features were included in the analysis. This allowed the capture of all relationships between the features with no missing links, which would be inevitable if only named metabolites were included, thereby reflecting the whole metabolome. Furthermore, annotation of metabolites is ongoing; metabolites identified as

important in this analysis will be prioritized for annotation and available for future analyses in this population.

.

Network Identification and association with Lung Function Traits

Weighted gene co-expression network analysis (WGCNA) [24] was used to identify transcriptomic and metabolomic networks based on correlation patterns. The correlation matrix quantifies interconnectedness between features (genes and metabolite features) and assigns them to co-expression modules. Highly correlated modules were then merged using a cut height (i.e. the Euclidean distance between clusters) of 0.75 for the transcriptomic data and 0.70 for the metabolomic data; these cut-heights were chosen using an iterative process to identify an optimal number of adequately sized modules for analysis. Features that do not show high enough co-expression metrics with any module are excluded from further analysis.

Modules were summarized by eigenvector (based on the first principal component of each module) for each participant. Associations between the modules with lung function traits were computed using multivariate regression models with adjustment for potential confounders (age, sex, height, weight and treatment regime (regular use to control chronic symptoms versus sporadic use for acute symptoms)). For the modules nominally significantly ($p \leq 0.05$) associated with at least one lung function phenotype, the 'hubs' were identified. Hubs are the features that are most highly connected within a module,

and therefore driving module formation. WGCNA computes a module-membership value and associated p-value for each feature within a module, which is a measure of how connected or co-expressed that feature is with others within the same module. Features with a module-membership p-value that retained significance after Bonferroni-correction were considered to be hubs.

Pathway Enrichment Analysis

For the transcriptomic data, pathway enrichment analysis was performed using the g:GOSt tool within the g:Profiler web server (http://biit.cs.ut.ee/gprofiler/). Metabolomic pathway analysis was performed using MetaboAnalyst v.3.0 (http://www.metaboanalyst.ca/). Metabolomic pathway analysis was limited to those metabolites that could be assigned Human Metabolome Database (HMDB) IDs, therefore it was utilized here as a hypothesis-generating tool (for full details see **e-Methods**).

Integrated Omics Analysis

Relationships between the WGCNA generated transcriptomic and metabolomic modules were explored by computing the correlation between the eigenvector of the two sets of modules. Constituent hub features of correlated pairs were submitted for integrated pathway analysis using IMPaLA: Integrated Molecular Pathway Level Analysis (http://impala.molgen.mpg.de/) to identify pathways simultaneously dysregulated on both a transcriptional and metabolic level. **(e-Methods**). These pathways were then prioritized for further follow up.

Replication

Replication of the most interesting findings was performed in a comparable childhood cohort of mild to moderate asthmatics who were phenotyped in the same way as the Costa Rica Cohort; The Childhood Asthma Management Program population (CAMP). CAMP is a multi-center, randomized, double-masked, clinical trial designed to determine the long-term (~16.5 years of follow-up) effects of three inhaled treatments for mild to moderate asthma in children aged 5 to 12 at baseline: placebo, nedocromil, or budesonide[25] A follow-up study to the primary trial extracted blood samples from 620 CAMP subjects at early adulthood (after trial completion) for gene expression profiling using Illumina HumanHT chip and the same protocols and methods as the Costa Rica study. Genome-wide SNP genotyping was as performed by Illumina, Inc. on the HumanHap550v3 BeadChip was also available for these children [26]. A total of 207 children additionally had metabolomics profiling conducted on the same blood samples at the Broad Institute. Metabolomic profiling was conducted on serum using the same four platforms and the same methodology as the Costa Rica plasma samples (full details in **supplementary methods**). A total of 14587 metabolite features including 324 named metabolites passed the sample QC and data processing pipeline as applied to the Costa Rica Metabolomics data. Two-hundred and twenty two named metabolites were common to the CAMP and Costa Rica populations. Each child's parent or guardian signed a consent statement approved by the clinic's institutional review board (IRB). The clinics also obtained each child's assent.

