CHEST Online Supplement

An Integrative Transcriptomic and Metabolomic Study of Lung Function in Children With Asthma

Rachel S. Kelly, PhD; Bo L. Chawes, MD, PhD; Kevin Blighe, PhD; Yamini V. Virkud, MD, MPH; Damien C. Croteau-Chonka, PhD; Michael J. McGeachie, PhD; Clary B. Clish, PhD; Kevin Bullock, MS; Juan C. Celedón, MD, DrPH; Scott T. Weiss, MD; and Jessica A. Lasky-Su, ScD

CHEST 2018; 154(2):335-348

 DOI: 10.1016/j.chest.2018.05.038

e-Appendix 1.

e-Methods

Spirometry

Baseline lung function was measured by spirometry using a Survey Tach Spirometer (Warren E. Collins; Braintree, MA) in accordance with the American Thoracic Society recommendations [1]. Prior to assessment, the children were told to withhold short-acting bronchodilators for at least 4 hours. The spirometric maneuvers were conducted with the children seated and wearing a nose clip. The best forced expiratory volume in the 1st second (FEV_1) and corresponding forced vital capacity (FVC) values from 3-5 acceptable flow-volume curves were selected for analysis of baseline FEV_1 and baseline FEV_1/FVC ratio. The measurements were calibrated for gender, age and height according to reference values for Mexican Americans [2].

Bronchodilator Response

After completing baseline spirometry, the children were given 200 µg (2 puffs) of an albuterol pressurized metered-dose inhaler (pMDI) using a spacer device. Spirometry was repeated after 15 min and the bronchodilator response was calculated as the percentage difference in FEV_1 from baseline.

Methacholine Challenge Testing

On a separate visit, a methacholine challenge test [3] was performed in children with an FEV_1 of at least 65% of predicted. The inhalation protocol consisted of five breaths of saline solution followed by one breath of a 1 mg/mL methacholine solution, one and four breaths of a 5 mg/mL methacholine solution, and one breath of a 25 mg/mL methacholine solution provided with a DeVilbiss 646 nebulizer (Sunrise Medical; Carlsbad, CA). Spirometry was performed at baseline and following each subsequent inhalation of methacholine. Airway responsiveness was calculated as the provocative dose of methacholine resulting in a 20% drop in FEV_1 from baseline (PD20). PD20 was log-transformed prior to analysis.

Metabolomic Profiling Details

Metabolite profiles were measured using four distinct liquid chromatography tandem mass spectrometry (LC-MS) methods designed to measure complementary sets of metabolites: polar metabolites measured in the positive ion mode (HILIC-positive), polar metabolites measured in the negative ion mode (HILIC-negative), metabolites of intermediate polarity (e.g. free fatty acids and bile acids; C18-negative), and lipids (C8-positive). Briefly,

hydrophilic interaction liquid chromatography (HILIC) analyses of water soluble metabolites in the positive ionization mode were conducted using an LC-MS system comprised of a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp.; Marlborough, MA) coupled to a Q Exactive hybrid quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Plasma samples (10 µL) were extracted using 90 µL of 74.9:24.9:0.2 v/v/v acetonitrile/methanol/formic acid containing stable isotope-labeled internal standards (valine-d8, Sigma-Aldrich; St. Louis, MO; and phenylalanine-d8, Cambridge Isotope Laboratories; Andover, MA). The samples were centrifuged (10 min, 9,000 x g, 4°C), and the supernatants were injected directly onto a 150 x 2 mm, 3 µm Atlantis HILIC column (Waters; Milford, MA). The column was eluted isocratically at a flow rate of 250 µL/min with 5% mobile phase A (10 mM ammonium formate and 0.1% formic acid in water) for 0.5 minute followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) over 10 minutes. MS analyses were carried out using electrospray ionization in the positive ion mode using full scan analysis over 70-800 m/z at 70,000 resolution and 3 Hz data acquisition rate. Other MS settings were: spray voltage 3.5 kV, capillary temperature 350°C, and heater temperature 300°C. HILIC analyses of water soluble metabolites in the negative ionization mode were conducted using an LC-MS system comprised of an AQUITY UPLC system (Waters; Milford, MA and a 5500 QTRAP mass spectrometer (SCIEX; Framingham, MA). Plasma samples (30 µL) were extracted using 120 µL of 80% methanol containing inosine-15N4, thymine-d4 and glycocholate-d4 internal standards (Cambridge Isotope Laboratories; Andover, MA). The samples were centrifuged (10 min, 9,000 x g, 4°C), and the supernatants were injected directly onto a 150 x 2.0 mm Luna NH2 column (Phenomenex; Torrance, CA). The column was eluted at a flow rate of 400 µL/min with initial conditions of 10% mobile phase A (20 mM ammonium acetate and 20 mM ammonium hydroxide in water) and 90% mobile phase B (10 mM ammonium hydroxide in 75:25 v/v acetonitrile/methanol) followed by a 10 min linear gradient to 100% mobile phase A. MS analyses were carried out using electrospray ionization and selective multiple reaction monitoring scans in the negative ion mode. To create the method, declustering potentials and collision energies were optimized for each metabolite by infusion of reference standards. The ion spray voltage was -4.5 kV and the source temperature was 500°C. Reversed-phase C18 chromatography/negative ion mode MS analyses of free fatty acids and bile acids were conducted using an LC-MS system comprised of a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp.; Marlborough, MA) coupled to a Q Exactive hybrid quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Plasma samples (30 µL) were extracted using 90 uL of methanol containing PGE_2-d4 (Cayman Chemical Co.; Ann Arbor, MI) and centrifuged (10 min, 9,000 x g, 4°C). The samples were injected onto a 150 x 2 mm ACQUITY T3 column (Waters; Milford, MA). The column was eluted isocratically at a

