Online Data Supplement

Metabolomic Differences in Lung Function Metrics: Evidence from two Cohorts

Rachel S. Kelly*, Isobel D. Stewart*, Haley Bayne, Priyadarshini Kachroo, Avron Spiro, III, Pantel Vokonas, David Sparrow, Scott T. Weiss, Hanna Knihtilä, Augusto A. Litonjua, Nicholas J. Wareham[2] Claudia Langenberg*, Jessica A. Lasky-Su*

Supplementary Methods

Spirometry/ respiratory disease status

Epic-Norfolk: FEV_1 and FVC pre-bronchodilator were measured at the time of blood collection (at baseline), using an electronic turbine spirometer (Micro Medical Instruments, Rochester, UK). Measurements were taken twice, and the higher of the two measures analyzed [1]. Asthma status at baseline was ascertained using self-report of a previous doctor's diagnosis.

NAS: Spirometry was repeated up to a maximum of eight spirograms, allowing at least three acceptable spirograms, at least 2 of which were reproducible with FEV_1 and FEV_1/FVC pre-bronchodilator measurements within 5% of each spirogram; the best of these 2 values was selected from a given encounter [2]. Acceptability of spirograms was judged according to American Thoracic Society standards [3, 4] The measures at the time of, or closest to blood draw were selected for analysis. Self-report of a doctors diagnosis of asthma status was adjusted for in the model, only two men reported a doctors diagnosis of emphysema so this variable was not included in the model, and information on COPD status as blood draw was not available.

Metabolite Data QC and processing

In EPIC-Norfolk, blood samples were collected at recruitment between 1993 and 1998 and underwent metabolomic profiling between 2015 and 2107 making them between 17 and 24 years old at profiling. In NAS, blood samples were collected between the years 2000 and 2008 and profiled in 2018, making them between 10 and 18 years old at the time of profiling.

In EPIC-Norfolk, measurements were made in citrated plasma samples taken at baseline, in two sets each consisting of approximately 6,000 quasi-randomly selected individuals. Blood samples were collected at recruitment between 1993 and 1998 and underwent metabolomic profiling between 2015 and 2107 making them between 17 and 24 years old at profiling.

Data were processed according to the standard quality control pipeline for each cohort. In EPIC-Norfolk, metabolite measures were median normalized across run days (with medians set to 1), without imputation of missing values. Within each measurement batch, metabolites measured in <70% of individuals were excluded and remaining missing values were imputed with half the minimum value for that metabolite. Metabolite measures were natural log transformed, windsorized at 5 SD and standardized ($\mu = 0$, SD = 1). We note that standardization was performed in men and women together.

NAS is comprised only of men. Subjects were selected for this study on the basis of an available sample plasma suitable for metabolomic profiling [5], and concurrently measured spirometry. Metabolite intensities were log transformed and *pareto* scaled. Those metabolites with a variance

of 0 were excluded from further analysis (n=144). Imputation was performed by assigning missing/unquantified values half the lowest value across all samples for each metabolite.

In EPIC-Norfolk, association analyses were performed individually within each measurement set, applying a minimum sample size threshold of >30 observations in each. Individuals who reported a previous diagnosis of bronchitis/emphysema were excluded. A fixed effects inverse variance weighted meta-analysis was performed to pool the results of the two measurement sets. Metabolites were matched between measurement batches by the Metabolon-assigned "Chemical ID" or by metabolite name for those without chemical identification. Only metabolites analysed in both measurement sets were considered.

Metabolon utilizes a tiered identification system. The tiered identifications are based on the associated analytical parameters for each metabolite. A tier one identification is predicated on match of retention time, accurate mass, and MS/MS fragmentation profile to a reference standard for unequivocal metabolite identification.

Of the 693 metabolites included in these analyses, 457 (66%) were tier one identified. Sixty-nine (10%) were classified as tier two, indicating they have not been confirmed based on a standard, but metabolon are confident in its identity based on a subset of analytical parameters.

The remainder (n=147, 24%) represent spectral features that Metabolon see recurring in biological samples that they can track and quantify. Because of this, Metabolon are confident that they represent biologically relevant molecules and not analytical artifacts. These identifiers

are in accordance with the proposed minimum reporting standards from the Metabolomics Standards Initiative [6].