RESULTS

Baseline Characteristics

Of 1165 asthmatic children, 328 had RNA available for transcriptomic profiling and sufficient plasma for metabolomic profiling. Initial clustering based on the transcriptomic data profiles identified three outliers who formed a distinct cluster that could be separated from the rest of the population using a cut-height of 0.9 (**e-Figure 1**). No metabolomic outliers were detected, and in total 325 children were included in all analyses (**e-Figure**

2). All children were Hispanic/Latino; 139 (42.8%) were female; mean age was 9.1 years (range: 4.5-13.3); and 90% reported asthma controller or reliever treatment (**Table 1**). Although distinct, there was significant correlation between the measured lung function phenotypes (**e-Table 1).**

The 325 children with omics profiling data were representative of the entire population of 1165 children (**e-Table 2**).

Transcriptomic Network Analysis

WGCNA clustered the 25,060 gene probes into 39 different modules. After merging highly correlated modules using a cut height of 0.75 (**e-Figure 3**) a total of eight modules, spanning 19,581 genes, were characterized (**e-Table 3**). A nominal significance threshold of $p \leq 0.05$ was utilized to capture all potentially biologically interesting relationships. Six modules were significantly correlated with at least one measure of lung function (**e-Figure 4**). Four of these modules retained significance at a 95% confidence interval after adjustment for age, weight, height, gender and asthma treatment (**Table 2**).

Lung function-associated modules were enriched for distinct groups of defined biological processes (**Table 3**): the dark green and plum modules for immune processes (renamed the “adaptive immunity” and “innate immunity” modules); the dark grey for cell cycle and mitotic processes (“cell cycle” module); the dark olive-green for asthma-related microRNA targets sites (“asthma microRNAs” module; g:Profiler allows functional interpretation of gene lists in the context of computationally predicted microRNA target

sites from the MicroCosm database); the pink for processes relating to translation ("translational" module); and the yellow for multiple transcriptional and metabolic processes ("transcriptional" module) (**Table 3**).

Metabolomic Network Analysis

WGCNA clustered 8,185 metabolite features (including 574 annotated metabolites) into 44 different modules. After merging (cut height 0.7; **E-Figure 5**), there were eight modules including 7,473 metabolite features (excluding the grey module) (**E-Table 4**). Six modules were correlated at a nominal significance level of 95% with at least one measure of lung function (**e-Figure 6**); and five were robust to confounder adjustment ($p \leq 0.05$) (**Table 4**).

All pathways with a nominally significant p-value <0.1 are reported for hypothesis-generating purposes. The medium-purple module (renamed the "lipid" module) was enriched for glycosylphosphatidylinositol (GPI)-anchor biosynthesis (p=0.034), sphingolipid metabolism (p=0.061), glycerolipid metabolism (p=0.077), and glycerophospholipid metabolism (p=0.093). The tan module (renamed the "purine" module) was enriched for caffeine (p=0.017) and purine metabolism (p=0.075) (**Table 5**).

Integrated Omics Analysis

There were ten nominally significant ($p \leq 0.05$) metabolite-gene module associations based on their eigenvector (**Figure 1**). Of these, seven were between modules that both correlated with lung function phenotypes.

Integrated analyses of the "asthma micoRNAs" transcriptomic and "lipid" metabolomic modules; which correlated with FEV_1 and FEV_1/FVC ratio, identified 186 pathways including 51 pathways robust to false discovery rate correction (top pathways in **Table 6**, full list in **E-Table 5**), the majority were for the dark olive-green module, most of which relate to immune regulation. This integrative analysis extends upon the findings of the metabolomic analysis, where sphingolipid, glycerolipid and glycerophospholipid metabolism were all enriched, by identifying the upstream genes including *ORMDL3, , SLC1A2* and *AKR1B1* driving the dysregulation of these pathways.

The "asthma microRNAs" module was also correlated with the red "Ubiquinone" metabolite module and both associated with FEV_1/FVC ratio (**Table 6**); integration of these modules identified a large number of pathways involved in cellular respiration. However, none were significant after correction for multiple testing.
Cellular respiration was also captured by the integration of the "translational" gene module and the red "Ubiquinone" metabolite module, which also both associated with FEV_1/FVC ratio

ORMDL3

ORMDL3 represents one of the most interesting genes in childhood asthma [27]. Given its role as a hub gene in the asthma micoRNAs" transcriptomic module, we utilized genotype data available for the Costa Rica subjects, to identify SNPs that may influence the expression of this gene and which also associate with the lipid metabolite module. Three relevant SNP were available in the Costa Rica population rs2872507, rs7216389 and rs8079416. While all three associated with at least one of the lipids in the lipid metabolite module, the strongest results were observed for rs8079416 which was associated with 165 (31%) of the 537 metabolites comprising the module, as well as with the module eigenvector (p=0.024), with an increasing number of T alleles associated with lower levels of these lipids (**e-Figure 7**). Importantly, the T allele was also strongly associated with decreased expression of *ORMDL3* in this population (p=0.0006 in an additive model). Providing a further link between *ORMDL3* and dysregulation of lipid metabolism.

