Consistent Multi-Omic Relationships Uncover Molecular Basis of Pediatric Asthma IgE Regulation

Tara Eicher[1,2], Rachel S. Kelly[3,4], John Braisted[1], Jalal K. Siddiqui[5], Juan Celedón[6], Clary Clish[7], Robert Gerszten[4,7,8], Scott T. Weiss[3,4], Michael McGeachie[3,4], Raghu Machiraju[2,9,10], Jessica Lasky-Su[3,4], Ewy A. Mathé[1*]

[1] Division of Preclinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD USA
[2] Department of Computer Science and Engineering, College of Engineering, The Ohio State University, Columbus, OH USA
[3] Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA USA
[4] Harvard Medical School, Boston, MA USA
[5] Comprehensive Cancer Center, The Ohio State University, Columbus, OH USA
[6] Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA USA
[7] Broad Institute, Cambridge, MA USA
[8] Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Boston, MA USA
[9] Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH USA
[10] Department of Pathology, The Ohio State University Wexner Medical Center, Columbus, OH USA
* To whom correspondence should be addressed (ewy.mathe@nih.gov)

Abstract

Serum total immunoglobulin E levels (total IgE) capture the state of the immune system in relation to allergic sensitization. High levels are associated with airway obstruction and poor clinical outcomes in pediatric asthma. Inconsistent patient response to anti-IgE therapies motivates discovery of molecular mechanisms underlying serum IgE level differences in children with asthma. To uncover these mechanisms using complementary metabolomic and transcriptomic data, abundance levels of 529 named metabolites and expression levels of 22,772 genes were measured among children with asthma in the Childhood Asthma Management Program (CAMP, N=564) and the Genetic Epidemiology of Asthma in Costa Rica Study (GACRS, N=309) via the TOPMed initiative. Gene-metabolite associations dependent on IgE were identified within each cohort using multivariate linear models and were interpreted in a biochemical context using network topology, pathway and chemical enrichment, and representation within reactions. A total of 1,617 total IgE-dependent gene-metabolite associations from GACRS and 29,885 from CAMP met significance cutoffs. Of these, glycine and guanidinoacetic acid (GAA) were associated with the most genes in both cohorts, and the associations represented reactions central to glycine, serine, and threonine metabolism and arginine and proline metabolism. Pathway and chemical enrichment analysis further highlighted additional related pathways of interest. The results of this study suggest that GAA may modulate total IgE levels in two independent pediatric asthma cohorts with different characteristics, supporting the use of L-Arginine as a potential therapeutic for asthma exacerbation. Other potentially new targetable pathways are also uncovered.

Key Words: allergic asthma; IgE; metabolome; multi-omics; transcriptomics

1

Introduction

Immunoglobulin E (IgE) is an antibody that binds to mast cells and basophils in response to allergen exposure, contributing to asthma exacerbation in children [1]. Exacerbations result in episodes of labored breathing, wheezing, coughing, and bronchial constriction, and are a major cause of disability resulting in poor quality of life and high health care costs [2]. To mitigate exacerbations caused by IgE, anti-IgE therapeutics such as omalizumab block the binding of IgE to receptors [3]. However, response to anti-IgE therapeutics is inconsistent and depends on multiple factors, including periostin, eosinophil levels in serum [4] and allergic sensitization [5,6]. Therefore, investigating the molecular underpinnings associated with higher serum total IgE may help illuminate biological and chemical pathways associated with clinical response, ultimately informing the best treatment approaches.

To this end, several single-omic studies have elucidated genes and metabolites associated with total IgE levels in both children and adults. These studies implicated, among others, the metabolites valine and tyrosine [7], Interleukin 13 [8,9], inflammatory and defense response, and cytokine pathways [10], and leukocyte, lymphocyte, and mononuclear cell proliferation [11], all of which are broadly associated with primarily a T helper 2 (Th2) inflammatory response.

However, the understanding of pathways associated with total IgE levels in asthma is still incomplete given the context of high inter-individual heterogeneity in demographics, exposure, and genetics [2,12]. Notably, the transcriptome and metabolome capture complementary information relevant to these sources of heterogeneity [13]. Thus, integrating the metabolome with the transcriptome should yield a more holistic picture of the complex metabolic processes underlying allergy responses and total IgE levels in pediatric asthma across heterogeneous groups. However, no studies have evaluated the metabolome and transcriptome in tandem in the context of total IgE. In this study, we evaluated the interplay between the plasma metabolome and transcriptome as it relates to total IgE levels among children with asthma from two independent cohorts: 1) the Childhood Asthma Management Program (CAMP) cohort [14] (N=564); and 2) the Genetic Epidemiology of Asthma in Costa Rica Study (GACRS) [15] (N=309). In each cohort, total IgE-dependent gene-metabolite associations were identified using a linear modeling approach that included an interaction term between gene expression and serum total IgE and adjusted for clinical covariates [16,17]. Associations were validated using a null (data permutation) model, and pathway and chemical class enrichment analyses were performed using a comprehensive resource of metabolite and gene annotations [18,19]. Analyses revealed total IgE-dependent gene-metabolite associations shared between the two cohorts that underlie key metabolic processes relevant to allergic and inflammatory responses. These include the glycine, serine, and threonine metabolism pathway, phospholipid pathways, and possible crosstalk between these pathways and Interleukin-1 signaling.

Materials and Methods

The Childhood Asthma Management Program (CAMP)

CAMP [14], available from ClinicalTrials.gov register NCT00000575, was a multi-center, randomized, double-masked, clinical trial designed to determine the long-term effects of inhaled treatments for mild to moderate asthma in children. From December 1993 to September 1995, CAMP recruited 1,041 children aged 5 to 12 years at baseline with mild to moderate asthma from 8 sites in North America [20]. Treatments were randomized to either nedocromil, budesonide, or the placebo arm, and then followed up over the 5 to 6 years of the trial period. All children completed a protocol which included questionnaires and collection of blood. A follow-up study to the primary

trial included age, sex, height, and weight and extracted blood samples from 620 CAMP subjects at early adulthood (after trial completion) for gene expression profiling when asthma was not exacerbated, with an average age of 16 years at follow-up [21].

