Supplementary Material

Urinary metabotype of severe asthma evidences decreased carnitine metabolism independent of oral corticosteroid treatment in the U-BIOPRED Study

Stacey N. Reinke[1,2,†], Shama Naz[1,†], Romanas Chaleckis[1,3], Hector Gallart-Ayala[1], Johan Kolmert[1,4], Nazanin Z. Kermani[5], Angelica Tiotiu[5,6], David I. Broadhurst[2], Anders Lundqvist[7], Henric Olsson[8] Marika Ström[9,10], Åsa M. Wheelock[9,10], Cristina Gómez[1,4], Magnus Ericsson[11], Ana R. Sousa[12], John H. Riley[12], Stewart Bates[12], James Scholfield[13], Matthew Loza[14], Frédéric Baribaud[14], Per S. Bakke[15], Massimo Caruso[16], Pascal Chanez[17], Stephen J. Fowler[18], Thomas Geiser[19], Peter Howarth[13], Ildikó Horváth[20], Norbert Krug[21], Paolo Montuschi[22], Annelie Behndig[23], Florian Singer[24], Jacek Musial[25], Dominick E. Shaw[26], Barbro Dahlén[10], Sile Hu[27], Jessica Lasky-Su[28], Peter J. Sterk[29], Kian Fan Chung[5], Ratko Djukanovic[13], Sven-Erik Dahlén[4,27], Ian M. Adcock[5], *Craig E. Wheelock[1,3,10], on behalf of the U-BIOPRED Study Group#

Affiliations:

1) Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden.
2) Centre for Integrative Metabolomics & Computational Biology, School of Science, Edith Cowan University, Perth, Australia
3) Gunma Initiative for Advanced Research (GIAR), Gunma University, Maebashi, Japan
4) The Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
5) National Heart and Lung Institute, Imperial College, London, U.K.
6) Department of Pulmonology, University Hospital of Nancy, Nancy, France
7) DMPK, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden
8) Translational Science and Experimental Medicine, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden
9) Respiratory Medicine Unit, K2 Department of Medicine Solna and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden
10) Department of Respiratory Medicine and Allergy, Karolinska University Hospital, Stockholm, Sweden
11) Department of Clinical Pharmacology, Huddinge Campus, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden
12) Glaxo Smith Kline, London, U.K.
13) Faculty of Medicine, Southampton University and NIHR Southampton Respiratory Biomedical Research Center, University Hospital Southampton, Southampton, U.K.
14) Janssen Research and Development, High Wycombe, U.K.
15) Institute of Medicine, University of Bergen, Bergen, Norway
16) Department of Biomedical and Biotechnological Sciences and Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy
17) Assistance Publique des Hôpitaux de Marseille, Clinique des Bronches, Allergies et Sommeil, Aix Marseille Université, Marseille, France
18) Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, and

Manchester Academic Health Science Centre and NIHR Biomedical Research Centre, Manchester University Hospitals NHS Foundation Trust, Manchester, U.K.
19) Department of Pulmonary Medicine, University Hospital, University of Bern, Switzerland
20) Department of Pulmonology, Semmelweis University, Budapest, Hungary
21) Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany
22) Pharmacology, Catholic University of the Sacred Heart, Rome, Italy
23) Department of Public Health and Clinical Medicine, Section of Medicine, Umeå University, Umeå, Sweden
24) Division of Paediatric Respiratory Medicine and Allergology, Department of Paediatrics, Inselspital, Bern University Hospital, University of Bern, Switzerland
25) Dept of Medicine, Jagiellonian University Medical College, Krakow, Poland
26) Nottingham NIHR Biomedical Research Centre, University of Nottingham, U.K.
27) Data Science Institute, Imperial College, London, U.K.
28) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
29) Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

†equal contribution

*corresponding author
Craig E Wheelock
Division of Physiological Chemistry 2
Department of Medical Biochemistry and Biophysics
Karolinska Institutet
Solnavägen 9, Biomedicum Quartier 9A
171 77 Stockholm, Sweden
Telephone: +46-852487630
Email: craig.wheelock@ki.se

Methods

Study Subjects and Design

A total of 605 participants from the pan-European U-BIOPRED study (Unbiased Biomarkers for the Prediction of Respiratory Disease outcomes) were included (**Table 1**). Participants from 15 clinical sites were included and asthma was classified according to international guidelines on severe asthma with the following groups; healthy controls (HC, n=100), mild-to-moderate asthmatics (MMA, n=87), non-smoking severe asthmatics (SAns, n=310), and smoking/ex-smoking severe asthmatics (SAs, n=108) [1]. All participants provided a urine sample within 28 days of initial screening (baseline visit); an additional urine sample was provided by 225 SAns and 80 SAs participants at a longitudinal follow-up visit 12-18 months later. A brief overview of baseline and longitudinal clinical and demographic characteristics is shown in (Table 1), with a more detailed baseline description found elsewhere [1]. Ethics approval was obtained from each participating clinical institution and all participants provided written informed consent. U-BIOPRED adhered to standards outlined by the International Conference on Harmonisation and Good Clinical Practice and is registered on ClinicalTrials.gov (identifier: NCT01976767).

Medication use

All MMA subjects were on ≤500 µg inhaled fluticasone equivalents/day (ICS), while all SA subjects received ≥1000 µg fluticasone equivalents/day. Of the severe asthmatics, 50% were prescribed oral corticosteroids (OCS), 18% were prescribed theophylline, and 12% were treated with anti-IgE therapy (*i.e.,* omalizumab). Regular use of non-steroidal anti-inflammatory drugs (NSAIDs) was part of the exclusion criteria. Reliever medication, such as short/long acting β_2 agonists (SABA/LABA) or combination

therapy was used by all asthmatic subjects. In sub-groups of asthmatics anticholinergics and chromoglycate was used among others, of which use was reported to be 18% and 2%, respectively.

Participant stratification by asthma treatment

Severe asthmatic non-smoker participants were stratified by treatment. OCS stratification was previously described [2]. Participants reporting at least daily use of OCS and had detectable OCS metabolites in their urine were classified as confirmed OCS users. Participants reporting never or previous use of OCS and did not have detectable OCS metabolites in their urine were classified as confirmed non-users of OCS. Theophylline stratifications were based on participant reported use. Participants reporting at least daily use of theophylline were considered users; those reporting no prior use were considered non-users. Omalizumab stratification was based on medical records of administration. Serum-IgE matched individuals with no prior use were considered non-users using a 2:1 nested case-control design as described by Kolmert et al. [2].

Metabolomics analysis

Metabolomics data were acquired by liquid chromatography – high resolution mass spectrometry (LC-HRMS) using previously described methods [3]. The analytical sequence (injection order) was randomised by clinical group, sex, age, BMI, collection site, and ethnicity to avoid analytical bias [4]; matching baseline and longitudinal samples were placed together in the analytical sequence in alternating order. To normalise for urine concentration and to reduce matrix effects [5], the specific gravity (SG) was measured. Prior to analysis, batches of 100 x 5 ml urine samples were

thawed at 4°C, then centrifuged for 5 minutes at 250 rcf to pellet any precipitate. For each sample, 100 μL was used to measure SG on a refractometer (Atago UG-a), 100 μL was used to create a pooled quality control (QC) sample and 5 x 500 μl aliquots were prepared and returned to -80°C. A pooled QC was made for each daily batch. After the final batch, all daily pooled QC samples were thawed to prepare a final pooled QC; the SG was measured for this sample and sub-aliquots were prepared for each analytical batch. In total, samples were analysed in 17 batches with pooled quality control (QC) samples analysed after every 5th sample to monitor analytical drift and measure precision [6]. On the day of analysis, urine samples were diluted with LC-MS grade water (Sigma-Aldrich, St. Louis, MO, USA) to the lowest SG (1.00x) measurements of the sample set and prepared as described [3]. Metabolite extraction was then performed by adding 180 μL of LC-MS grade acetonitrile (Fisher Scientific, Loughborough, UK) containing internal standards to 20 μL of SG-diluted urine. Samples were vortexed briefly then centrifuged at 13,000 x *g* for 15 minutes at 4°C; 40 μL of the supernatant was transferred to an LC-MS vial containing an insert for analysis.