Replication of the Integrated Analysis

In order to explore the validity of the nominally significant integrative findings a replication analysis was conducted, focusing on the relationship between the lung function associated "asthma microRNAs" transcriptomic module and the "lipid" metabolomic module.

Whole genome transcriptomic profiling and metabolomic data were available on blood samples from 207 participants from the Childhood Asthma Management Program (CAMP) population; a randomized clinical trial designed to determine the long-term

effects of inhaled treatments for mild to moderate asthma in children [28] (see **e-methods and e-Table 6**).

A summary score based on the first principal component was generated utilizing the expression levels in CAMP of 3926 hub genes from the Costa Rica "asthma microRNAs" module. A significant association was observed between this score and FEV_1 BDR in a regression model adjusting for age, sex, height, weight and treatment group (p= 0.027). Metabolomic profiling information was available for 564 named metabolites; including 30 of the named metabolites identified in the lipid metabolomic module. A score based on the first principal component of these 30 metabolites had a borderline significant association with baseline FEV_1 (p=0.073). Furthermore, there was evidence of a correlation between the transcriptomic and metabolomics modules (r=0.13, p=0.065). Finally, replication of the finding for the *ORMDL3* SNP rs8079416 was attempted in this population. It was again shown that the T allele was strongly associated with increased expression of *ORMDL3* ($p=5.2x10^{-10}$), and with four of the investigated lipids. Thereby providing a measure of replication of the Costa Rica metabolomics, transcriptomic and integrative-omic findings.

DISCUSSION

This study represents one of the first integrative-omics studies of children with asthma. A network approach was taken to identify and integrate modules of highly co-regulated genes and metabolites that correlated with lung function metrics. Specifically, metrics relating to $FEV_{1,}$ and FEV_1/FVC ratio, which associate with childhood asthma severity and have been shown to be predictive of future lung function [16]. Interrogation of the biological pathways and processes underlying these modules allowed the exploration of asthmatic lung function mechanisms. These findings strengthen the evidence for the role of sphingolipids, lipids and fatty acids, and demonstrate the potential of network based methods for integrating large-scale omic datasets.

The transcriptomic network analysis revealed six gene-modules that associated with lung function in terms of airway obstruction, airway responsiveness to methacholine, and BDR. Pathway analysis determined these modules were enriched for distinct processes with biologically plausible relationships with lung function, including adaptive and innate immunity.

Mitosis and processes relating to cell division characterized the "cell cycle" transcriptomic module. Interestingly, a previous WGCNA study of gene expression data from asthmatic airway epithelial cells also identified a "mitosis" module characterized by *PTTG1, BIRC5, NCAPG, CDCA2, FANCI* [29]. All these genes were hubs within our "cell cycle" module. Dysregulation of cell division can alter the composition of the airway epithelium, initiate airway remodeling and lead to chronic airway obstruction and more severe asthma [30]. Intriguingly, the "cell cycle" module was significantly associated with FEV_1/FVC ratio pre and post-bronchodilator, but not with FEV_1. It has been shown that children with asthma can have an abnormal FEV_1/FVC ratio despite a normal FEV_1 and FVC; this condition, known as dysanapsis, occurs when growth in lung volume and airway length outpaces the increase in airway caliber [31]. This may explain the association with a module enriched for dysregulated cell cycle processes in children.

The "asthma microRNAs" gene module, based on microRNA target sits, was enriched for seven microRNA regulatory motifs, which have previously been implicated in lung

function and asthma severity: hsa-miRNA-339-5p has been shown to be differentially expressed in the blood of asthmatics [32] and serum miR-331-3p has been associated with FEV_1/FVC ratio in childhood asthma [33]. Additionally, expression of miR-874 has been shown to be upregulated in patients with allergic rhinitis [34], while miR-874 and miR-423 have both been demonstrated to be upregulated in response to antigens in mouse models [35]. Crucially, in an independent population a summary score based on the expression levels of the genes of this module was again associated with FEV_1 BDR.