The Genetic Epidemiology of Asthma in Costa Rica Study (GACRS)

GACRS [15] recruited a total of 1,165 children aged 6-14 years with mild to moderate asthma from 140 schools across the Central Valley of Costa Rica between February 2001 and August 2008. At enrollment, all children completed a protocol similar to that of the children in CAMP, including blood collection and quantification of serum total IgE at the same timepoint. The questionnaire was a translated version (from English to Spanish) of the one used in the Collaborative Study on the Genetics of Asthma [22]. Unlike CAMP, GACRS was a cross-sectional observational study. Written parental and participating child consent was obtained. The study was approved by the Partners Human Research Committee at Brigham and Women's Hospital (Boston, USA) (Partners Human Research Committee; Protocol#: 2000-P-001130/55) and the Hospital Nacional de Niños (San José, Costa Rica).

Data Collection

The metabolomic samples used in our analysis were collected from plasma in 953 children and the transcriptomic samples from whole blood in 609 children during the follow-up timepoint 4 years later. Serum total IgE level was determined using the UniCAP 250 system and converted to a $\log_{10}$ scale for analysis. The CAMP study was approved by the Institutional Review Board of Partners Healthcare (Partners Human Research Committee; Protocol#: 1999-P-001549/29), by all 8 CAMP clinical centers and by the CAMP Data Coordinating Center. Each child's parent or guardian signed a consent statement, and the clinics also obtained each child's assent.

Metabolomic Profiling

Plasma metabolomic profiling was conducted by the Broad Institute using 4 complementary liquid chromatography mass spectrometry (LC-MS) platforms as part of the Trans-Omics for Precision Medicine (TOPMed) initiative [23], i.e., Reversed-Phase C8 Chromatography/Positive Ion Mode (C8-pos), Reversed-Phase C18 Chromatography/Negative Ion Mode (C18-neg), Hydrophilic Interaction Liquid Chromatography/Positive Ion Mode (HILIC-pos), and Targeted Negative Ion Mode (Amide-neg) [24,25] in 953 children in CAMP and 1,155 children in GACRS. Data processing and quality control (QC) was performed using methods previously described by Kelly et al [26]. Briefly, metabolite features with undetectable/missing levels for >75% of study samples or a coefficient of variance within QC samples greater than 25% were excluded. All missing values were imputed using the k-nearest-neighbor imputation method with $k = 3$. To confirm IDs during analytical runs, authentic standards as well as pooled QC samples were analyzed. Metabolites were analyzed as measured LC-MS peak areas, and log10-transformed and unit-scaled prior to analysis. After these QC and scaling steps, we filtered out metabolites with > 20% minimum-imputed values across samples, resulting in 3,756 metabolites in the GACRS cohort and 3,742 in the CAMP cohort, of which 583 named metabolites were present in both cohorts. Named metabolites with variance at or below the 5th percentile were removed. A total of 553 metabolites passed this threshold in the GACRS cohort and 545 in the CAMP cohort, 529 of which were in common and used in our statistical analyses.

Transcriptomic Profiling

Whole-blood gene expression profiles were generated with 47,009 probes from the Illumina HumanHT-12 v4 Expression BeadChip, all of which passed stringent and commonly used quality control (QC) metrics (i.e., removal of failed arrays, probes with low outlying $\log_2$ intensities (<5), and probes with poor signal-to-noise ratios (95th percentile / 5th percentile) [27]. Expression data were $\log_2$-transformed and quantile-normalized as a single batch using the *lumiT* and *lumiN* functions, respectively, from the R package *lumi* (version 2.22) [28,29]. A standard non-specific variance filter was applied to the expression data using the *nsFilter* function from the R package *genefilter* (version 1.52) [30]. Data were collapsed to a single probe per gene based on the largest interquartile range of expression variance. Genes not annotated with a valid Entrez gene identifier or Human Genome Organization (HUGO) gene symbol or with variance at or below the 5th percentile were removed. A total of 23,722 genes passed this threshold in the GACRS cohort and 23,723 in the CAMP cohort, 22,772 of which were in common and included in our statistical analyses.

Data Pretreatment

Supplementary Figure 1 illustrates the data pretreatment procedure. Only samples that could be run on all platforms and that included serum total IgE levels, weight, height, age, race/ethnicity (CAMP only) and study arm (CAMP only) as well as metabolomic and transcriptomic data were retained. In the CAMP clinical data, a single variable combined race and ethnicity (i.e., "Hispanic" was represented as a racial group). Race/ethnicity was omitted from GACRS analyses because the study population was ethnically homogeneous. Levels of 2-deoxyuridine, reduced glutathione, and indole-3-propanoic acid had variances in GACRS that exceeded the maximum variance of any metabolite in CAMP. Upon further inspection, these outlier values were attributed to 10 samples in GACRS, which were removed as they did not pass the generalized extreme Studentized deviate (ESD) test for outliers with $\alpha = 0.0001$. In total, 309 GACRS and 564 CAMP samples were analyzed.

Identification of Gene-Metabolite Relationships Dependent on Total IgE

Pairwise linear models were used to capture IgE-dependent gene-metabolite relationships using Integration of Omics Data Using Linear Modeling (IntLIM) 2.0 [17], described in Equations 1-2 for GACRS and CAMP, respectively.

$$metab = \beta_0 + \beta_1 gene + \beta_2 IgE + \beta_3 (gene:IgE) + \beta_4 age + \beta_5\ sex + \beta_6 weight + \beta_7 height + \varepsilon \quad \text{(Equation 1, GACRS cohort)}$$

$$metab = \beta_0 + \beta_1 gene + \beta_2 IgE + \beta_3 (gene:IgE) + \beta_4 age + \beta_5\ sex + \beta_6 weight + \beta_7 height + \beta_8 treatment + \beta_9 race/ethnicity + \varepsilon \quad \text{(Equation 2, CAMP cohort)}$$

The significance of the statistical interaction between gene expression and serum IgE *(β_3)* denotes the strength of the total IgE-dependent gene-metabolite relationship. All models are adjusted for age, sex, weight, height, and for CAMP only race/ethnicity and treatment arm. The *P*-value of the β_3 coefficients were corrected for multiple comparisons (Benjamini-Hochberg False Discovery Rate (FDR)) [31]. Models with significant FDR-adjusted *P*-values (< 0.05), β_3 coefficients in > 80th percentile; and model $R^2 > 0.1$ were considered statistically significant in each cohort.