Data were acquired using a 1290 Infinity II ultra-high performance liquid chromatography system coupled to an Agilent 6550 iFunnel Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). Metabolites were separated using hydrophilic interaction liquid chromatography (SeQuant ZIC-HILIC column 100 Å, 100 × 2.1 mm, 3.5 μm particle size) coupled to a 2.1 × 2 mm, 3.5 μm particle size guard column (Merck, Darmstadt, Germany) and an inline-filter. Mass spectral data were acquired in both electrospray ionisation (ESI) positive and negative modes. The mobile phases for ESI-positive ionization mode were water containing

0.1% formic acid (solvent A) and acetonitrile containing 0.1% formic acid (solvent B), and for ESI-negative ionization mode were 10 mM ammonium acetate pH 6.7 (solvent A) and acetonitrile (solvent B). The elution gradient was as follows: 1.5 min at 95% [B], 95 to 40% [B] in 12 min, maintained at 40% [B] for 2 min, then decreasing to 25% [B] at 14.2 min, maintained for 2.8 min, then returned to initial conditions over 1 min, and then the column was equilibrated at initial conditions for 7 min. The flow rate was 0.3 mL/min; injection volume was 2 µL, and the column oven was maintained at 25 °C

Data were acquired in a mass range of 40–1200 *m/z* using the following settings: sheath gas, N_2, 8 L/min; drying gas, N_2, 15 L/min; gas temperature, 250°C; nebulizer pressure, 35 psi; voltage, 3000 V; fragmentor voltage, 380 V. All data were acquired using all ions fragmentation (AIF) mode; this included three sequential experiments at three alternating collision energies (0 eV, 10 eV, and 30 eV). The data acquisition rate was 6 scans/s.

Peak deconvolution and metabolite identification were performed using Agilent TOF-Quant software (version B.07.00, Agilent Technologies) as described [3]. To ensure accurate metabolite identification, metabolites were matched against retention time, accurate mass, and MS/MS fragmentation patterns of 408 chemical reference standards in an in-house database. Metabolites were only included for statistical analysis if the accurate mass, retention time, and MS/MS fragmentation pattern matched to an authentic standard; thus, all metabolites reported have a Level 1 identification level as defined by the Metabolite Standards Initiative [7].

Systematic experimental within- and between-batch variation was corrected using the QC- Robust Spline Correction (QC-RSC) algorithm [8]. Metabolite abundances were then plotted against injection order and visually inspected to identify deconvolution problems; deconvolution was optimised and repeated as necessary. Following this procedure, the relative standard deviation of the pooled QC samples (QC_{RSD}) and the ratio of QC variance to sample variance (D-ratio) were calculated for each metabolite, aligning to community quality control best practice [6]. Quality assessment revealed high quality data as evidenced by an average relative standard deviation (RSD_{QC}) of 3.3%, an average D-Ratio of 8.2%, and 1.38% total missing values.

Tryptophan quantification

Tryptophan and 6 of its metabolites were quantified by reversed-phase liquid chromatography coupled to mass spectrometry (LC-MS/MS). Briefly, the urine samples were diluted 100 times in purified water and centrifuged (Eppendorf Centrifuge 5430 R) at 15,000 rcf for 10 min at 4°C. Part of the supernatant (200 µl) was transferred to a 96 deep well plate (Thermo Scientific 26052) and capped with a mat. Calibration curves were diluted in purified water and run together with the samples. Samples were analysed on an Agilent 1290 Infinity II system with multiwash function and an Agilent 6490 Ion Funnel triple quadrupole mass spectrometer. 5 µl of the sample extract was injected into a Zorbax Eclipse RRHD C18 column (50 × 2.1 mm, 1.8 µm particle size). A short gradient (0.5 ml/min flow rate, 40°C column oven) using 0.1 % formic acid in HPLC water (mobile phase A) (Milli-Q, Millipore) and 0.1 % formic acid in HPLC acetonitrile (mobile phase B) (Rathburn Chemicals) was applied. The gradient started at 2 % B increasing to 40 % B after 2 min and directly to a washing step at 95 % B for 0.6 minutes and then returned to initial conditions followed by an

0.8-minute column re-equilibration. Mass spectrometry data (MRM, multiple reaction monitoring) were acquired in positive electrospray ionization mode, using the transitions 225.09>109.9, 209.09>146.0, 205.1>188.1, 192.07>145.9, 190.05>144.0, 177.1>160.0 and 168.03>105.9 for 3-hydroxykynurenine, kyrunenine, tryptophan, 5-hydroxyindoleacetic acid, kynurenic acid, serotonin and quinolinic acid, respectively. Fragmentor voltage was set to 380 V and the Collision energies ranged from 10 to 18 V. In positive mode, the capillary voltage was 4.0 kV with a sheath gas temperature of 400°C and gas flow of 12 l/min. The Ion Funnel parameters were 200 and 110 for the high and low pressure radio frequencies. Data handling and quantification was performed using Agilent MassHunter B06.00 software. Samples were randomized across each batch to prevent potential confounding signal drift.

Genotyping

Sputum and bronchial brushing cis-eQTL summary statistic data were obtained in U-BIOPRED. The U-BIOPRED genotype data was imputed by IMPUTE2 [9] using 1000Genome phase 3 data [10] as the reference panel. The sputum eQTL analysis was performed on 91 U-BIOPRED participants that had both genetic and gene expression data in sputum available, and the bronchial brushing eQTL analysis was performed on 118 U-BIOPRED participants that had both genetic and gene expression data in bronchial brushings available. The eQTL analysis was performed with matrixEQTL in R [11] using age, sex and 10 principle components as covariates. The asthma GWAS summary statistics were also downloaded from a recent large scale GWAS study [12]. The number of cases and controls in the adult-onset GWAS were 26,582 and 327,253, respectively. The genome build used for both the eQTL and GWAS summary statistic was GRCh37. To determine if any of the SLC22A5 cis-eQTLs

overlapped with the asthma GWAS hits, a p-value threshold of 0.05 was applied for sputum and bronchial brushing eQTLs and genome-wide threshold of 5e-8 for GWAS SNPs. A lenient threshold was applied for the eQTL due to the fact that the eQTLs in sputum and bronchial brushing were underpowered due to the small sample size. We then subset the eQTLs with their asthma GWAS summary statistics also available, and aligned them to have the same effect allele with their GWAS summary statistics.

Statistical analysis

Missing values were imputed using the K-nearest neighbour (K=3) method, as is standard practice for metabolomics [13]. Metabolomics data, expressed as relative abundances, were not normally distributed (Lilliefors test); non-parametric univariate statistical tests were subsequently used. The null hypothesis (H_0), that the distribution of each metabolite was the same across outcomes, was tested using the Wilcoxon Rank-sum (2 outcomes) and Kruskal-Wallis test (more than 2 outcomes). The H_0, that baseline and longitudinal distributions were the same for each metabolite, was tested using the Wilcoxon signed-rank test. The Storey positive false discovery rate (FDR) [14] was calculated for all univariate analyses. Median fold-changes and confidence intervals were estimated using 500 iterations of bootstrap resampling [15] as previously reported [16]. Confounder correction was performed on the log urinary carnitine abundances using multiple linear regression, adjusting the clinical outcome for sex, age, and BMI. To provide robustness against heteroscedasticity, Huber's sandwich estimator for the regression coefficient standard errors was used [17]. Collection site-specific batch effects upon the observed metabotypes were evaluated and found to vary with metabolite.

To identify similarities between metabolites, hierarchical cluster analysis (HCA) was performed using a multivariate Spearman correlation distance metric and Ward's group linkage. The most similar metabolites form the lowest linkages in the resulting circular dendrogram; thus, emergent clusters represent similar trends. The mean of the log-transformed and z-scaled data of the resulting clusters were plotted against clinical groups to qualitatively visualise metabolite patterns across clinical groups. Principal Components – Canonical Variate Analysis (PC-CVA) was then performed, as previously described [16], to assess the multifactorial and correlated discrimination between clinical groups. Leave-one-out cross validation was used to determine the optimal number of PCs to be used in the model (**Figure E1**).