flow rate of 400 µL/min with 60% mobile phase A (0.1% formic acid in water) for 4 minutes followed by a linear gradient to 100% mobile phase B (acetonitrile with 0.1% formic acid) over 8 minutes. MS analyses were carried out in the negative ion mode using electrospray ionization, full scan MS acquisition over 200-550 m/z, and a resolution setting of 70,000. Metabolite identities were confirmed using authentic reference standards. Other MS settings were: spray voltage -3.5 kV, capillary temperature 320°C, and heater temperature 300°C. C8 chromatography analyses of polar and non-polar plasma lipids were conducted using an LC-MS system conducted using an LC-MS system comprised of a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp.; Marlborough, MA) coupled to a Exactive Plus orbitrap mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Plasma samples (10 µL) were extracted using 190 µL of isopropanol containing 1,2-didodecanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids; Alabaster, AL). After centrifugation, supernatants were injected directly onto a 100 x 2.1 mm, 1.7 µm ACQUITY BEH C8 column (Waters; Milford, MA). The column was eluted isocratically with 80% mobile phase A (95:5:0.1 vol/vol/vol 10mM ammonium acetate/methanol/formic acid) for 1 minute followed by a linear gradient to 80% mobile-phase B (99.9:0.1 vol/vol methanol/formic acid) over 2 minutes, a linear gradient to 100% mobile phase B over 7 minutes, then 3 minutes at 100% mobile-phase B. MS analyses were carried out using electrospray ionization in the positive ion mode using full scan analysis over 200–1000 m/z at 70,000 resolution and 3 Hz data acquisition rate. Other MS settings were: spray voltage 3 kV, capillary temperature 300°C, and heater temperature 300°C. Lipid identities were determined based on comparison to reference plasma extracts and were denoted by total number of carbons in the lipid acyl chain(s) and total number of double bonds in the lipid acyl chain(s).

To account for potential batch effect a cassette of two pooled plasma samples was run at intervals of 20 study samples. Each pooled plasma sample cassette included a pooled sample prepared from the study samples for data standardization using a "nearest neighbor" approach as well as a pooled plasma sample for determination of data quality and precision both before and after data standardization.

Raw data from Q Exactive/Exactive Plus MS systems were processed using TraceFinder software (Thermo Fisher Scientific; Waltham, MA) and Progenesis QI (Nonlinear Dynamics; Newcastle upon Tyne, UK). Raw data collected using the HILIC-negative method were processed using MultiQuant 2.1 software (SCIEX; Framingham, MA). Compounds were identified by their exact mass and by matching their retention times to authentic reference standards. In many cases, isomeric compounds were analyzed and in cases where the compound could not be resolved by chromatography, a general name for the compound is

reported (e.g. pentose phosphate for ribulose 5-phosphate/ribose 5-phosphate). Only identified metabolites are included in these analyses.

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/) [4]. A minimum intersect of three between the curated gene lists and the module gene list was specified, the 25,060 genes included in the initial analyses were imputed as a custom background list, gene sets curated via *in silico* means were excluded and Bonferroni correction was applied.

For the metabolomic data, pathway analysis was performed using MetaboAnalyst v.3.0 [5] The hypergeometric test was specified for the over-representation analysis and relative 'betweeness' centrality for the pathway topology analysis. Metabolomic pathway analysis was limited as for each module only those metabolites that could be assigned Human Metabolome Database IDs could be included in the pathway analysis, therefore it was utilized here as a hypothesis-generating tool.

For the integrated pathway analysis using IMPaLA: Integrated Molecular Pathway Level Analysis [6] the hypergeometric distribution is used to assess the significance of pathway overlap; testing 3073 pathways from 11 public databases.

e-Table 1: Correlation between lung function metrics among 325 children

	Gender	Age at recruitment	Height	Weight	BMI	Controller treatment	Baseline FEV1	Baseline FEV_1/ FVC ratio	Post-bronchodilator FEV_1/ FVC ratio	Airway responsiveness measured by methacholine challenge, PD20	$FEV1_1$ bronchodilator response
Gender	1										
Age at recruitment	0.05	1									
Height	−0.02	0.83 *	1								
Weight	−0.01	0.67 *	0.82 *	1							
BMI	0.01	0.40 *	0.49 *	0.89 *	1						
Controller treatment	−0.01	−0.01	0.02	0.06	0.08	1					
Baseline FEV1	0.04	0.88 *	0.99 *	0.82 *	0.49 *	0.03	1				
Baseline FEV_1/FVC ratio	−0.16 *	−0.09	−0.07	−0.06	−0.06	−0.19 *	−0.08	1			
Post-bronchodilator FEV_1/FVC ratio	−0.17 *	−0.11 *	−0.09	−0.11	−0.09	−0.18 *	−0.11	0.83 *	1		
Airway responsiveness measured by methacholine challenge, PD20	−0.08	−0.09	−0.14 *	−0.09	−0.01	−0.04	−0.14 *	0.13	0.07	1	
$FEV1_1$ bronchodilator response	0	−0.02	−0.02	−0.06	−0.09	0.1	−0.01	−0.42 *	−0.12 *	−0.13	1

**Significant ($p<0.05$) correlations*

e-Table 2: Baseline characteristics of the study population included in this study in comparison to the total Genetic epidemiology of asthma in Costa Rica population

Epidemiological and clinical characteristics		Omic population (n=325)	Total population (n=1164)[a]	*p-value*
Gender	**Male *n* [%]**	190 [58.5%]	501 [59.7%]	*0.746*
Age	**yrs *mean* [SD]**	9.1 [1.8]	9.3 [1.9]	*0.208*
Height	**cm *mean* [SD]**	132.1 [11.5]	132.9 [11.9]	*0.271*
Weight	**kg *mean* [SD]**	32.5 [11.3]	33.3 [11.5]	*0.302*
BMI	**kg/m2 *mean* [SD]**	18.1 [3.9]	18.3 [3.7]	*0.436*
Treatment	**Yes *n* [%]**	297 [91.4%]	793 [94.5%]	*0.067*
Baseline FEV1	***mean* [SD]**	1.8 [0.5]	1.8 [0.5]	*0.218*
Baseline FEV_1/FVC ratio	***mean* [SD]**	86.5 [7.7]	83.4 [7.8]	$9.4x10^{-10*}$
post-BD FEV_1/FVC ratio	***mean* [SD]**	89.1 [6.7]	86.5 [6.8]	$8.4x10^{-9*}$
Pd20	***mean* [SD]**	1.7 [2.1]	2.0 [2.7]	*0.093*
FEV_1 Bronchodilator Response	***mean* [SD]**	5.0 [8.7]	5.8 [10.5]	*0.175*