4

Model Validation

An additional model validation, based on randomly permutated data, was performed using IntLIM 2.0 to ensure that statistically significant IgE-dependent gene-metabolite pairs identified within each cohort were robust when validated against a null model (e.g., not uncovered in permuted data). To do this, serum IgE was randomly assigned to samples in 100 individual permutations while the levels of genes, metabolites, and other covariates retained their original assignment. For each permutation, IgE-dependent gene-metabolite association models were applied in the same manner as the original unpermuted data. Associations present in fewer than 10 random permutations (out of 100) were considered non-random and retained as significant.

Functional Analyses

For each cohort, a network was constructed, where metabolites and genes were nodes, and IgE-dependent gene-metabolite associations were the edges. Metabolite hubs were defined as metabolites with at least 10 edges (i.e., having IgE-dependent associations with at least 10 genes). Pathway, metabolite chemical class, and reaction-based functional analyses were carried out using Relational Database of Metabolomic Pathways (RaMP-DB) 2.0 [32]. For each cohort, metabolites and genes from significant IgE-dependent associations were utilized as input pathway and chemical class enrichment. Fisher's tests were calculated for enrichment analyses. We further evaluated each significant gene-metabolite pair in CAMP and GACRS to determine whether it corresponded to a known chemical reaction. Information regarding each reaction was extracted from the RaMP-DB reaction label for Rhea [33] reactions. For the Human Metabolome Database (HMDB) [34], the protein corresponding to the gene in the gene-metabolite pair was found using manual lookup on the National Center for Biotechnology Information website . The protein and metabolite were then used to further query the form of the reaction in HMDB.

Results

Two independent cohorts of children with mild to severe asthma, CAMP (N=564) and GACRS (N=309), were utilized in this study. **Table 1** provides a patient summary for both cohorts after QC and data processing. Notably, serum total IgE value distributions were similar between the two cohorts (**Supplementary Figure 2**). We note that height, and weight differed significantly by *t*-test which was likely driven by the fact that the CAMP population was significantly older at blood collection (*P*-value < 2.2e-16 for all).

5

	CAMP	GACRS
Total with all omics	564	309
Age, years – mean (sd)	12.7 (2.1)	9.1 (1.7)
Sex	352 Males, 212 Females	179 Males, 130 Females
Race/Ethnicity	388 Caucasian	309 Costa Rican
	85 African American	
	53 Hispanic	
	38 Other	
Treatment	159 Budesonide	NA
	174 Nedocromil	
	231 Placebo	
Height, cm – mean (sd)	154.8 (13.3)	131.7 (11.4)
Weight, kg – mean (sd)	52.3 (17.0)	32.4 (11.4)
Geometric mean (sd) total IgE level	2.5 (0.7)	2.5 (0.7)

Table 1. Characteristics of the CAMP and GACRS cohort participants.

Figure 1 illustrates our study workflow. Plasma metabolomic and transcriptomic profiles from the same samples were generated in both cohorts, resulting in 22,772 gene and 529 named metabolite measurements in each cohort. The measured metabolites consisted primarily of lipids and lipid-like molecules, especially glycerophospholipids (GPL), glycerolipids, and fatty acyls; however, other represented classes of molecules included organic acids and derivatives and organoheterocyclic compounds (**Supplementary Figure 3**). IgE-dependent gene-metabolite associations were elucidated using IntLIM 2.0 [17], which returns coefficients that describe the statistical interaction of genes and IgE levels (see Methods). A positive coefficient implies that the relationship between a gene and metabolite increases as IgE levels increase. Metabolites and genes from statistically significant associations were then subject to functional analyses to identify enriched pathways and chemical classes. Further, we evaluated whether significant metabolite-gene associations were supported by known reactions.
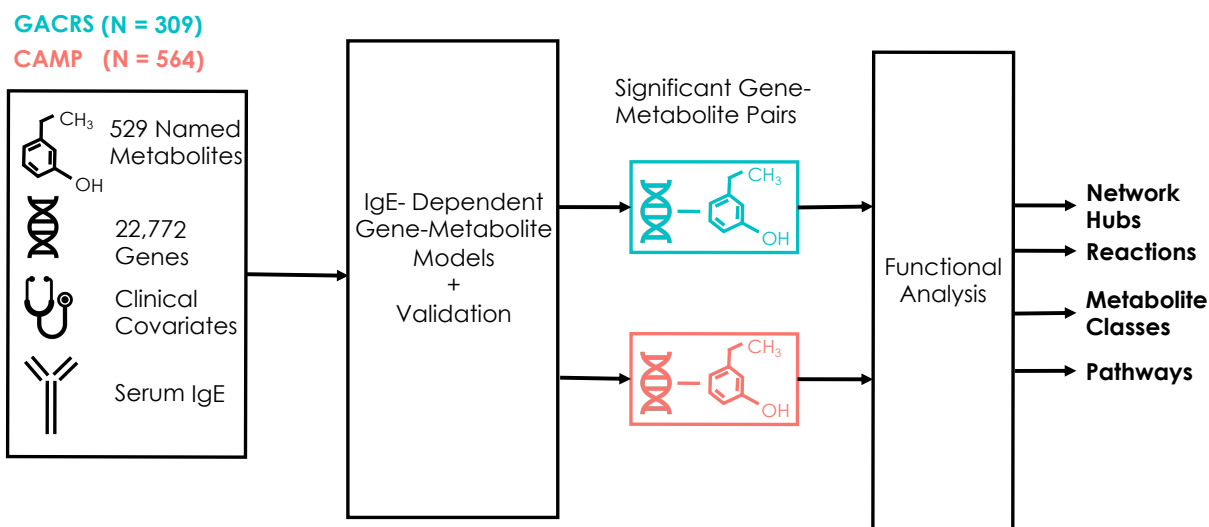
Figure 1. Study workflow. Measurements of 22,772 genes and 529 named metabolites from two independent cohorts, GACRS and CAMP, were used to identify total IgE-dependent gene-metabolite associations ($|\beta_3|_{perc} > 0.8$, $p < 0.05$, $R^2 > 0.1$). To help interpret these associations, pathway and chemical enrichment analyses were performed on metabolites and genes from these associations and reactions that include both the gene and metabolite in an association were identified.