The targeted tryptophan data included 7 metabolites, including tryptophan and 6 downstream intermediates along 3 different pathways. Univariate analyses were performed using non-parametric tests (as described above). To visualise and assess clinical group differences across each pathway, mean data were presented in a bar graph and MANOVA was performed. Prior to MANOVA, data were normalised to tryptophan, log-transformed, and z-scaled.

In order to investigate the genetic impact of SLC22A5 on gene expression, we applied a linear regression framework to identify the cis-eQLTs associated with expression of SLC22A5. SNPs that were up to 1 MB (megabase) away from the TSS (transcription start site) of the gene were used in the association test. In the regression model, the genetic effect was assumed to be "additive", and standard linear regression was considered to model the genetic association with gene expression in SLC22A5:

$$E = \beta_0 + \beta_G G + \sum_{i=1}^{C} \beta_i C_i$$

Where *E* is log2-scaled gene expression, β_0 is the intercept term, G is the copies of effect allele, $C_{i=1,2\ldots C}$ are the covariates including age, sex, asthma groups and 10 principal components extracted from the genetic related matrix. Our null hypothesis (H_0) was that effect size β_G=0, *i.e.,* there was no genetic association to the gene expression. A p-value of β_G less than a given threshold was then used to determine the significant genetic association. We used a p-value threshold 0.05 here without considering the multiple testing correction due to the smaller sample size, and we have therefore emphasised caution when interpreting the result.

Individual subject clinical and biochemical data were collected from the U-BIOPRED TranSMART platform (eTRIKS). All statistical analyses were performed using the MATLAB scripting language (Mathworks, Natick, MA, USA), R (3.4.4, CRAN Network), and STATA v14 (StataCorp LLC, College Station, TX, USA).

Results of the tryptophan pathway analyses

Targeted quantitative analyses were performed to further elucidate tryptophan metabolite alterations (**Figure E9, Table E10**). No changes were observed in tryptophan levels. Given the variability in the response, downstream metabolites were normalized to tryptophan levels on an intra-individual basis (**Figure E9**). Tryptophan metabolism was divided into 3 metabolic pathways: Monoamine oxidase (MAO), Indoleamine 2,3-dioxygenase (IDO), and Kynurenine 3-monooxygenase (KMO). Mild-to-moderate asthmatics had similar levels in all 3 pathways relative to healthy participants, while severe asthmatics were dysregulated relative to healthy participants. Smoking status exerted a minor effect upon the pathways. The observed patterns reached significance for the IDO and KMO pathways ($p=0.01$ and 0.03, respectively) and were replicated at the 12-to-18-month follow-up. Tryptophan metabolism was further probed at the mRNA level (**Table E11**). Of the detected transcripts from enzymes in these pathways, the majority exhibited no change in association with asthma status. However, IDO1 transcripts significantly increased in bronchial brushings, PBMCs, and sputum. The magnitude of the observed shifts for other transcripts was nominal and most likely not of importance.

Discussion of the tryptophan pathway analyses

There is significant interest in the potential role of tryptophan and its metabolites in multiple inflammatory diseases [18], including obstructive lung disease [19]. However, the literature is unclear in relation to asthma. We accordingly used the scale of the U-BIOPRED study to address this question. Tryptophan itself was unchanged in association with asthma severity or smoking status. However, the downstream metabolites were dysregulated with asthma severity and further perturbed by smoking

status (**Figure E9**). In the current study, the strongest observed changes were in the IDO pathway, which has previously been reported to be associated with allergic airway inflammation [19]. Of particular interest is that the transcript levels of IDO1 increased in bronchial brushings, sputum, and PBMCs, suggesting a systemic upregulation in this pathway that was reflected in the urinary metabotype. These profiles were generally stable at the 12-18 months longitudinal sampling. Serotonin was affected by smoking status as previously reported [20, 21], as well as OCS treatment. An intriguing hypothesis is that the tryptophan metabolites are produced by microbiota involved in the gut-lung axis [22], reflecting asthma-associated dysbiosis.

List of Supplementary Tables:

Table E1. Metabolite Information and Statistical Analysis by Cohort
Table E2. Baseline versus Longitudinal Analysis
Table E3. Effects of OCS on Metabolite Abundance
Table E4. Effects of Theophylline on Metabolite Abundance
Table E5. Effects of Omalizumab on Metabolite Abundance
Table E6. Effects of Anticholinergics on Metabolite Abundance
Table E7. Effects of Leukotriene Modifiers on Metabolite Abundance
Table E8. Summary of All Metabolites Significantly Affected by Therapeutics or Smoking in the SAns Group
Table E9. Confounder Correction on log(Urinary Carnitine Abundane)
Table E10. Analysis of Quantified Tryptophan Metabolites
Table E11. Tryptophan Metabolism Transcripts

Each supplementary table is presented in a separate sheet of the 'UBIOPRED_supplementary_tables.xlsx' file.