Analysis of the metabolomic profiles from the same children at the same time-point also identified six modules associated with metrics of FEV_1 and FEV_1/FVC ratio. The "lipid" metabolite module is of particular interest, as lipid mediators influence asthmatic airway inflammation, in which the conversion of arachidonic acid in membrane phospholipids to eicosanoids, leukotrienes and prostaglandins results in bronchoconstriction and inflammation [36]. It has also been shown that essential omega n-3 fatty acids in foods and oily fish such as eicosapentanoic acid and docosapentanoic acid are capable of displacing arachidonic acid from the cell membrane and promote resolution of inflammation and dampening of airway hyper-responsiveness [37,38]. Both GPI and glycerophospholipids are key constituents of cell membranes, and the identification of these pathways in a metabolite module associated with FEV_1 further highlights the importance of fatty acids and lipid mediators in the pathogenesis of asthma.

Sphingolipids have also been implicated in asthma pathogenesis [39]. *ORMDL3*; one of the most replicated genes for childhood asthma, is thought to exert its effect through the

sphingolipid metabolism pathway via negative regulation of serine palmitoyltransferase (*SPT*), which catalyzes the initial step of sphingolipid biosynthesis [27]. In mouse models, knockout of *SPT* has been shown to result in alterations in *de novo* lung epithelial tissue sphingolipid biosynthesis and an increase in inflammation and airway hyper-responsiveness [40]. However, direct evidence of an association in human studies is lacking.

In the integrated-omics analysis the "lipid" metabolite module correlated with the "asthma microRNAs" transcriptomic module that included *ORMDL3,* providing evidence for a mechanistic connection between *ORMDL3*, microRNA regulatory motifs, and sphingolipid metabolism in asthma. Furthermore, we were able to demonstrate a SNP within the 17q21 locus; rs8079416 which has previously been associated with asthma [41] was significantly associated with both the expression of *ORMDL3* and with the lipid metabolomics module and its constituent metabolites. The "asthma microRNAs" module also encompassed several other asthma genes which, like *ORMDL3*, map to the 17q21 locus including *CCL21* and *GSDMB*. This suggests variants in this region systematically contribute to the pathogenesis of asthma through a dysregulation of sphingolipid metabolism, potentially due to the binding of these genes to the regulatory microRNAs. Intriguingly, the "asthma microRNAs" module also included *CRISPLD2;* previously identified as an asthma candidate gene due to its role as a regulator of the anti-inflammatory effects of glucocorticoids in the airway smooth muscle [42]. Sphingolipids have been shown to mediate the effects of glucocorticoids [43], further supporting the observed link between the "asthma microRNAs" and "lipid" modules and suggesting a possible connection between

the *CRISPLD2* gene and sphingolipid metabolism. Full replication and validation of these findings was not possible as there was incomplete crossover between the known metabolites measured in Costa Rica and CAMP, and the crossover between the unknown metabolites cannot be computed. However, in the independent CAMP population there was evidence that a lung function associated lipid module correlated with this module in blood samples taken at the same timepoint. This adds weight to the theory that the genes enriched for microRNA targets may be acting on lung function though the dysregulation of lipid metabolism.

To date, relatively few studies have attempted to combine metabolomic and transcriptomic data in any human disease, and no analytical gold-standards exist. However, statistical integration using pathway and network-based approaches has shown promising results [44]. This study demonstrated relationships between transcriptomic and metabolomic modules, which were generated independently and which were independently associated with lung function. The integrative analysis enhanced single-omics analysis and improved pathway recovery; capturing associations that may be missed based on a significance threshold when a single-omics data type is analyzed. The downstream metabolites "anchored" the transcriptome [44], providing a functional readout of the changes taking place at a transcriptional level. This makes these transcriptomic changes more biologically interpretable, and provided mechanistic evidence to support the role of common asthma genes such as *ORMDL3* and *CRISPLD2.*

A major strength of this study included the innovative use of validated [29] network methods to perform–for the first time–integrative-omics in a large well-characterized cohort of asthmatic children, allowing the simultaneous exploration of multiple clinically relevant phenotypic characteristics of lung function. Blood is well suited for future clinical translation; mounting research suggests that plasma is an excellent resource for metabolic profiling in asthma and it is known that gene expression patterns in peripheral blood show systematic changes when asthma exacerbations occur [45]. However, it is difficult to determine how the mixture of different cell types in whole blood may affect gene expression.