Total IgE-Dependent Gene-Metabolite Associations in GACRS and CAMP Highlight Role of Guanidinoacetic Acid

For both GACRS and the follow-up timepoint in CAMP, all possible gene-metabolite pairs (12,046,388 total combinations of 529 metabolites and 22,772 genes) were evaluated to identify those that were IgE-dependent. After filtering ($|\beta_3|_{perc} > 0.8$, $p < 0.05$, $R^2 > 0.1$), 29,885 pairs were significant in CAMP, and 1,617 pairs were significant in GACRS after permutation validation **(Supplementary Tables 1, 2, and 3)**. While the overall range of R^2 and the distribution of β_3 did not differ considerably between GACRS and CAMP, CAMP had more significant P-values overall **(Supplementary Figures 4-7)**, which may be due to CAMP having a higher sample size than GACRS. The similarity in R^2 range implies that the learned models fit the data equally well in both cohorts.

From the significant pairs, we identified 136 CAMP and 72 GACRS metabolites involved in multiple IgE-dependent gene associations. Metabolite hubs, defined as metabolites associated with more than 10 genes in both cohorts, are listed in **Table 2**. The hubs identified are glycine, diacetylspermine, guanidinoacetic acid (GAA), Lysophosphatidylcholine (LPC)(18:1), pregnenolone sulfate, and Phosphatidylcholine (PC)(P-34:1)/(O-34:2) and are likely to be involved in determining serum total IgE.

Metabolite Name	Number of Associations in CAMP	Number of Associations in GACRS
Glycine	573	43
Diacetylspermine	385	100
Guanidinoacetic acid (GAA)	333	283
LPC(18:1)	145	19
Pregnenolone sulfate	75	135
PC(P-34:1)/PC(O-34:2)	27	32

Table 2. Number of significant IgE-dependent associations found in the 6 metabolite hubs, defined as metabolites associated with > 10 genes in both CAMP and GACRS.

Further, consistency between the cohorts was observed at the individual gene-metabolite pair level and the individual genes and metabolites contained therein. Specifically, 9 gene-metabolite pairs had IgE-dependent associations in the same direction in both cohorts (**Table 3**).

Metabolite	Gene	CAMP		GACRS	
		Interaction β (95% CI)	FDR	Interaction β (95% CI)	FDR
GAA	PTPN12	0.06 (0.03, 0.08)	<0.01	0.08 (0.04, 0.11)	0.04
	HNRNPH2	0.06 (0.03, 0.08)	<0.01	0.09 (0.05, 0.13)	0.04
	PELI1	0.05 (0.03, 0.07)	<0.01	0.08 (0.04, 0.12)	0.03
	CHMP1B	0.05 (0.03, 0.08)	<0.01	0.08 (0.04, 0.12)	0.03
	C9orf6	0.06 (0.02, 0.09)	0.04	0.12 (0.07, 0.18)	0.03
	RCHY1	0.07 (0.02, 0.11)	0.05	0.13 (0.07, 0.18)	0.02
	VANGL2	0.08 (0.03, 0.12)	0.05	0.17 (0.10, 0.24)	0.02
Hydroxyproline	MGC52282	0.06 (0.03, 0.09)	0.01	0.08 (0.04, 0.12)	0.04
Diacetylspermine	DNAJC11	-0.09 (-0.15, -0.04)	0.04	-0.27 (-0.39, -0.16)	0.01

Table 3. Statistically significant gene-metabolite pairs shared between CAMP and GACRS cohorts. * β – coefficient of the statistical interaction between gene expression level and serum IgE. CI – confidence interval.

Notably, GAA is involved in 7 of these 9 IgE-dependent gene-metabolite associations. The genes associated with GAA are involved in Interleukin-1 signaling (PTPN12 and PELI1), mRNA processing and splicing (HNRNPH2), and DNA damage bypass (RCHY1). Because all these pathways are involved in inflammatory response, the associations suggest possible crosstalk between GAA and inflammatory response. Lastly, we note that a total of 770 genes and 50 metabolites involved in IgE-dependent associations were shared between both GACRS and CAMP. For comparison, 12,221 genes and 172 metabolites were involved in IgE-dependent associations in CAMP while 1,283 genes and 123 metabolites were involved in IgE-dependent associations in GACRS, respectively.

Functional Analysis Reveals Perturbation of Glycine, Serine, and Arginine Metabolism in Serum IgE-dependent Gene-Metabolite Pairs

Significant IgE-dependent pairs in CAMP (29,885) comprise 47 HMBD reactions and 4 Rhea reactions, while those significant in GACRS represented a single HMDB reaction, i.e., each reaction includes both the gene and the metabolite from a significant gene-metabolite pair **(Supplementary Tables 4-5)**. The reaction represented in GACRS pairs (1,617) was the conversion of L-Arginine and glycine into ornithine and GAA, catalyzed by glycine amidinotransferase (GATM). Notably, this reaction was also significant ($|\beta_3|_{perc} > 0.8$, $p < 0.05$, $R^2 > 0.1$) in CAMP but was removed after the permutation testing. Other CAMP reactions related to this single GACRS reaction included conversion of glycine to serine and two reactions belonging to the citric acid cycle, which is downstream of glycine to serine conversion in the glycine, serine, and threonine metabolism and of the arginine and proline metabolism pathways. These reactions included conversion of Adenosine Triphosphate (ATP) to Adenosine Diphosphate (ADP) via creatine and conversion of Acyl Coenzyme A to Coenzyme A. We also note that the conversion of glycine to serine was borderline significant in GACRS (FDR = 0.13), suggesting that it may also be perturbed with serum IgE level in both cohorts.