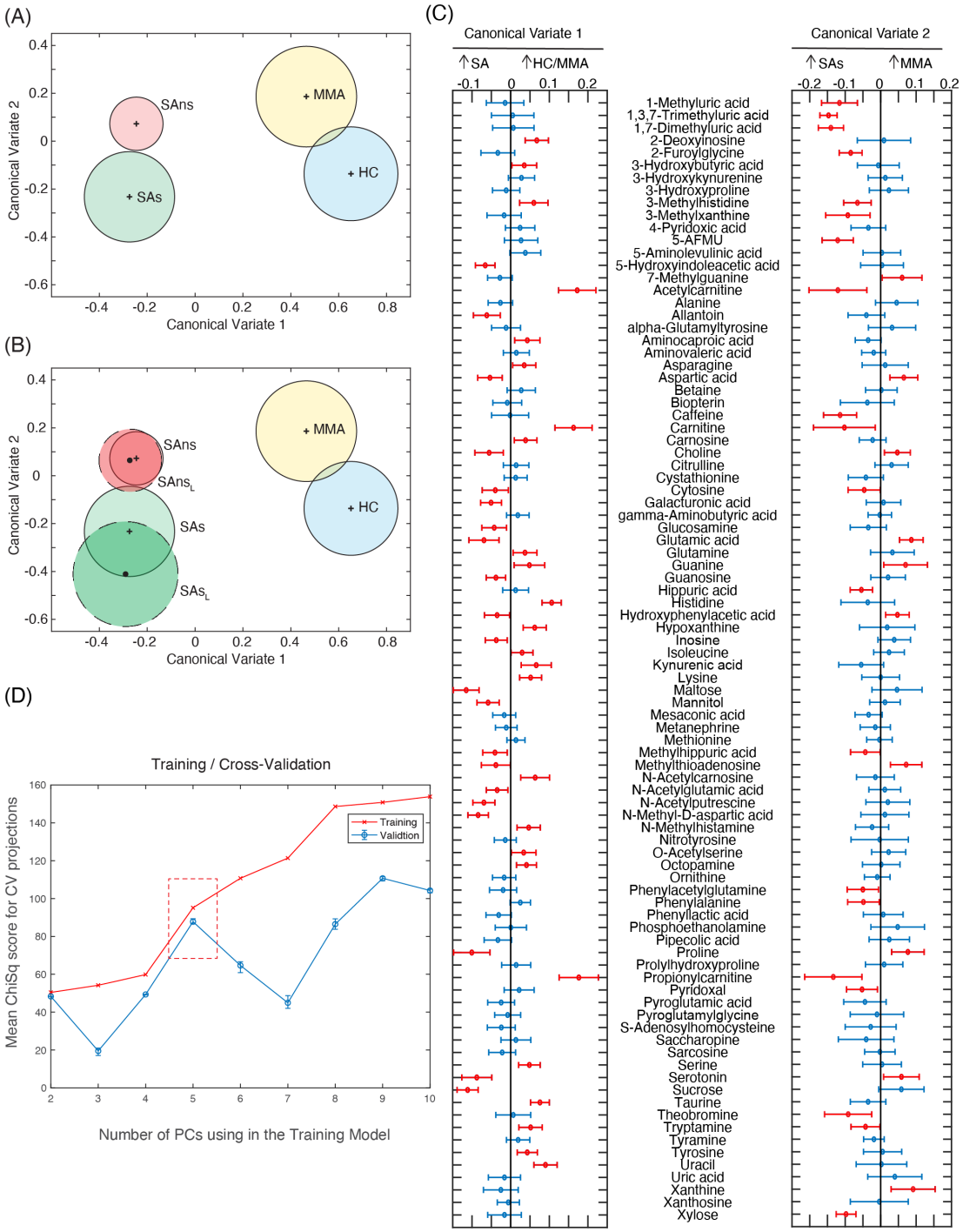


Figure E1. Principal Components – Canonical Variate Analysis (PC-CVA). (A) Scores plot of baseline data, labelled by clinical class. Blue, healthy controls; yellow, mild-to-moderate asthma (MMA); red, severe asthma non-smokers (SAns); green, severe asthma smokers (SAs). **(B)** Longitudinal data for severe asthma groups projected into the baseline model. $_B$, baseline data; $_L$, longitudinal data. +, mean of each baseline group; •, mean of each longitudinal group; solid circles, 95% confidence intervals of the mean of baseline groups; dashed circles, 95% confidence interval of the mean of longitudinal groups. **(C)** Loadings plots for Canonical Variate 1 (CV1, left panel) and CV2 (right panel). Red, metabolites that significantly ($p<0.05$) contribute to separation in the CV; blue, metabolites that do not significantly contribute to the separation in the CV. **(D)** Leave-one-out cross validation; 5 principal components were chosen as the number of optimal components to use in the model.

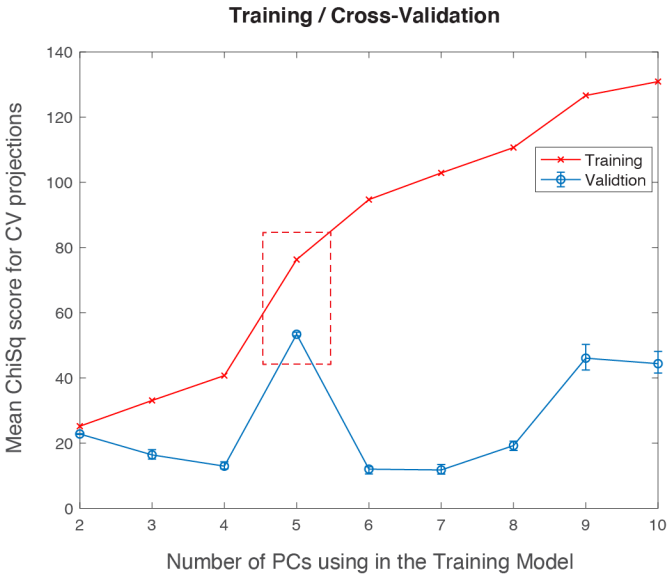


Figure E2. Cross Validation for the PC-CVA model presented in Figure 2. Leave-one-out cross validation was performed, identifying 5 principal components as the optimal number for the model.

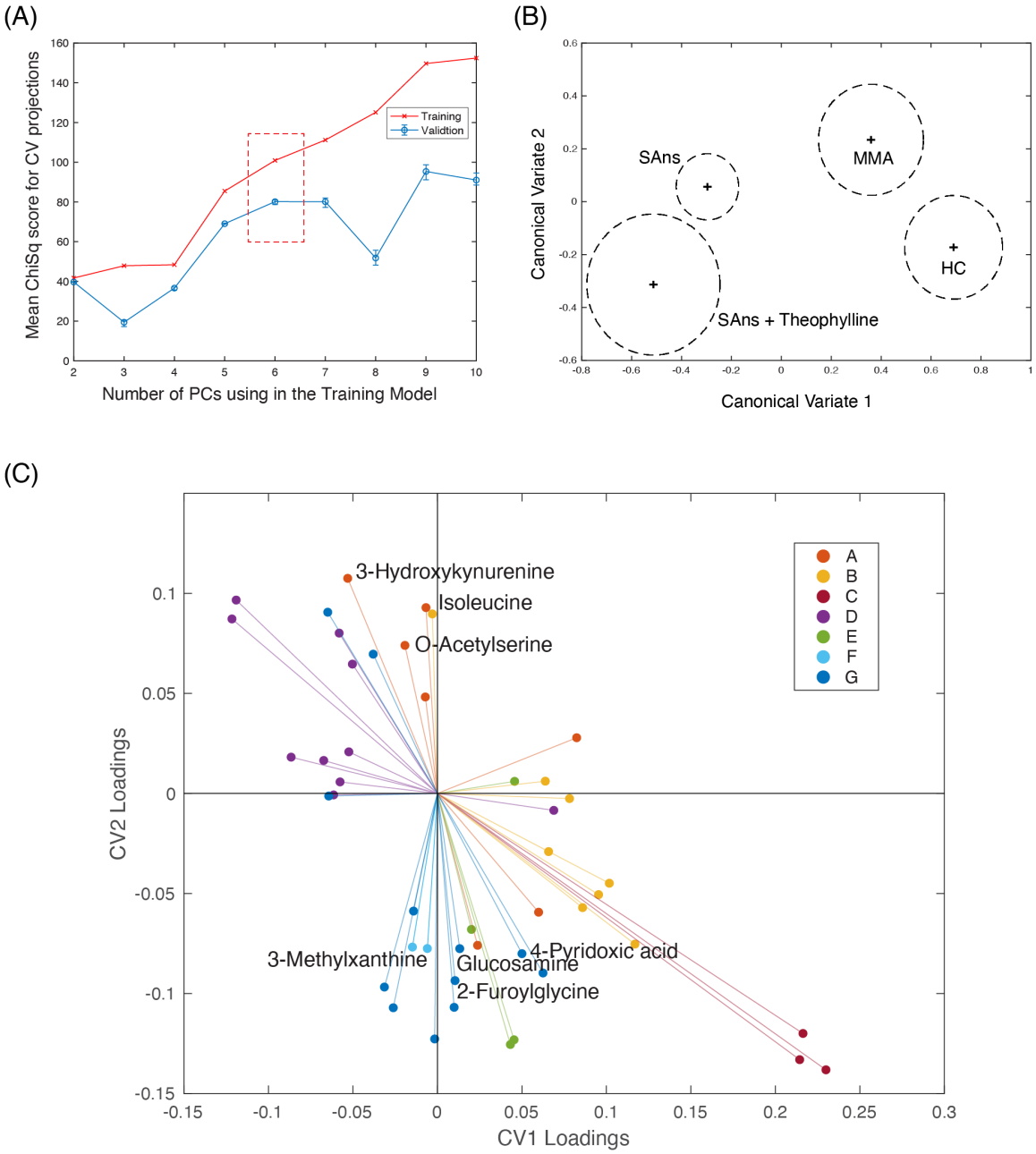


Figure E3. Principal Components – Canonical Variate Analysis (PC-CVA) with non-smoking severe asthmatics stratified by theophylline use. (A) Leave-one-out cross validation; 6 principal components were chosen as the number of optimal components to use in the model. **(B)** Scores plot labelled by outcome. HC, healthy control participants; MMA, mild-to-moderate asthmatics; SAns, severe asthma non-smokers; SAns + Theophylline, severe asthma non-smokers taking theophylline treatment. +, mean of each group; dashed circles, 95% confidence interval of the mean of each group. **(C)** Loadings plot displaying metabolites that significantly ($p<0.05$) contribute to the model. Metabolite position displays the magnitude and direction of affect in CV1 (x-axis) and CV2 (y-axis). The quadrant positions of metabolites are related to those of the clinical groups in the scores plots. In other words, metabolites are most abundant in the clinical groups with which they share a quadrant. Metabolites are colour-coded based on corresponding cluster as identified in **Figure 1** and according to the figure legend.