There were some other limitations. Both known and unknown metabolites are required to build a complete metabolomic network. In this study, only 574 (7%) of 8185 metabolite features could be annotated to known metabolites; subsequent pathway enrichment analysis was therefore limited, and thus additional dysregulated metabolomic pathways within the modules could not be identified. Although there is likely some redundancy in the metabolite features, we consider the inclusion of all features necessary to generate the truest metabolic network possible. If only named metabolites were included in an unsupervised clustering approach such a WGCNA, multiple connections would be missed leading to metabolites not being assigned to their optimal module. Although nominal significance thresholds were employed to consider module relationships, reflecting the exploratory nature of the analysis, the biologically plausibility of the reported results render them worthy of further exploration.

The samples were not originally collected for metabolomic and transcriptomic profiling; and therefore there is the potential that some of the methods employed may have

impacted on our omic profiles and subsequently on the results. However, the majority of samples were processed within four hours of collection and additionally both the metabolomic and transcriptomic data underwent rigorous QC and processing, which aimed to reduce noise, identify outliers and eliminate systematic bias. There was a lack of extreme phenotypes in this population; nevertheless, a variety of studies have demonstrated that omics profiling can capture phenotypic differences within populations of non-severe asthmatics [10,11,20,29,45]. Children from the central valley of Costa Rica represent a genetically homogenous population, which may limit generalizability, however there is abundant evidence that the results of previous genetic analyses in this population can be replicated in outbred populations from different geographical locations [28,46,47]. Crucially, some of the most intriguing findings could be replicated in an independent cohort. The ability to replicate between studies remains an ongoing issue in metabolomics due to heterogeneity in approach, technology and the fact that no one method is capable of capturing the complete metabolome. Accordingly, the current replication is limited by the fact that information was not available in CAMP for all of the relevant metabolites from Costa Rica and thereforeshould be interpreted with caution. Furthermore the differeing biological media used for metabolomics profiling in the two studies may have damped the ability to replicate [48]. Further replication and functional validation is still necessary, and should considered targeted metabolomics profiling of the most interesting findings to obtain absolute metabolite quantification Such work is necessary before the potential for the clinical translation of these findings into biomarkers for asthma prognosis or endotyping can be considered.

In conclusion, this hypothesis-generating study demonstrates how integrating multiple omics technologies provides a more informative picture of asthmatic lung function biology than a single-omics approach, and suggests that network-based methods represent viable integrative strategies.

FIGURE LEGEND

Figure 1: Correlation between the eigenvectors of the transcriptomic and metabolomics modules

Correlation coefficients are shown for each module-trait pair and the associated p-value in brackets; colors indicate direction of association and darker colors indicate more significant associations

Acknowledgements:

RSK takes full responsibility for this manuscript and its contents. The authors wish to thank the participants of the Genetic Epidemiology of Asthma in Costa Rica Study, and all those involved in conducting the study.

Author Contributions:

RSK assisted with the conception of the project performed the statistical analysis and wrote the initial draft of the manuscript. BLC and JLA assisted with the conception of the project and the statistical analysis and edited the manuscript. KB, YVY, DCC and MJM assisted and advised on the statistical analysis and edited the manuscript. JCC and STW conceptualized the initial Costa Rica study, assisted with the conception of this project and edited the manuscript.

F**unding**

The Genetics of Asthma in Costa Rica Study was supported by grants HL066289 and HL04370 from the U.S. NIH. The metabolomics work was supported by 1R01HL123915-01. Dr. Celedón's contribution was supported by grants HL117191 and HL119952 from the U.S. NIH. The funding source played no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. The corresponding author had full access to all the data in the study and final responsibility for the decision to submit for publication.