Pathways

Pathways that included the 12,221 genes and 172 metabolites from significant total IgE-dependent gene-metabolite pairs in CAMP and the 1,283 genes and 123 metabolites in GACRS in both cohorts were also evaluated, with all measured genes and metabolites included as the background. While none of the pathways achieved statistical significance (Benjamini-Hochberg-adjusted P-value < 0.05), we found that the top 50 pathways (sorted by nominal P-value) represented in each cohort included 29 shared pathways **(Figure 2)**. **Supplementary Tables 6-7** include all pathways. Shared pathways included glycine metabolism, creatine pathways, serine metabolism, and amino acid metabolism, further supporting the relationship between glycine and serine metabolism and serum IgE level. Moreover, shared glutathione, creatine, and proline pathways support the relationship between arginine and proline metabolism and serum IgE level.
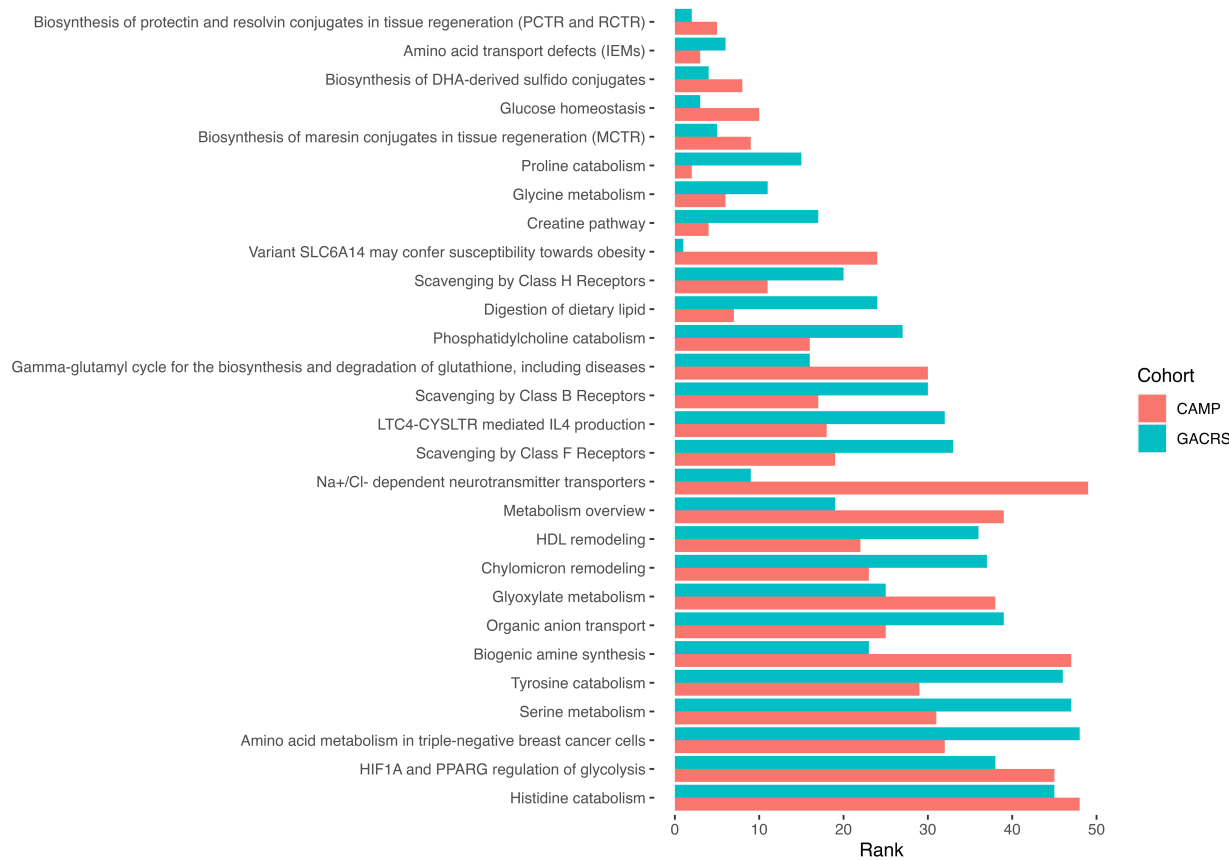
9

Figure 2. Top 50 pathways, sorted by enrichment *P*-value rank, represented in the GACRS and CAMP IgE-dependent gene-metabolite pairs.

Chemical Classes

We further performed class enrichment analysis to provide an overview of the chemical classes of metabolites represented in total IgE-specific gene-metabolite pairs. We found that the significant associations in CAMP were enriched for glycerophospholipids and glycerolipids with Benjamini-Hochberg adjusted *P*-value < 0.05 **(Supplementary Table 8)**. The significant associations in GACRS were enriched for lipids and lipid-like molecules, with glycerolipids significant prior to adjustment **(Supplementary Table 9)**. Notably, glycerophospholipid metabolism is a downstream pathway of glycine, serine, and threonine metabolism that also overlaps with glycerolipid metabolism, suggesting dysregulation of the glycerophospholipid and glycerolipid pathways via glycine, serine, and threonine metabolism could be involved in modulation of total IgE levels.

Discussion

Through integration of metabolomic and transcriptomic profiles, we uncovered associations that implicate a potential role of GAA and glycine in modulating serum total IgE levels of children with asthma. Prior literature shows that with increased GAA (supported by the positive β_3 values), the levels of arginine needed for creatine synthesis decrease, freeing arginine for other functions (such as its role in reducing inflammation) [35]. This presents a possible indirect mechanism by which GAA influences total IgE levels. Additionally, mitochondrial, inflammatory and DNA damage pathways were represented in our uncovered IgE-dependent gene-metabolite pairs, which

10

may reflect activity of creatine (derived from GAA) in inflammatory response via mitochondrial dysfunction and DNA damage. The role of creatine in these processes has previously been studied outside of the context of asthma [36-40].