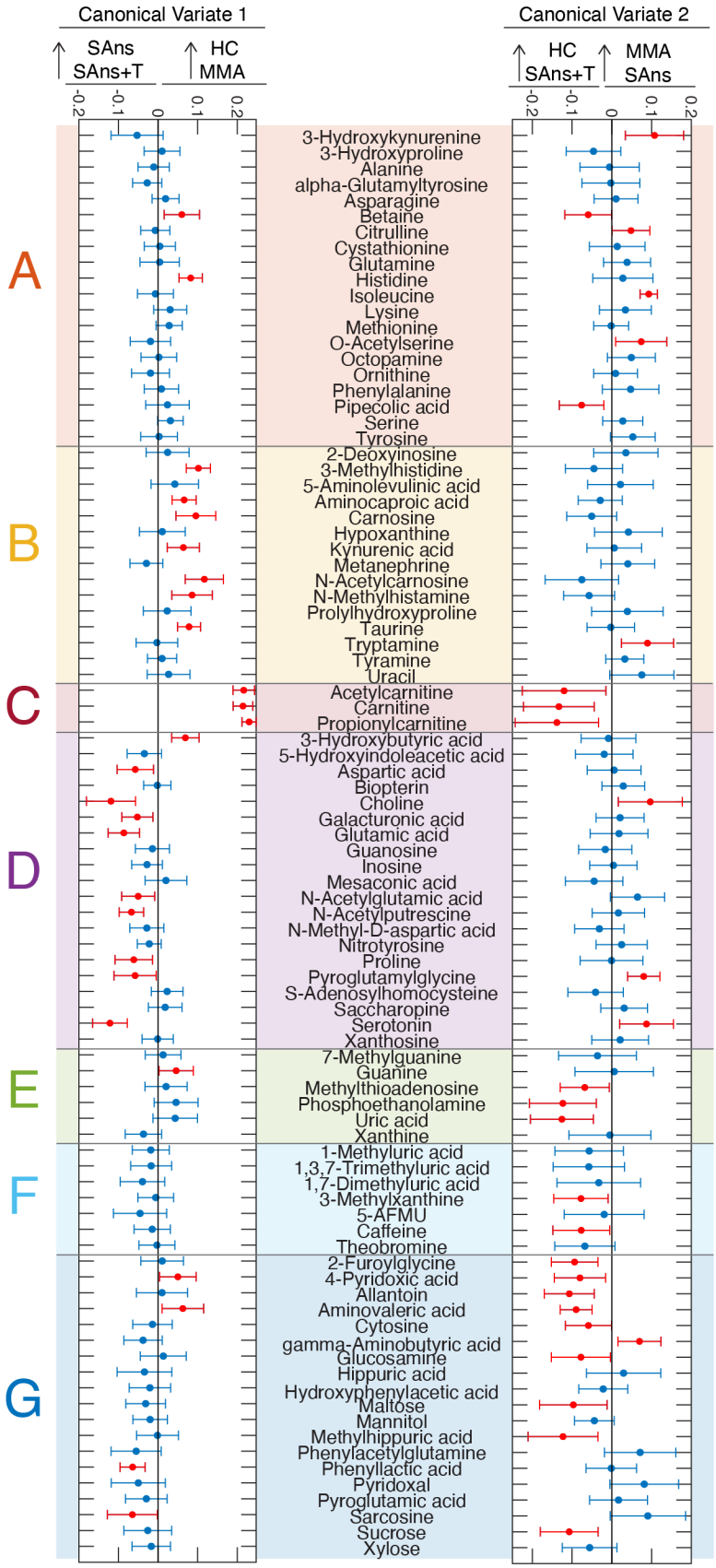
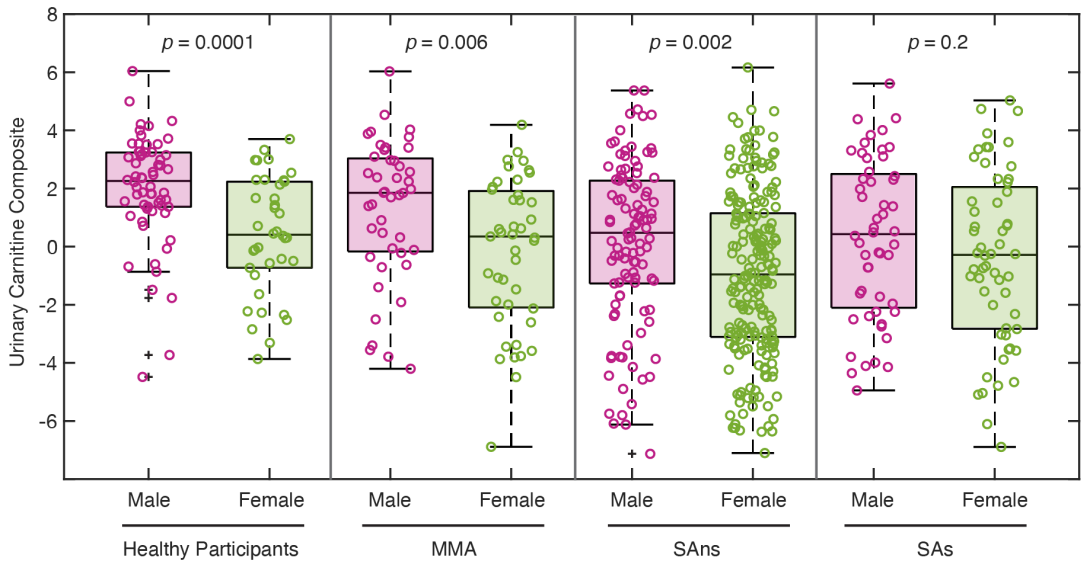
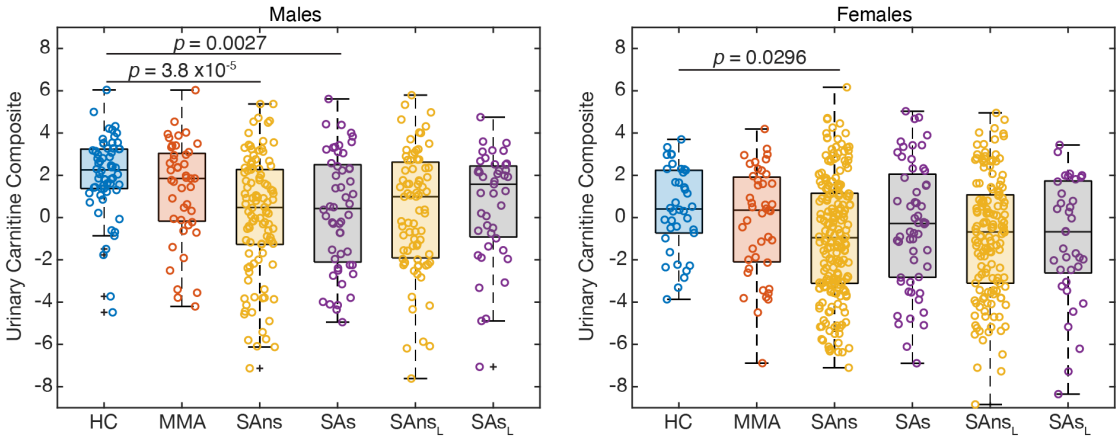


Figure E4. Individual Canonical Variate (CV) loadings for the PC-CVA with non-smoking severe asthmatics stratified by theophylline use. Loadings plots for Canonical Variate 1 (CV1, left panel) and CV2 (right panel) are shown. Red, metabolites that significantly ($p<0.05$) contribute to separation in the CV based on 500 iterations of bootstrap resampling / remodelling; blue, metabolites that do not significantly contribute to the separation in the CV. Metabolites are ordered and colour-coded by cluster as defined in **Figure 1**. The cluster label is presented on the left side of the figure. HC, healthy control participants; MMA, mild-to-moderate asthma; SAns, severe asthma non-smokers; SAs, severe asthma ex/smokers; T, theophylline.