[a]*one subject coded incorrectly for age was excluded from analysis*
**Significant at the 95% confidence interval*

e-Table 3: Transcriptomic modules

Module	n genes
darkolivegreen	7493
yellow	7047
grey	*5479*
pink	4359
cyan	214
darkred	212
darkgreen	113
darkgrey	107
plum1	36

Grey module contains genes that cannot be assigned to any other modules and is excluded from further analysis

e-Table 4: Metabolomic modules

Module	n metabolites
ivory	4517
brown	1208
grey	*712*
mediumpurple3	537
red	449
purple	411
tan	147
lightcyan	111
darkred	93

Grey module contains genes that cannot be assigned to any other modules and is excluded from further analysis

e-Table 5: Pathways jointly enriched by both genes of the dark olive green "asthma microRNAs" module and metabolites of the medium purple "lipid" module

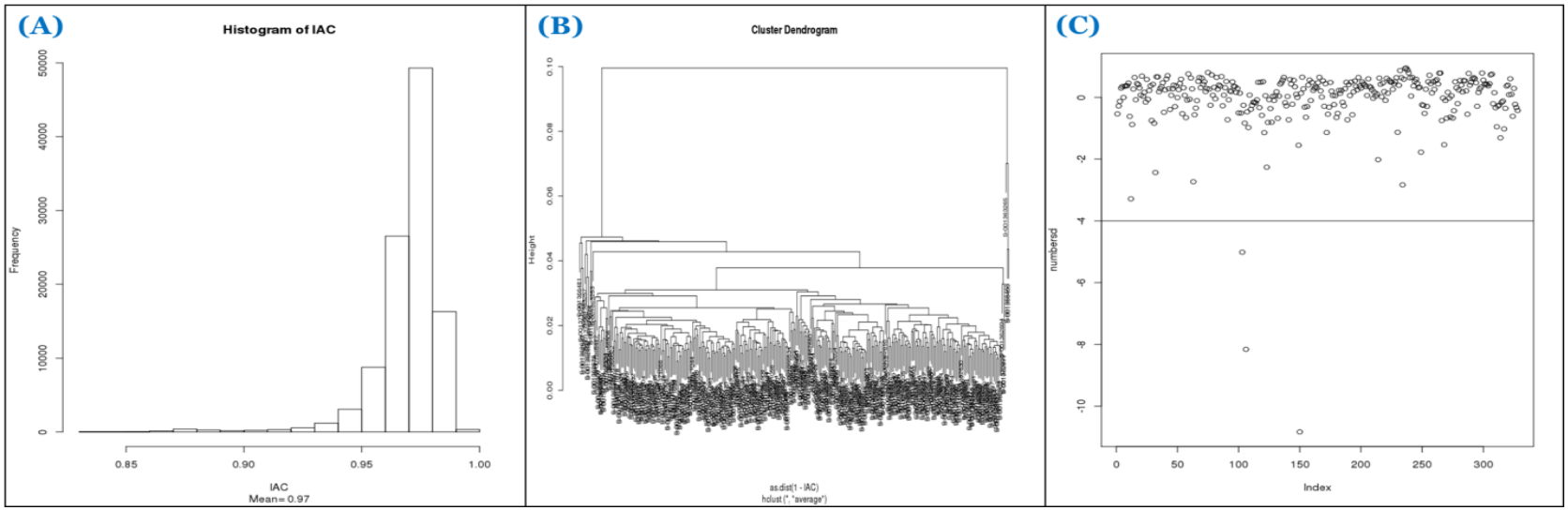
Pathway name	n. over-lapping genes	n. over-lapping metabolites	Joint p-value	Joint q-value
Immune System	409	4	9.38x10-11	2.78x10-7
Membrane Trafficking	152	1	1.46x10-8	2.17x10-5
B cell receptor signaling pathway - Homo sapiens (human)	31	1	5.12x10-8	5.05x10-5
Vesicle-mediated transport	157	1	1.22x10-7	9.00x10-5
Adipocytokine signaling pathway - Homo sapiens (human)	23	2	6.64x10-7	3.27x10-4
Signalling by NGF	106	2	1.03x10-6	4.36x10-4
Adaptive Immune System	168	3	1.42x10-6	4.83x10-4
TNF alpha Signaling Pathway	33	1	1.47x10-6	4.83x10-4
Neurotrophin signaling pathway - Homo sapiens (human)	34	2	1.84x10-6	5.43x10-4
Insulin resistance - Homo sapiens (human)	27	3	2.23x10-6	5.99x10-4
Innate Immune System	276	2	3.88x10-6	8.83x10-4
Chemokine signaling pathway	50	1	4.52x10-6	9.56x10-4
Fc gamma R-mediated phagocytosis - Homo sapiens (human)	32	1	9.57x10-6	1.89x10-3
Lipid storage and perilipins in skeletal muscle	3	2	1.76x10-5	3.26x10-3
Metabolism	415	4	1.96x10-5	3.42x10-3
Chemokine signaling pathway - Homo sapiens (human)	52	1	3.30x10-5	5.43x10-3
Metabolism of proteins	310	2	4.32x10-5	6.72x10-3
Post-translational protein modification	218	2	5.97x10-5	8.84x10-3
CD4 T cell receptor signaling	38	1	6.59x10-5	9.29x10-3
NF-kappa B signaling pathway - Homo sapiens (human)	29	1	8.12x10-5	1.09x10-2
Sphingolipid signaling pathway - Homo sapiens (human)	31	2	9.56x10-5	1.18x10-2
p75 NTR receptor-mediated signalling	30	1	1.77x10-4	2.01x10-2
Cell Cycle	126	1	2.21x10-4	2.42x10-2
Chylomicron-mediated lipid transport	4	2	2.43x10-4	2.48x10-2
Phospholipid metabolism	44	3	2.83x10-4	2.70x10-2
Hemostasis	154	1	3.07x10-4	2.75x10-2
Acyl chain remodeling of DAG and TAG	2	2	3.25x10-4	2.82x10-2
ceramide signaling pathway	12	1	3.91x10-4	3.13x10-2
Glycerophospholipid metabolism	30	3	4.21x10-4	3.19x10-2
Cargo recognition for clathrin-mediated endocytosis	29	1	4.52x10-4	3.28x10-2
GnRH signaling pathway - Homo sapiens (human)	27	1	4.63x10-4	3.28x10-2