Associations shared between the two cohorts reflected reactions participating in or downstream of pathways in which GAA and glycine are integral: namely, (1) amino acid metabolism, (2) glycine, serine, and threonine metabolism, and (3) arginine and proline metabolism. Specifically, the conversion of L-arginine and glycine into ornithine and GAA as represented in GACRS is central to glycine, serine, and threonine metabolism as well as arginine and proline metabolism, and the conversion of glycine to serine as represented in CAMP is central to glycine, serine, and threonine metabolism. The Tricarboxylic Acid (TCA) cycle reactions represented in CAMP are downstream of all three pathways. Moreover, amino acid, glycine, serine, creatine, glutathione, and proline pathways were shared between both cohorts, and the glycerophospholipid class was enriched in CAMP. These findings uncover a possible mechanistic basis for the total IgE levels observed in children with asthma via GAA and its pathways, and thus provide opportunities for targeting these mechanisms.

The relationship between these relevant pathways and total IgE level is consistent with previous work showing that that glycine and serine reduce inflammation [41] and that levels of creatine and glycine in plasma decrease after inhalation of budesonide and salbutamol by children with asthma [42], indicating increased glycine and serine metabolism. Further, amino acids have previously been implicated in asthma status and lung function [43,44], and serine protease inhibitors have been shown to mitigate inflammatory cytokine production and to suppress pro-inflammatory genes in a mouse model of asthma [45]. Finally, l-Arginine supplementation has been shown to inhibit nitric oxide producing enzymes (associated with inflammation) via the NOS and arginase pathways in animal and cell culture models [46,47]. L-Arginine supplementation also decreases asthma exacerbations in a subset of clinical trial subjects driven by metabolites that include creatine, taurine, linoleic acid, and α-Glutamylcysteine (Glu-Cys) [48].

The strengths of our analyses include the following: (1) we integrated metabolomics and transcriptomics data in an IgE-dependent context in pediatric asthma, which have not been investigated previously, (2) transcriptomic and metabolomic measurements were performed on the same blood samples in GACRS and on samples taken 4 years later in CAMP, (3) we replicated the results in two separate cohorts with markedly different clinical and population characteristics, and (4) we evaluated relevant biological and chemical pathways, which are complementary and circumnavigate the issue of sparse pathway annotations for metabolites in knowledgebases [49]. One limitation of our analysis is that it is restricted to linear models and thus we are not detecting non-linear relationships which could be biologically relevant. Further, our analysis is correlative and does not reveal causal relationships between metabolites and transcripts or directionality of pathway and reaction dysregulation with increasing serum IgE level. We also note the lack of statistical significance in pathway enrichment analyses, which is likely due to the high number of input metabolites and genes. Despite these limitations, our identified IgE-dependent gene-metabolite pairs were supported by observed chemical reactions and highlight key metabolites, notably GAA, that could modulate serum IgE levels. These observations at a population level offers support for further mechanistic and validation studies that aim to identify potential therapies for asthma exacerbation.

Our integrative analysis of metabolomic and transcriptomic profiles support the potential involvement of GAA, glycine, and related pathways in modulating serum IgE levels of children

with asthma. Notably, our results were consistent across two diverse and independent cohorts, CAMP and GACRS, cohorts that have stark differences in population characteristics.

Conflict of Interest Statement

STW receives royalties from UpToDate Inc. and is on the Board of Histolix Inc. JLS is a scientific advisor to TruDiagnostic Inc, Precion Inc, and Ahara Corp and is on the Metabolomics Society Board. RSK received payment from AAAAI for an invited lecture pertaining to the subject matter of the manuscript.

Data Availability

All data produced are available online at https://topmed.nhlbi.nih.gov/.

Financial Disclosures

Acknowledgments

Author Contribution Statement

EAM and JL-S conceptualized the study; STW, RG, MM, JC, RSK, and JL-S contributed to the design and collection of the TOPMed data described herein. CC performed metabolomic profiling. TE and JKS performed the data pretreatment and statistical analysis. RSK, JL-S, EAM, RM, TE, and JKS conceptualized the analytical workflow. RSK, JL-S, EAM, and TE contributed to data interpretation. JB built a custom version of the RaMP-DB database for the analytical workflow. TE, EAM, RSK, JL-S, and RM drafted the manuscript and contributed to revision of manuscript and editing. All authors had final responsibility for the decision to submit for publication.

References

1.   Rosenwasser LJ. Mechanisms of IgE Inflammation. *Current Allergy and Asthma Reports.* 2011;11(2):178-183.

2.   Carr TF, Bleecker E. Asthma heterogeneity and severity. *World Allergy Organization Journal.* 2016;9:41.

3.   Samitas K, Delimpoura V, Zervas E, Gaga M. Anti-IgE treatment, airway inflammation and remodelling in severe allergic asthma: current knowledge and future perspectives. *European Respiratory Review.* 2015;24(138):594-601.
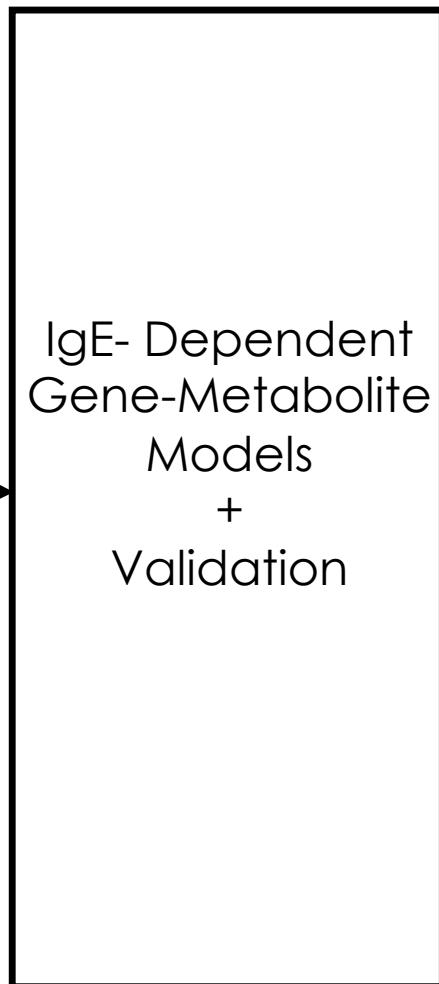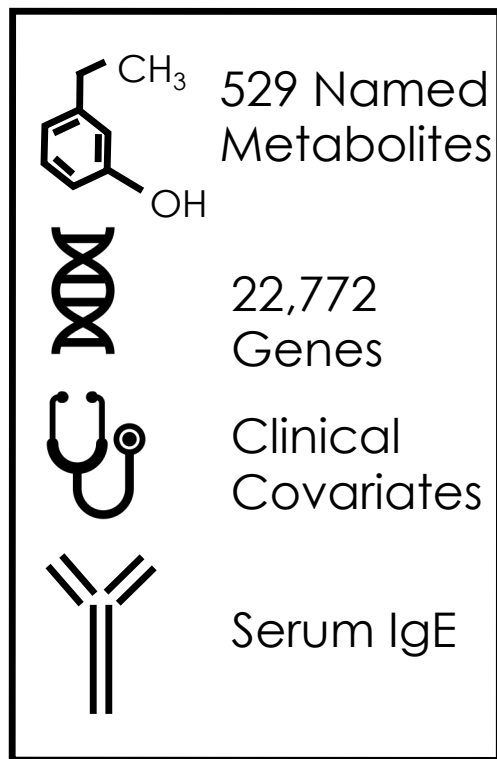
4.   Pajno GB, Castagnoli R, Arasi S, Licari A, Caminiti L, Marseglia GL. Pediatric use of omalizumab for allergic asthma. *Expert Opinion on Biological Therapy.* 2020;20(7):695-703.