(A)



(B)



(C)

	Males					
	Baseline				Longitudinal	
	HC	MMA	SAns	SAs	SAns	SAs
Acetylcarnitine	1 (0.81,1.22)	0.77 (0.54,1.14)	0.53 (0.43,0.68)	0.58 (0.31,0.88)	0.52 (0.31,0.76)	0.84 (0.39,1.21)
Carnitine	1 (0.83,1.22)	1.05 (0.77,1.21)	0.73 (0.59,0.87)	0.8 (0.58,1.02)	0.75 (0.59,1.02)	0.84 (0.59,1)
Propionylcarnitine	1 (0.77,1.29)	0.72 (0.37,1.06)	0.49 (0.38,0.63)	0.38 (0.29,0.79)	0.5 (0.29,0.73)	0.66 (0.37,0.9)

	Females					
	Baseline				Longitudinal	
	HC	MMA	SAns	SAs	SAns	SAs
Acetylcarnitine	1 (0.66,1.55)	0.69 (0.39,1.24)	0.47 (0.32,0.67)	0.58 (0.35,1.16)	0.53 (0.33,0.74)	0.59 (0.23,0.94)
Carnitine	1 (0.71,1.39)	1.02 (0.72,1.28)	0.84 (0.61,1)	0.94 (0.66,1.28)	0.87 (0.62,1.08)	0.97 (0.68,1.31)
Propionylcarnitine	1.01 (0.7,1.42)	0.95 (0.51,1.6)	0.6 (0.4,0.81)	0.62 (0.44,0.98)	0.66 (0.46,0.9)	0.66 (0.32,1.16)

Figure E5. Sex-specific differences in urinary carnitine levels. (A) The four U-BIOPRED groups stratified by sex. **(B)** The four U-BIOPRED groups compared separately for each sex. **C)** The fold-change estimates with 95% confidence intervals of the 3 carnitine species within each U-BIOPRED group stratified by sex relative to healthy participants. The urinary carnitine composite value was calculated by summing the log-transformed, z-scaled relative abundances of acetylcarnitine, carnitine, and propionylcarnitine. MMA, mild-to-moderate asthma; SAns, severe asthma non-smokers; SAs, severe asthma smokers and ex-smokers; $SAns_L$, severe asthma non-smokers longitudinal; SAs_L, severe asthma smokers and ex-smokers longitudinal.

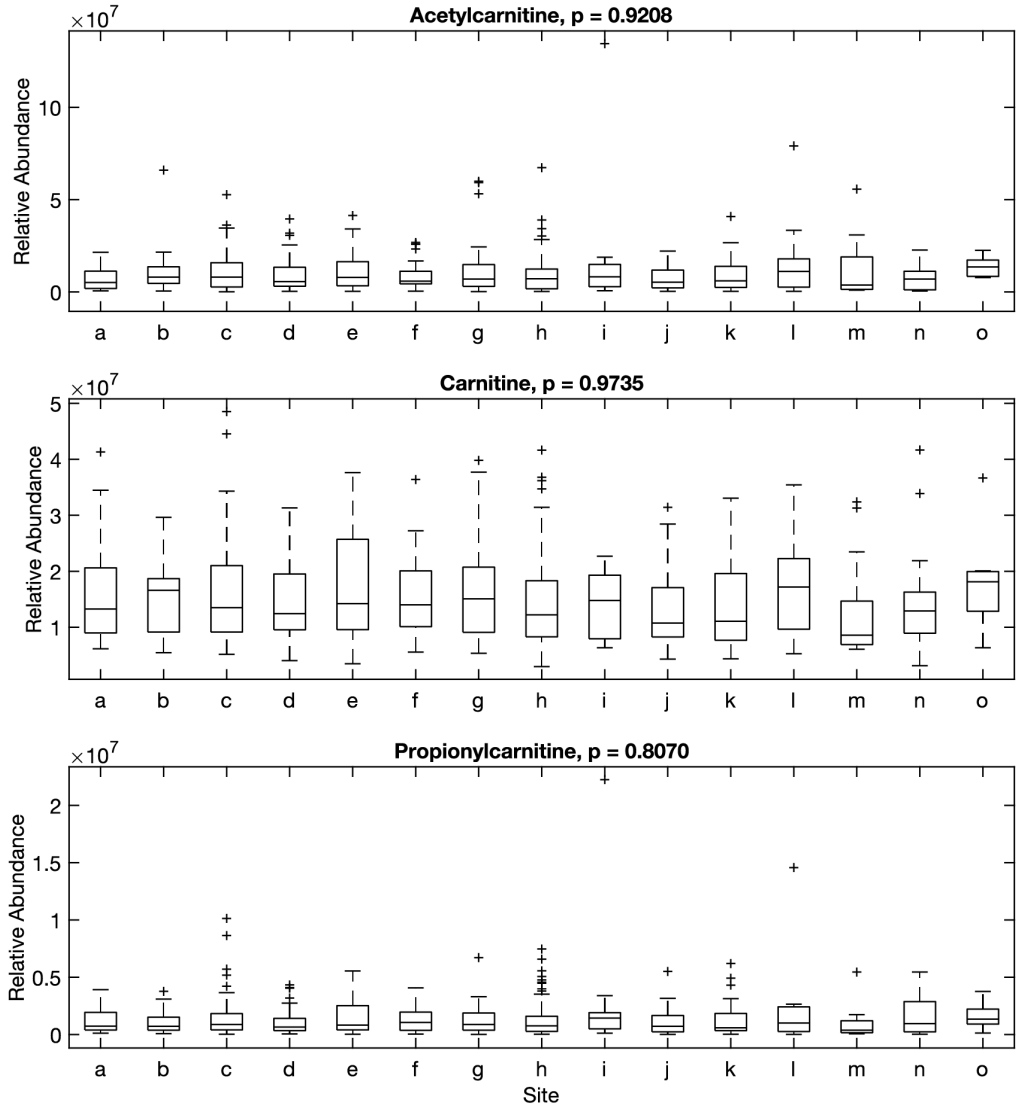


Figure E6. Dependency of urinary carnitine levels upon clinical recruitment site. The U-BIOPRED study consisted of 15 different patient recruitment sites across Europe. Boxplots are shown of the relative abundance of each carnitine metabolite stratified by clinical recruitment site code. Data are shown as post QC-corrected intensity values from the mass spectrometer in units of relative abundance. The Kruskal-Wallis p-values are reported in the figure.