AGE-RAGE signaling pathway in diabetic complications - Homo sapiens (human)	18	2	4.67x10-4	3.28x10-2
Phosphatidylinositol phosphate metabolism	29	1	4.77x10-4	3.28x10-2
Clathrin-mediated endocytosis	38	1	5.26x10-4	3.46x10-2
Aldosterone synthesis and secretion - Homo sapiens (human)	26	1	5.45x10-4	3.51x10-2
Lipoprotein metabolism	16	2	5.58x10-4	3.51x10-2
Triacylglyceride Synthesis	4	2	5.90x10-4	3.64x10-2
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	34	2	6.28x10-4	3.79x10-2
GPCR GroupI metabotropic glutamate receptor	11	1	6.56x10-4	3.88x10-2
Glioma - Homo sapiens (human)	20	1	6.85x10-4	3.97x10-2
VEGF	45	1	7.48x10-4	4.17x10-2
Phospholipid Biosynthesis	9	2	7.81x10-4	4.17x10-2
Platelet activation - Homo sapiens (human)	34	1	8.12x10-4	4.21x10-2
CLEC7A (Dectin-1) signaling	15	1	8.31x10-4	4.24x10-2
Fat digestion and absorption - Homo sapiens (human)	3	2	9.06x10-4	4.54x10-2
Visual phototransduction	15	3	1.02x10-3	5.01x10-2
ErbB signaling pathway - Homo sapiens (human)	24	1	1.09x10-3	5.21x10-2
Synthesis of glycosylphosphatidylinositol (GPI)	4	2	1.10x10-3	5.21x10-2
Cell Cycle_ Mitotic	105	1	1.13x10-3	5.21x10-2
Glycerophospholipid biosynthesis	23	3	1.17x10-3	5.26x10-2
NGF signalling via TRKA from the plasma membrane	79	1	1.17x10-3	5.26x10-2

e-Table 6: Characteristics of the CAMP Replication Population

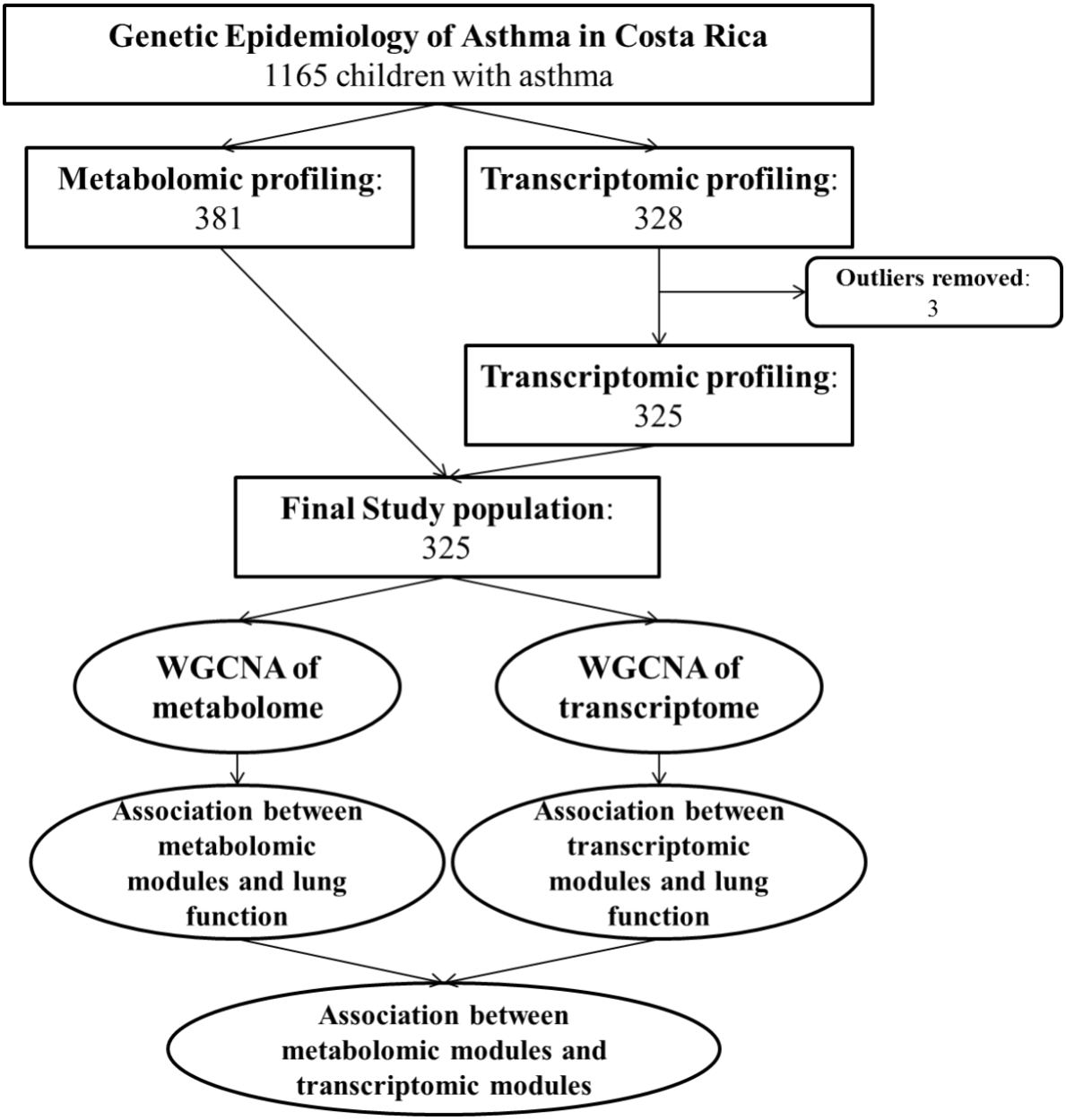
Epidemiological and clinical characteristics		**CAMP population (n=207)**
Gender	**Male *n* [%]**	138 [66.7%]
Age	**yrs *mean* [SD]**	16.9 [2.9]
Treatment	**Placebo *n* [%]**	93 [44.9%]
	Budesonide *n [%]*	54 [26.1%]
	Nedocromil *n [%]*	60 [29.0%]
Baseline FEV1	***mean* [SD]**	3.41 [0.81]