5.   Vanik P, Novosad J, Kirchnerova O, Krcmova I, Terl M, Czech Anti-Ig ERc. Effect of individual allergen sensitization on omalizumab treatment outcomes in patients with severe allergic asthma determined using data from the Czech Anti-IgE Registry. *Allergy, Asthma, & Clinical Immunology.* 2020;16:81.

6.   Sheehan WJ, Krouse RZ, Calatroni A, et al. Aeroallergen Sensitization, Serum IgE, and Eosinophilia as Predictors of Response to Omalizumab Therapy During the Fall Season Among Children with Persistent Asthma. *Journal of Allergy and Clinical Immunology: In Practice.* 2020;8(9):3021-3028 e3022.

7.   Chiu C-Y, Cheng M-L, Chiang M-H, Wang C-J, Tsai M-H, Lin G. Metabolomic Analysis Reveals Distinct Profiles in the Plasma and Urine Associated with IgE Reactions in Childhood Asthma. *Journal of Clinical Medicine.* 2020;9:887.

8.   Moffatt MF, Gut IG, Demenais F, et al. A Large-Scale, Consortium-Based Genomewide Association Study of Asthma. *New England Journal of Medicine.* 2010;363:1211-1221.

9.   Granada M, Wilk JB, Tuzova M, et al. A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study. *Journal of Allergy and Clinical Immunology.* 2012;129(3):840-845 e821.

10.   Recto K, Huan T, Lee DH, et al. Transcriptome-wide Association Study of Circulating IgE Levels Identifies Novel Targets for Asthma and Allergic Diseases. *Frontiers in Immunology.* 2023;14.

11.   Virkud YV, Kelly RS, Croteau-Chonka DC, et al. Novel eosinophilic gene expression networks associated with IgE in two distinct asthma populations. *Clinical and Experimental Allergy.* 2018;48:1654-1664.

12.   Siroux V, Bouzigon E. Asthma heterogeneity: the increasing genetic evidence. *The Lancet Respiratory Medicine.* 2019;7:469-471.

13.   Chu SH, Huang M, Kelly RS, et al. Integration of Metabolomic and Other Omics Data in Population-Based Study Designs: An Epidemiological Perspective. *Metabolites.* 2019;9.

14.   Childhood Asthma Management Program Research Group. The childhood asthma management program (CAMP): Design, rationale, and methods. *Controlled Clinical Trials.* 1999;20:91-120.

15.   Kelly RS, Virkud Y, Giorgio R, Celedón JC, Weiss ST, Lasky-Su J. Metabolomic profiling of lung function in Costa-Rican children with asthma. *Biochimica et Biophysica Acta Molecular Basis of Disease.* 2017;1863:1590-1595.

16.   Siddiqui JK, Baskin E, Liu M, et al. IntLIM: Integration using linear models of metabolomics and gene expression data. *BMC Bioinformatics.* 2018;19:81.

17.   Eicher T, Spencer KD, Siddiqui JK, Machiraju R, Mathe EA. IntLIM 2.0: identifying multi-omic relationships dependent on discrete or continuous phenotypic measurements. *Bioinformatics Advances.* 2023;3(1):vbad009.

18.   Zhang B, Hu S, Baskin E, Patt A, Siddiqui JK, Mathé EA. RaMP: A Comprehensive Relational Database of Metabolomics Pathways for Pathway Enrichment Analysis of Genes and Metabolites. *Metabolites.* 2018;22(8).

19.   Braisted J, Patt A, Tindall C, et al. RaMP-DB 2.0: a renovated knowledgebase for deriving biological and chemical insight from genes, proteins, and metabolites. *Bioinfomatics.* 2022;39(1).

20. Childhood Asthma Management Program Research Group. Recruitment of Participants in the Childhood Asthma Management Program (CAMP). I. Description of Methods: Childhood Asthma Management Program Research Group. *Journal of Asthma.* 1999;36:217-237.

21. Croteau-Chonka DC, Qiu W, Martinez FD, et al. Gene expression profiling in blood provides reproducible molecular insights into asthma control. *American Journal of Respiratory and Critical Care Medicine.* 2017;195:179-188.

22. Marsh DG, Maestri NE, Freidhoff LR, et al. A genome-wide search for asthma susceptibility loci in ethnically diverse populations. *Nature Genetics.* 1997;15:389-392.

23. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.* 2021;590(7845):290-299.

24. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature.* 2019;569(7758):655-662.