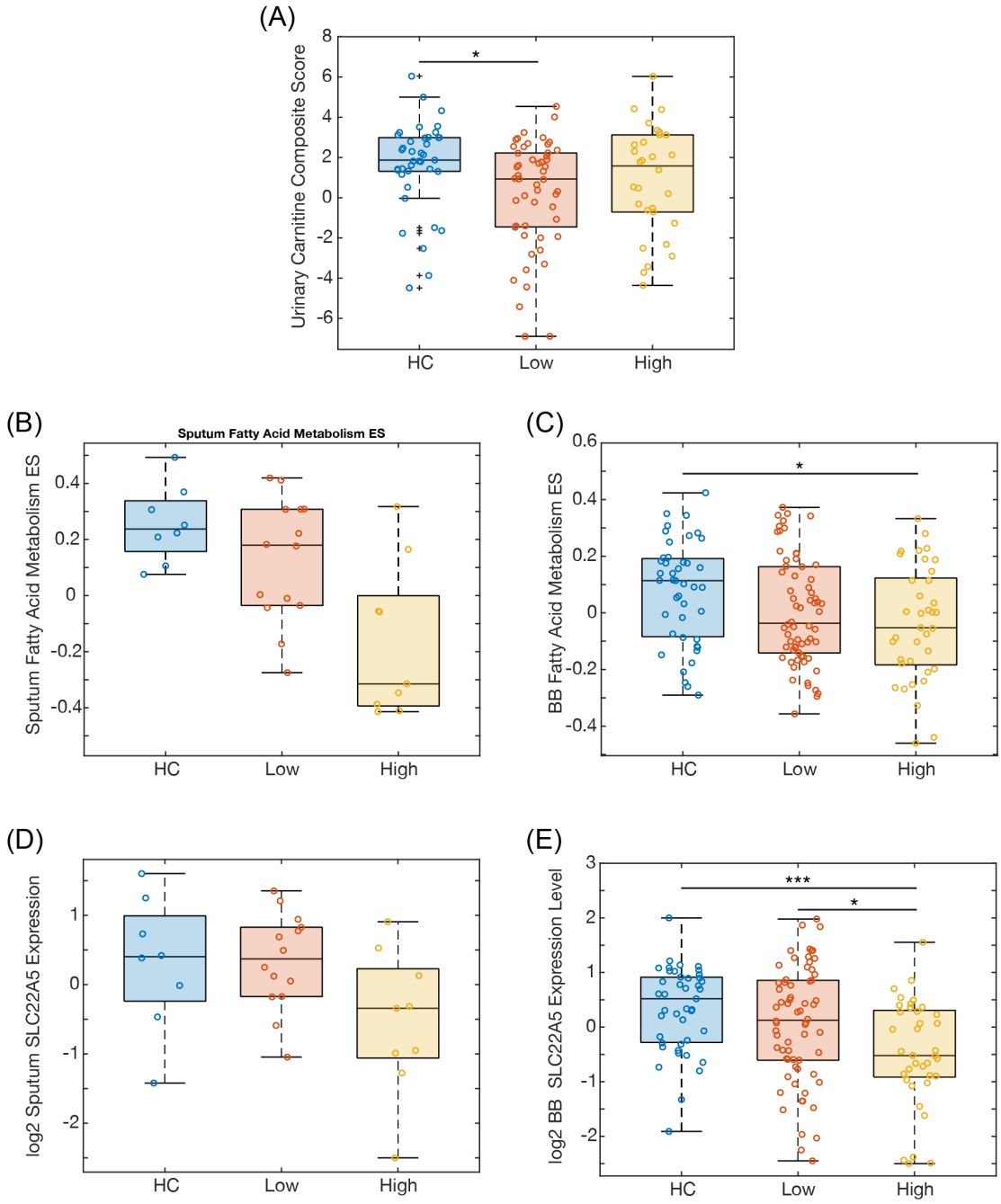


Figure E7. Molecular signatures of carnitine metabolism. Scatter-overlaid boxplots stratified by Type-2 classification. **(A)** Urinary carnitine composite variable. Relative abundances of carnitine, acetylcarnitine, and propionylcarnitine were log-transformed, z-scaled, and summed (p=0.031). **(B)** Sputum and **(C)** Bronchial brushings (BB) fatty acid metabolism enrichment score (ES) (p=0.038). **(D)** Sputum SLC22A5 expression levels. **(E)** Bronchial brushings (BB) SLC22A5 expression levels ($p=9.7x10^{-4}$). Inferential statistics were not performed for **(B)** and **(D)** due to small sample sizes. The Type-2 patient stratification was based on the ES of the IL-13-induced gene expression patterns in human bronchial epithelial cells using GSVA [23, 24]. Open circles, observations; box, median and interquartile range (IQR) of the data; whiskers, range of data up to 1.5 times of IQR above Q3 or below Q1; +, outliers. Kruskal-wallis p-values are reported, with posthoc pairwise comparisons shown on the figure. *, p<0.05; ***, p<0.001. HC, healthy control participants; Low, Type-2 low; High, Type-2 high.

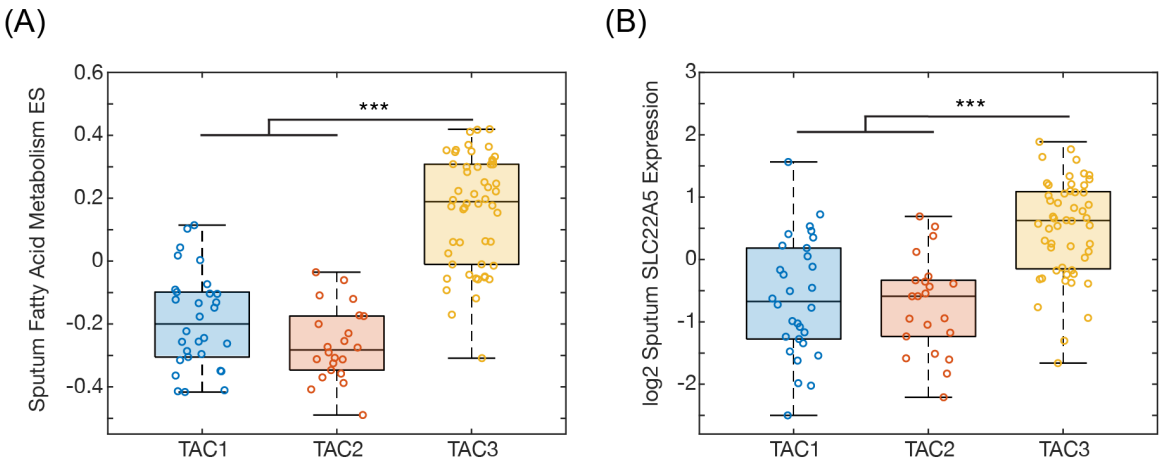


Figure E8. Molecular signatures of carnitine metabolism. Scatter-overlaid boxplots stratified by transcriptome-associated cluster (TAC) classification membership [23]. **(A)** Sputum fatty acid metabolism enrichment score (ES) ($p=2.1x10^{-14}$). **(B)** Sputum SLC22A5 expression levels ($p=1.2x10^{-8}$). Open circles, observations; box, median and interquartile range (IQR) of the data; whiskers, range of data up to 1.5 times of IQR above Q3 or below Q1; +, outliers. Kruskal-wallis p-values are reported here, with posthoc pairwise comparisons shown on the figure. ***, $p<0.001$.