e-Figure 1: Visualization of outliers in the gene expression data



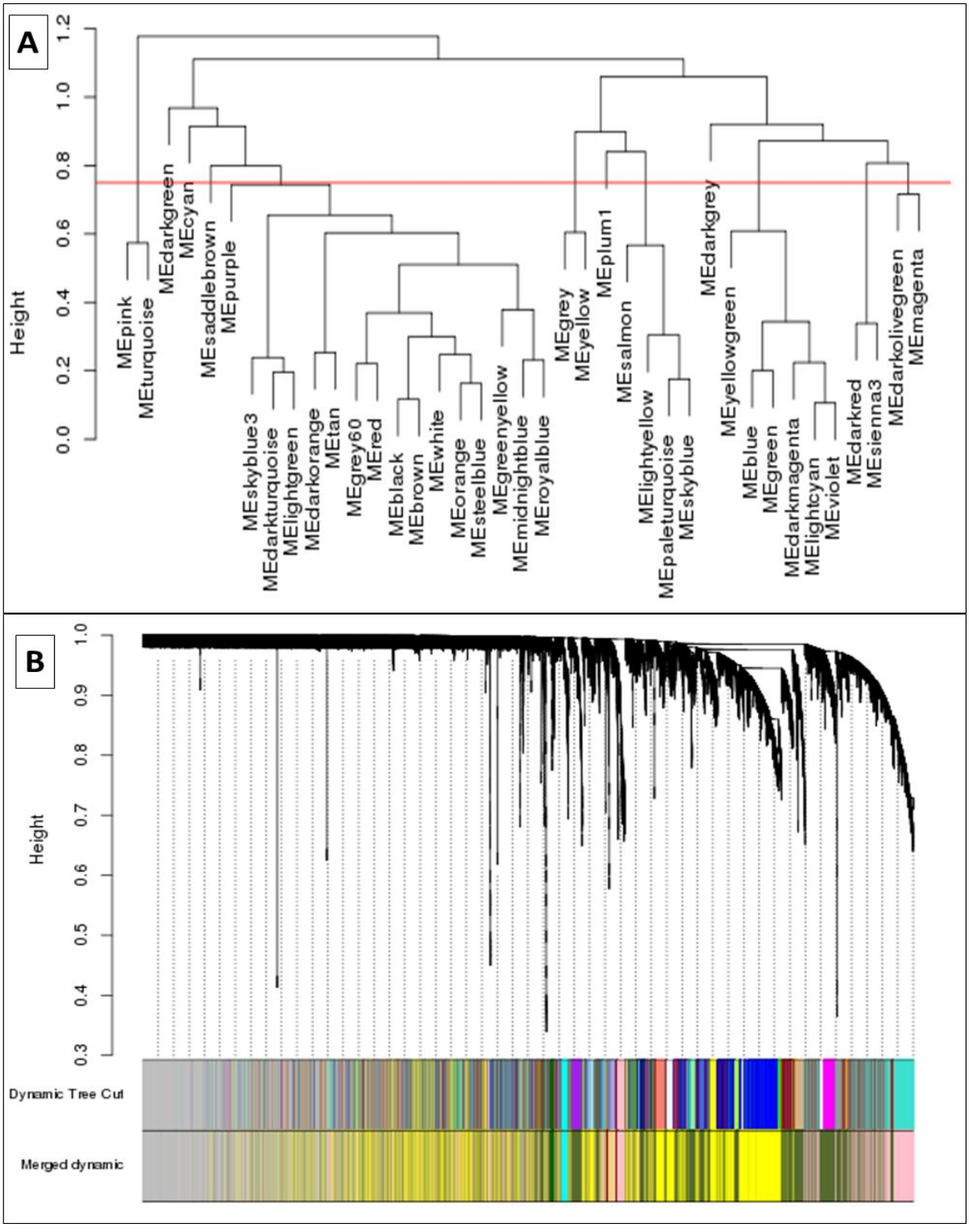
Outlying samples are identified using the inter-array correlation (IAC), which is defined as the Pearson correlation coefficient of the expression levels for a given pair of samples. In the figure below (A) displays a histogram of all IAC values, the tail to the left confirms the presence of outliers (B) presents the relationships between arrays as a dendrogram using average linkage hierarchical clustering with 1-IAC as a distance metric, three outlying non-clustering samples can be seen; (C) shows the distribution of the mean IAC,; the same three outliers are identified, and consequently these three samples were excluded from further analysis