25. Mazzilli KM, McClain KM, Lipworth L, et al. Identification of 102 Correlations between Serum Metabolites and Habitual Diet in a Metabolomics Study of the Prostate, Lung, Colorectal, and Ovarian Cancer Trial. *The Journal of Nutrition.* 2020;150(4):694-703.

26. Kelly RS, Mendez KM, Huang M, et al. Metabo-Endotypes of Asthma Reveal Differences in Lung Function: Discovery and Validation in Two TOPMed Cohorts. *American Journal of Respiratory and Critical Care Medicine.* 2022;205(3):288-299.

27. Croteau-Chonka DC, Chen Z, Barnes KC, et al. Gene Coexpression Networks in Whole Blood Implicate Multiple Interrelated Molecular Pathways in Obesity in People with Asthma. *Obesity (Silver Spring).* 2018;26(12):1938-1948.

28. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008;24:1547-1548.

29. Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research.* 2008;36:e11.

30. *genefilter: methods for filtering genes from high-throughput experiments* [computer program]. Bioconductor; 2020.

31. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society.* 1995;57:289-300.

32. Braisted J, Patt A, Tindall C, et al. RaMP-DB 2.0: a renovated knowledgebase for deriving biological and chemical insight from metabolites, proteins, and genes. *Bioinformatics.* 2023;39(1).

33. Alcantara R, Axelsen KB, Morgat A, et al. Rhea--a manually curated resource of biochemical reactions. *Nucleic Acids Research.* 2012;40(Database issue):D754-760.

34. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research.* 2018;46(D1):D608-D617.

35. Ostojic SM, Jorga J. Guanidinoacetic acid in human nutrition: Beyond creatine synthesis. *Food Science & Nutrition.* 2023;11(4):1606-1611.

36. Marshall RP, Droste JN, Giessing J, Kreider RB. Role of Creatine Supplementation in Conditions Involving Mitochondrial Dysfunction: A Narrative Review. *Nutrients.* 2022;14(3).

37. Magnani ND, Marchini T, Calabro V, Alvarez S, Evelson P. Role of Mitochondria in the Redox Signaling Network and Its Outcomes in High Impact Inflammatory Syndromes. *Frontiers in Endocrinology Lausanne.* 2020;11:568305.
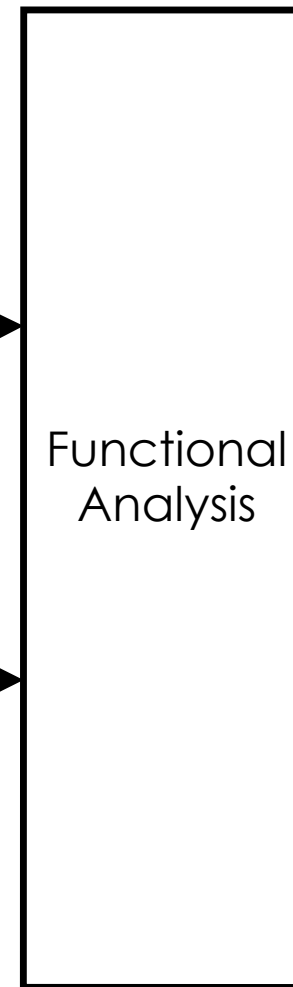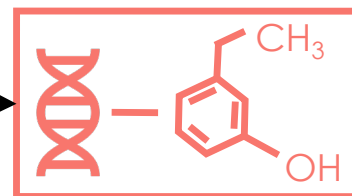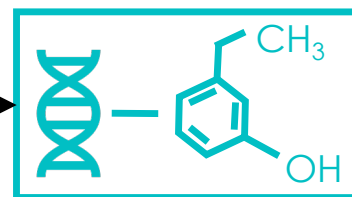
14

38. Rahimi R. Creatine supplementation decreases oxidative DNA damage and lipid peroxidation induced by a single bout of resistance exercise. *The Journal of Strength and Conditioning Research.* 2011;25(12):3448-3455.
39. Valero T. Mitochondrial Biogenesis: Pharmacological Approaches. *Current Pharmaceutical Design.* 2014;20(35):5507-5509.
40. Gowayed MA, Mahmoud SA, El-Sayed Y, Abu-Samra N, Kamel MA. Enhanced mitochondrial biogenesis is associated with the ameliorative action of creatine supplementation in rat soleus and cardiac muscles. *Experimental and Therapeutic Medicine.* 2020;19(1):384-392.
41. Ueland PM, McCann A, Midttun O, Ulvik A. Inflammation, vitamin B6 and related pathways. *Molecular Aspects of Medicine.* 2017;53:10-27.
42. Yang Q-J, Zhang J-P, Zhang J-H, et al. Distinct Metabolic Profile of Inhaled Budesonide and Salbutamol in Asthmatic Children during Acute Exacerbation. *Basic & Clinical Pharmacology & Toxicology.* 2017;120(3):303-311.
43. Kelly RS, McGeachie MJ, Lee-Sarwar KA, et al. Partial least squares discriminant analysis and bayesian networks for metabolomic prediction of childhood asthma. *Metabolites.* 2018;8.
44. Kelly RS, Chawes BL, Blighe K, et al. An Integrative Transcriptomic and Metabolomic Study of Lung Function in Children With Asthma. *Chest.* 2018;154:335-348.
45. Allam V, Waern I, Taha S, Akula S, Wernersson S, Pejler G. Nafamostat has anti-asthmatic effects associated with suppressed pro-inflammatory gene expression, eosinophil infiltration and airway hyperreactivity. *Frontiers in Immunology.* 2023;14:1136780.
46. Bratt JM, Zeki AA, Last JA, Kenyon NJ. Competitive metabolism of L-arginine: arginase as a therapeutic target in asthma. *Journal of Biomedical Research.* 2011;25(5).
47. Scott JA, Maarsingh H, Holguin F, Grasemann H. Arginine Therapy for Lung Diseases. *Frontiers in Pharmacology.* 2021;12.
48. Liao SY, Showalter MR, Linderholm AL, et al. l-Arginine supplementation in severe asthma. *JCI Insight.* 2020;5(13).
49. Barupal DK, Fiehn O. Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Nature Scientific Reports.* 2017;7:14567.