References

1. Eder W, Ege MJ, von Mutius E. The Asthma Epidemic. *New England Journal of Medicine.* 2006;355(21):2226-2235.
2. Boulet LP, FitzGerald JM, Reddel HK. The revised 2014 GINA strategy report: opportunities for change. *Curr Opin Pulm Med.* 2015;21(1):1-7.
3. Bisgaard H, Jensen SM, Bønnelykke K. Interaction between Asthma and Lung Function Growth in Early Life. *Am J Respir Crit Care Med.* 2012;185(11):1183-1189.
4. Martinez FD, Morgan WJ, Wright AL, Holberg CJ, Taussig LM, Personnel* TGHMA. Diminished Lung Function as a Predisposing Factor for Wheezing Respiratory Illness in Infants. *New England Journal of Medicine.* 1988;319(17):1112-1117.
5. McGeachie MJ, Yates KP, Zhou X, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *New England Journal of Medicine.* 2016;374(19):1842-1852.
6. Håland G, Carlsen KCL, Sandvik L, et al. Reduced Lung Function at Birth and the Risk of Asthma at 10 Years of Age. *New England Journal of Medicine.* 2006;355(16):1682-1689.
7. Chung KF, Wenzel SE, Brozek JL, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *European Respiratory Journal.* 2014;43(2):343-373.
8. Belgrave DCM, Buchan I, Bishop C, Lowe L, Simpson A, Custovic A. Trajectories of Lung Function during Childhood. *Am J Respir Crit Care Med.* 2014;189(9):1101-1109.
9. Renz H, Conrad M, Brand S, Teich R, Garn H, Pfefferle PI. Allergic diseases, gene–environment interactions. *Allergy.* 2011;66:10-12.
10. McGeachie MJ, Dahlin A, Qiu W, et al. The metabolomics of asthma control: a promising link between genetics and disease. *Immunity, inflammation and disease.* 2015;3(3):224-238.
11. Singh A, Yamamoto M, Kam SHY, et al. Gene-Metabolite Expression in Blood Can Discriminate Allergen-Induced Isolated Early from Dual Asthmatic Responses. *PLoS ONE.* 2013;8(7):e67907.
12. Kelly RS, Dahlin A, McGeachie MJ, et al. Asthma metabolomics and the potential for integrative omics in research and the clinic. *Chest.* 2016.
13. Celedon JC, Soto-Quiros ME, Avila L, et al. Significant linkage to airway responsiveness on chromosome 12q24 in families of children with asthma in Costa Rica. *Hum Genet.* 2007;120(5):691-699.
14. Hunninghake GM, Soto-Quiros ME, Avila L, et al. Sensitization to Ascaris lumbricoides and severity of childhood asthma in Costa Rica. *Journal of Allergy and Clinical Immunology.* 2007;119(3):654-661.
15. Asher MI, Montefort S, Bjorksten B, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema

in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *Lancet.* 2006;368(9537):733-743.
16. den Dekker HT, Sonnenschein-van der Voort AMM, de Jongste JC, et al. Early growth characteristics and the risk of reduced lung function and asthma: A meta-analysis of 25,000 children. *Journal of Allergy and Clinical Immunology.* 2016;137(4):1026-1035.
17. Kelly RS, Virkud Y, Giorgio R, Celedon JC, Weiss ST, Lasky-Su J. Metabolomic profiling of lung function in Costa-Rican children with asthma. *Biochimica et biophysica acta.* 2017.
18. Himes BE, Hunninghake GM, Baurley JW, et al. Genome-wide Association Analysis Identifies PDE4D as an Asthma-Susceptibility Gene. *Am J Hum Genet.* 2009;84(5):581-593.
19. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America.* 2010;107(21):9546-9551.
20. Croteau-Chonka DC, Qiu W, Martinez FD, et al. Gene Expression Profiling in Blood Provides Reproducible Molecular Insights into Asthma Control. *Am J Respir Crit Care Med.* 2017;195(2):179-188.
21. Kelly RS, Croteau-Chonka DC, Dahlin A, et al. Integration of metabolomic and transcriptomic networks in pregnant women reveals biological pathways and predictive signatures associated with preeclampsia. *Metabolomics.* 2016;13(1):7.
22. Mayers JR, Wu C, Clish CB, et al. Elevated circulating branched chain amino acids are an early event in pancreatic adenocarcinoma development. *Nature medicine.* 2014;20(10):1193-1198.
23. Townsend MK, Bao Y, Poole EM, et al. Impact of Pre-analytic Blood Sample Collection Factors on Metabolomics. *Cancer Epidemiol Biomarkers Prev.* 2016;25(5):823-829.
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559-559.
25. The Childhood Asthma Management Program (CAMP): Design, Rationale, and Methods. *Controlled Clinical Trials.* 1999;20(1):91-120.
26. Forno E, Lasky-Su J, Himes B, et al. Genome-wide association study of the age of onset of childhood asthma. *J Allergy Clin Immunol.* 2012;130(1):83-90.e84.
27. Paulenda T, Draber P. The role of ORMDL proteins, guardians of cellular sphingolipids, in asthma. *Allergy.* 2016;71(7):918-930.
28. Hersh CP, Raby BA, Soto-Quiros ME, et al. Comprehensive testing of positionally cloned asthma genes in two populations. *Am J Respir Crit Care Med.* 2007;176(9):849-857.
29. Modena DBD, Bleecker DER, Busse DWW, et al. Gene Expression Correlated to Severe Asthma Characteristics Reveals Heterogeneous Mechanisms of Severe Disease. *Am J Respir Crit Care Med.*0(ja):null.
30. Al-Muhsen S, Johnson JR, Hamid Q. Remodeling in asthma. *J Allergy Clin Immunol.* 2011;128(3):451-462; quiz 463-454.