e-Figure 2: Study Schematic



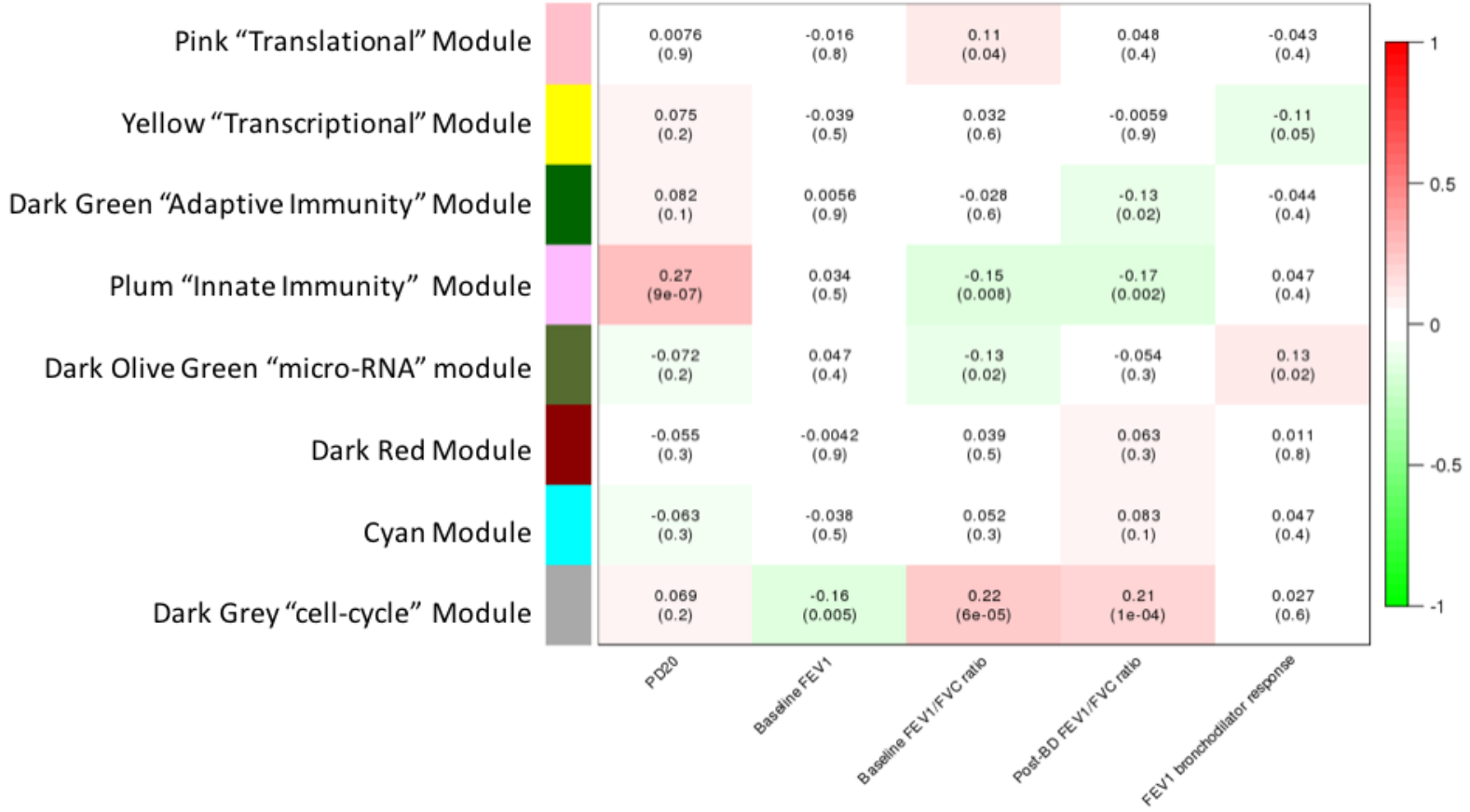
WGCNA-Weighted gene co-expression network analysis; outliers identified using hierarchical clustering