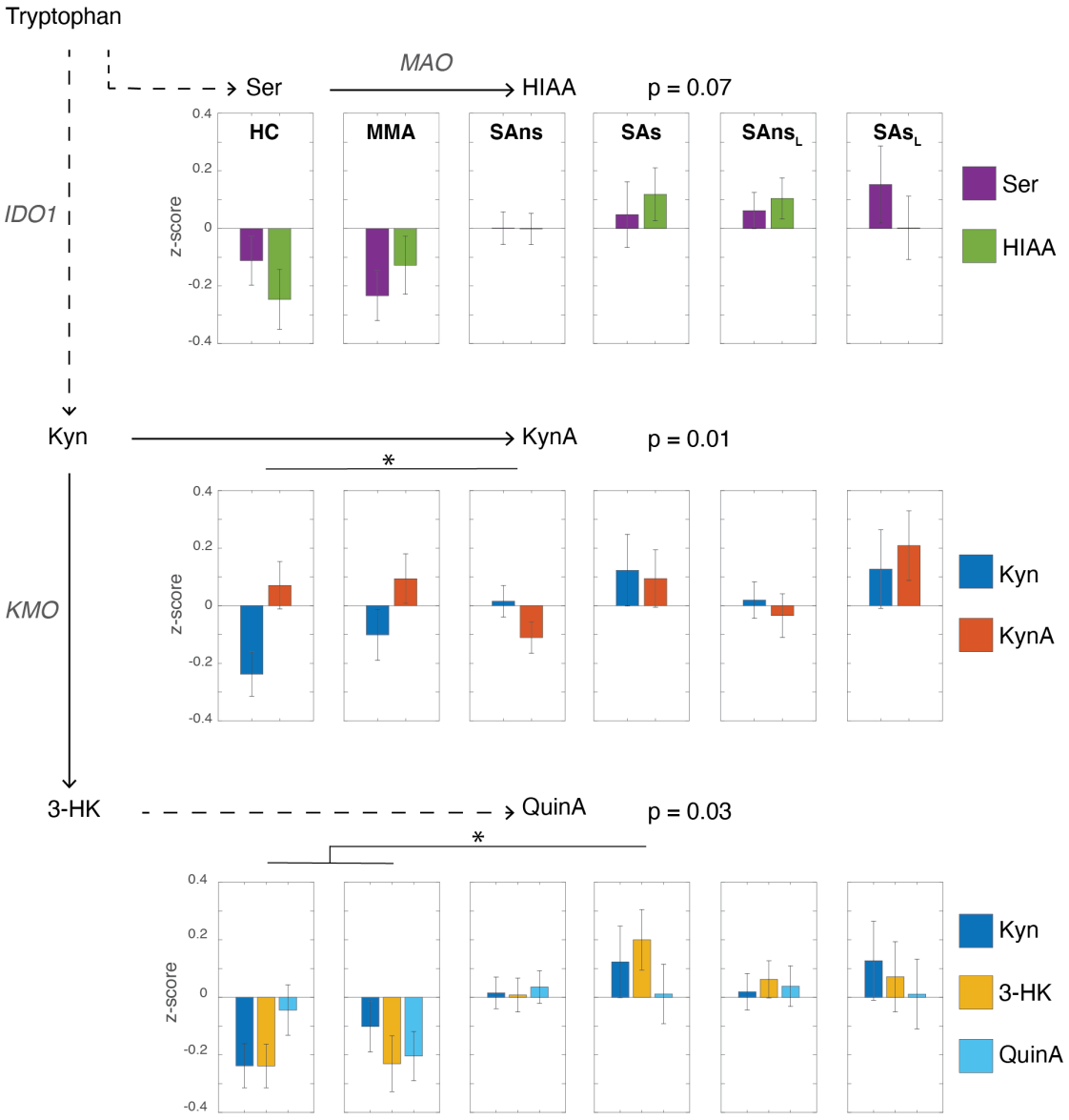


Figure E9. Tryptophan metabolism. Tryptophan and 6 downstream metabolites were quantified in urine and divided into 3 biochemical pathways: monoamine oxidase (MAO), indoleamine 2,3-dioxygenase (IDO), and kynurenine 3-monooxygenase (KMO). Downstream metabolites were normalised to tryptophan levels on an intra-individual basis, log-transformed, then z-scaled. To visualise and assess clinical group differences across each pathway, mean data were presented in a bar graph and MANOVA was performed. The quantified metabolite levels are provided in **Table E7**. Trp, tryptophan; Ser, serotonin; HIAA, 5-hydroxyindoleacetic acid; Kyn, kynurenine; KynA, kynurenic acid; QuinA, quinolinic acid. HC, healthy control participants; MMA, mild-to-moderate asthma; SAns, severe asthma non-smokers; SAs, severe asthma ex/smokers; $_{L}$, longitudinal data.

References

1. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, Pandis I, Bansal AT, Bel EH, Auffray C, Compton CH, Bisgaard H, Bucchioni E, Caruso M, Chanez P, Dahlen B, Dahlen SE, Dyson K, Frey U, Geiser T, Gerhardsson de Verdier M, Gibeon D, Guo YK, Hashimoto S, Hedlin G, Jeyasingham E, Hekking PP, Higenbottam T, Horvath I, Knox AJ, Krug N, Erpenbeck VJ, Larsson LX, Lazarinis N, Matthews JG, Middelveld R, Montuschi P, Musial J, Myles D, Pahus L, Sandstrom T, Seibold W, Singer F, Strandberg K, Vestbo J, Vissing N, von Garnier C, Adcock IM, Wagers S, Rowe A, Howarth P, Wagener AH, Djukanovic R, Sterk PJ, Chung KF, U-BIOPRED Study Group. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J* 2015: 46(5): 1308-1321.

2. Kolmert J, Gomez C, Balgoma D, Sjodin M, Bood J, Konradsen JR, Ericsson M, Thorngren JO, James A, Mikus M, Sousa AR, Riley JH, Bates S, Bakke PS, Pandis I, Caruso M, Chanez P, Fowler SJ, Geiser T, Howarth P, Horvath I, Krug N, Montuschi P, Sanak M, Behndig A, Shaw DE, Knowles RG, Holweg CTJ, Wheelock AM, Dahlen B, Nordlund B, Alving K, Hedlin G, Chung KF, Adcock IM, Sterk PJ, Djukanovic R, Dahlen SE, Wheelock CE, U-BIOPRED Study Group. Urinary Leukotriene E4 and Prostaglandin D2 Metabolites Increase in Adult and Childhood Severe Asthma Characterized by Type 2 Inflammation. A Clinical Observational Study. *Am J Respir Crit Care Med* 2021: 203(1): 37-53.

3. Naz S, Gallart-Ayala H, Reinke SN, Mathon C, Blankley R, Chaleckis R, Wheelock CE. Development of a Liquid Chromatography-High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition. *Anal Chem* 2017: 89(15): 7933-7942.

4. Dudzik D, Barbas-Bernardos C, Garcia A, Barbas C. Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *J Pharm Biomed Anal* 2018: 147: 149-173.

5. Edmands WM, Ferrari P, Scalbert A. Normalization to specific gravity prior to analysis improves information recovery from high resolution mass spectrometry metabolomic profiles of human urine. *Anal Chem* 2014: 86(21): 10925-10931.

6. Broadhurst D, Goodacre R, Reinke SN, Kuligowski J, Wilson ID, Lewis MR, Dunn WB. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* 2018: 14(6): 72.

7. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007: 3(3): 211-221.

8. Kirwan JA, Broadhurst DI, Davidson RL, Viant MR. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Anal Bioanal Chem* 2013: 405(15): 5147-5157.

9. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009: 5(6): e1000529.

10. Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015: 526(7571): 68-74.

11. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012: 28(10): 1353-1358.

12. Ferreira MAR, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, Brew BK, Ullemar V, Lu Y, Jiang Y, andMe Research T, e QC, Consortium B, Magnusson PKE, Karlsson R, Hinds DA, Paternoster L, Koppelman GH, Almqvist C. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am J Hum Genet* 2019: 104(4): 665-684.

13. Do KT, Wahl S, Raffler J, Molnos S, Laimighofer M, Adamski J, Suhre K, Strauch K, Peters A, Gieger C, Langenberg C, Stewart ID, Theis FJ, Grallert H, Kastenmuller G, Krumsiek J. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics* 2018: 14(10): 128.

14. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002: 64(3): 479-498.

15. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009.

16. Reinke SN, Gallart-Ayala H, Gomez C, Checa A, Fauland A, Naz S, Kamleh MA, Djukanovic R, Hinks TS, Wheelock CE. Metabolomics analysis identifies different metabotypes of asthma severity. *Eur Respir J* 2017: 49(3).

17. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkely Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. University of California Press, Berkeley, California, 1967; pp. 221-233.

18. Yeung AW, Terentis AC, King NJ, Thomas SR. Role of indoleamine 2,3-dioxygenase in health and disease. *Clin Sci (Lond)* 2015: 129(7): 601-672.

19. Xu H, Oriss TB, Fei M, Henry AC, Melgert BN, Chen L, Mellor AL, Munn DH, Irvin CG, Ray P, Ray A. Indoleamine 2,3-dioxygenase in lung dendritic cells promotes Th2 responses and allergic inflammation. *Proc Natl Acad Sci U S A* 2008: 105(18): 6690-6695.

20. Naz S, Bhat M, Stahl S, Forsslund H, Skold CM, Wheelock AM, Wheelock CE. Dysregulation of the Tryptophan Pathway Evidences Gender Differences in COPD. *Metabolites* 2019: 9(10).

21. Gu F, Derkach A, Freedman ND, Landi MT, Albanes D, Weinstein SJ, Mondul AM, Matthews CE, Guertin KA, Xiao Q, Zheng W, Shu XO, Sampson JN, Moore SC, Caporaso NE. Cigarette smoking behaviour and blood metabolomics. *Int J Epidemiol* 2016: 45(5): 1421-1432.

22. Budden KF, Shukla SD, Rehman SF, Bowerman KL, Keely S, Hugenholtz P, Armstrong-James DPH, Adcock IM, Chotirmall SH, Chung KF, Hansbro PM. Functional effects of the microbiota in chronic respiratory disease. *Lancet Respir Med* 2019: 7(10): 907-920.

23. Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, Sousa A, Corfield J, Djukanovic R, Lutter R, Sterk PJ, Auffray C, Guo Y, Adcock IM, Chung KF, U-BIOPRED Study Group. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J* 2017: 49(2).

24. Pavlidis S, Takahashi K, Ng Kee Kwong F, Xie J, Hoda U, Sun K, Elyasigomari V, Agapow P, Loza M, Baribaud F, Chanez P, Fowler SJ, Shaw DE, Fleming LJ, Howarth PH, Sousa AR, Corfield J, Auffray C, De Meulder B, Knowles R, Sterk PJ, Guo Y, Adcock IM, Djukanovic R, Fan Chung K, U-BIOPRED Study Group. "T2-high" in severe asthma related to blood eosinophil, exhaled nitric oxide and serum periostin. *Eur Respir J* 2019: 53(1).