31. Forno E, Weiner DJ, Mullen J, et al. Obesity and Airway Dysanapsis in Children with and without Asthma. *Am J Respir Crit Care Med.* 2016;195(3):314-323.
32. Pietrusinska M, Pajak A, Gorski P, et al. Preliminary studies: differences in microRNA expression in asthma and chronic obstructive pulmonary disease. *Postepy dermatologii i alergologii.* 2016;33(4):276-280.
33. Kho AT, Sharma S, Davis JS, et al. Circulating MicroRNAs: Association with Lung Function in Asthma. *PLOS ONE.* 2016;11(6):e0157998.
34. Wu G, Yang G, Zhang R, et al. Altered microRNA Expression Profiles of Extracellular Vesicles in Nasal Mucus From Patients With Allergic Rhinitis. *Allergy, Asthma & Immunology Research.* 2015;7(5):449-457.
35. Ha T-Y. The Role of MicroRNAs in Regulatory T Cells and in the Immune Response. *Immune Network.* 2011;11(1):11-41.
36. Dennis EA, Norris PC. Eicosanoid Storm in Infection and Inflammation. *Nature reviews. Immunology.* 2015;15(8):511-523.
37. Calder PC. Omega-3 polyunsaturated fatty acids and inflammatory processes: nutrition or pharmacology? *British Journal of Clinical Pharmacology.* 2013;75(3):645-662.
38. Serhan CN. Pro-resolving lipid mediators are leads for resolution physiology. *Nature.* 2014;510(7503):92-101.
39. Levy BD. Sphingolipids and susceptibility to asthma. *N Engl J Med.* 2013;369(10):976-978.
40. Worgall TS, Veerappan A, Sung B, et al. Impaired sphingolipid synthesis in the respiratory tract induces airway hyperreactivity. *Science translational medicine.* 2013;5(186):186ra167.
41. Toncheva AA, Potaczek DP, Schedel M, et al. Childhood asthma is associated with mutations and gene expression differences of ORMDL genes that can interact. *Allergy.* 2015;70(10):1288-1299.
42. Himes BE, Jiang X, Wagner P, et al. RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. *PLoS One.* 2014;9(6):e99625.
43. Lucki NC, Sewer MB. The Interplay Between Bioactive Sphingolipids and Steroid Hormones. *Steroids.* 2010;75(6):390-399.
44. Cavill R, Jennen D, Kleinjans J, Briedé JJ. Transcriptomic and metabolomic data integration. *Briefings in Bioinformatics.* 2016;17(5):891-901.
45. Bjornsdottir US, Holgate ST, Reddy PS, et al. Pathways activated during human asthma exacerbation as revealed by gene expression patterns in blood. *PloS one.* 2011;6(7):e21902.
46. Torgerson DG, Ampleford EJ, Chiu GY, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nature genetics.* 2011;43(9):887-892.
47. Du R, Litonjua AA, Tantisira KG, et al. Genome-wide association study reveals class I MHC-restricted T cell-associated molecule gene (CRTAM) variants interact with vitamin D levels to affect asthma exacerbations. *J Allergy Clin Immunol.* 2012;129(2):368-373, 373.e361-365.

48. Yu Z, Kastenmüller G, He Y, et al. Differences between Human Plasma and Serum Metabolite Profiles. *PLOS ONE.* 2011;6(7):e21230.