e-Figure 3: Transcriptomic module assignment for 25060 genes in 325 subjects



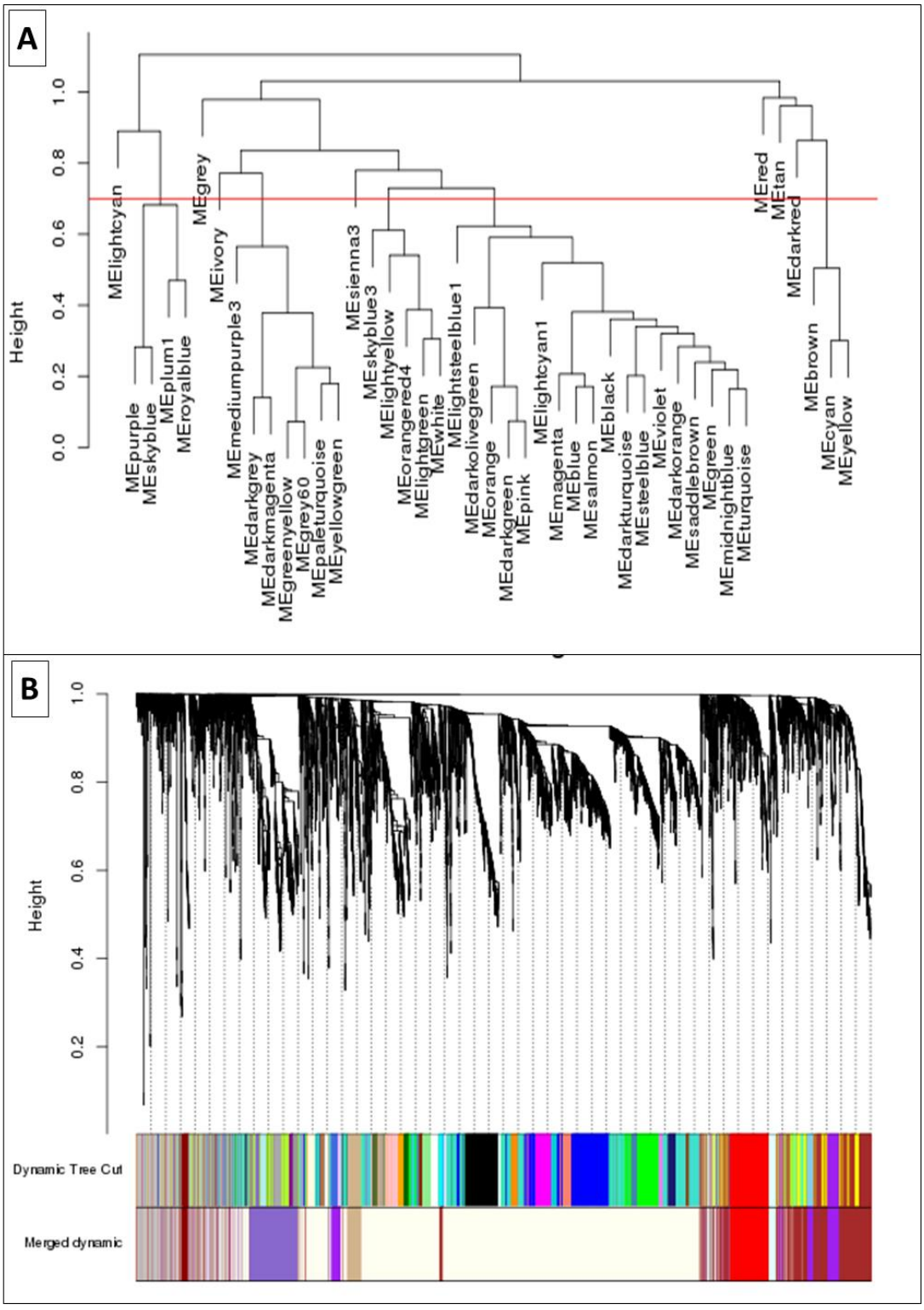
A) Sample Dendrogram and cut height for merging; B) Genes clustering and module assignment before and after merge

e-Figure 4: Correlation between transcriptomic modules and lung function traits



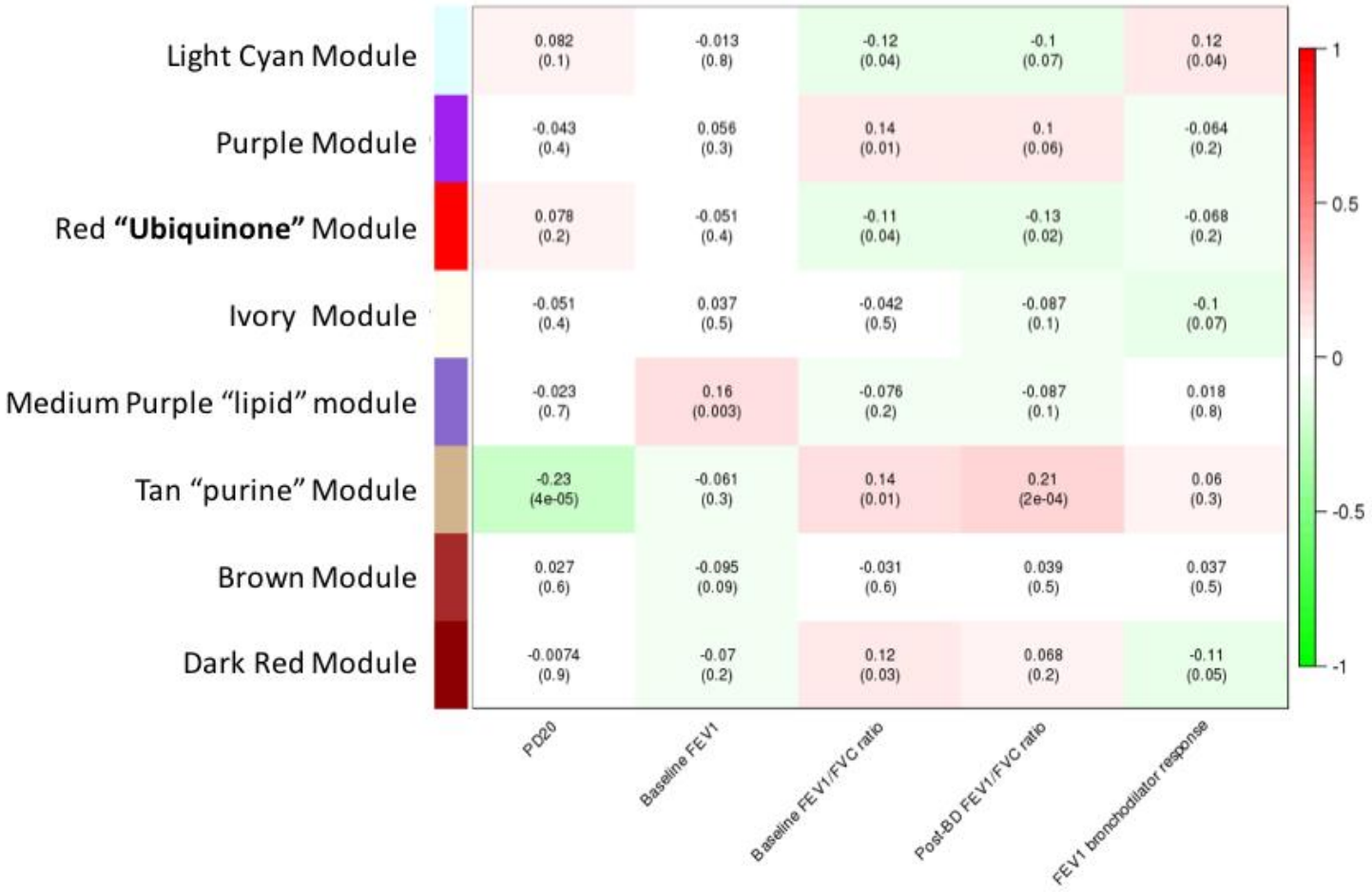
Correlation coefficients are shown for each module-trait pair and the associated p-value in brackets; colors indicate direction of association and darker colors indicate more significant associations
Functional names have been assigned to those modules that associated with lung function

e-Figure 5: Metabolomic module assignment for 8185 metabolite features in 325 subjects



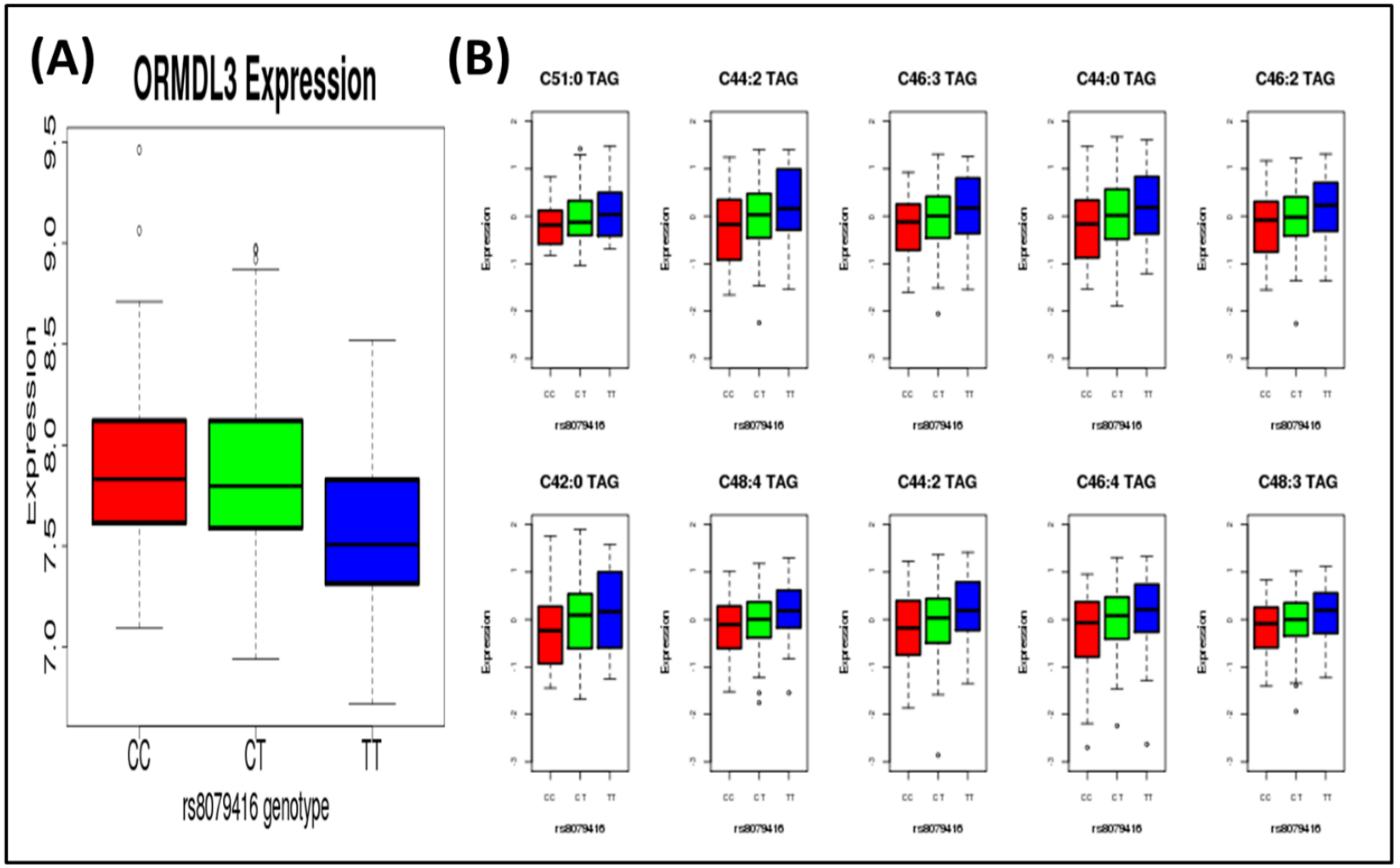
B) Sample Dendrogram and cut height for merging; B) Metabolite clustering and module assignment before and after merge

e-Figure 6: Correlation between metabolite modules with lung function traits



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 association

e-Figure 7: (A) Expression levels of ORMDL3 and (B) Peak Intensity of the top ten most strongly associated metabolites from the medium purple module according to rs8079416 genotype in 246 children from the Genetic Epidemiology of asthma in Costa Rica Study with genotype; gene expression and metabolomic profiling data



Number of children in each genotype strata; CC=75; CT= 131; TT=40
The top ten (of 165) metabolites significantly associated with rs8079416 are shown. All 165 were inversely associated with increasing #C alleles

References

1. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med.* 1995;152(3):1107-1136.
2. HANKINSON JL, ODENCRANTZ JR, FEDAN KB. Spirometric Reference Values from a Sample of the General U.S. Population. *Am J Respir Crit Care Med.* 1999;159(1):179-187.
3. Chatham M, Bleecker ER, Smith PL, Rosenthal RR, Mason P, Norman PS. A Comparison of Histamine, Methacholine, and Exercise Airway Reactivity in Normal and Asthmatic Subjects. *American Review of Respiratory Disease.* 1982;126(2):235-240.
4. Reimand J, Arak T, Adler P, et al. g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* 2016;44(W1):W83-89.
5. Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Research.* 2015.
6. Kamburov A, Cavill R, Ebbels TM, Herwig R, Keun HC. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics (Oxford, England).* 2011;27(20):2917-2918.