Supplementary Tables

Table E1. Baseline characteristics of the included EPIC-Norfolk Participants

	All participants	Men	Women	*P*-value*
	(n = 10,460)	(n = 4,868)	(n = 5,592)	
Age, mean (SD)	59.73 (8.95)	60.14 (8.97)	59.37 (8.91)	<0.001
Ethnicity/Ethnic Origin, n (%)				0.091
White	10392 (99.35%)	4832 (99.26%)	5560 (99.43%)	
Black Caribbean	6 (0.06%)	2 (0.04%)	4 (0.07%)	
Black Other	3 (0.03%)	3 (0.06%)	0	
Indian	4 (0.04%)	4 (0.08%)	0	
Pakistani	1 (0.01%)	1 (0.02%)	0	
Chinese	3 (0.03%)	2 (0.04%)	1 (0.02%)	
Other	8 (0.08%)	5 (0.10%)	3 (0.05%)	
Missing	43 (0.41%)	19 (0.39%)	24 (0.43%)	
Smoking, n (%)				<0.001
Current	1145 (10.95%)	573 (11.77%)	572 (10.23%)	
Former	4448 (42.52%)	2689 (55.24%)	1759 (31.46%)	
Never	4867 (46.53%)	1606 (32.99%)	3261 58.32%)	
BMI kg/m^2, mean (SD)	26.18 (3.70)	26.35 (3.12)	26.04 (4.14)	<0.001
Height cm, mean (SD)	166.99 (9.12)	173.96 (6.71)	160.93 (6.11)	<0.011
Asthma, n (%)				0.082
Yes	652 (6.23%)	282 (5.79%)	370 (6.62%)	
No	9808 (93.77%)	4586 (94.21%)	5222 (93.38%)	
FEV_1 (Liters), mean (SD)	2.52 (0.72)	2.93 (0.72)	2.16 (0.51)	<0.001
FEV_1/FVC, mean (SD)	0.82 (0.11)	0.81 (0.12)	0.82 (0.10)	<0.001
FEV_1/FVC <0.7, n (%)	1285 (12.3%)	768 (15.8%)	517 (9.2%)	<0.001
FEV1 percent predicted, mean % (SD) [a]	86.51 (16.31)	85.44 (16.64)	87.44 (15.96)	<0.001
FEV1 percent predicted (%)				<0.001
FEV_1≥ 80% predicted	7444 (65.52%)	3235 (66.95%)	3989 (71.74%)	
50% ≤ FEV_1 < 80% predicted	2893 (27.84%)	1449 (29.99%)	1444 (25.97%)	
30% ≤ FEV_1 < 50% predicted	230 (2.21%)	122 (2.52%)	108 (1.94%)	
FEV_1 < 30% predicted	45 (0.43%)	26 (0.54%)	19 (0.34%)	

**Significance of difference between men and women was evaluated using chi-squared test for categorical variables and two-sample t-test for continuous variables.*

FEV_1 is measured in liters per second; FEV_1/FVC is expressed as a proportion; FEV1 percent predicted is calculated as the % of an individual's FEV1 to their expected FEV1 based on their age, height, sex and race, it was computed using the R package 'rspiro'

[a] could not be computed for 68 subjects

Table E2. Baseline characteristics of included participants in NAS

	Variable info	**Number (%), Mean (SD)**
n		437
Age		75.13 (6.66)
Ethnicity	White	429 (98.2%)
	Black	6 (1.4%)
	Other	2 (0.5%)
Smoke status	Regular	19 (4.3%)
	Former	286 (65.6%)
	Never	132 (30.2%)
BMI (kg/m^2)		27.63 (4.12)
Height (inches)		68.43 (2.74)
Smoke status	Regular	19 (4.3%)
	Former	286 (65.6%)
	Never	132 (30.2%)
Asthma status	Current	12 (2.8%)
	No Current Asthma	425 (97.3%)
FEV_1 (Liters)		2.50 (0.60)
FEV_1/FVC		0.74 (0.08)
FEV_1/FVC <0.7 (n)		114 (26.0%)
FEV1 percent predicted (%)[a]		86.3% (16.6)
FEV1 percent predicted (%)[a]	$FEV_1 \geq$ 80% predicted	283 (64.9%)
	50% $\leq FEV_1 <$ 80% predicted	137 (31.4%)
	30% $\leq FEV_1 <$ 50% predicted	15 (3.4%)
	$FEV_1 <$ 30% predicted	1 (2.3%)

FEV_1 is measured in liters per second; FEV_1/FVC is expressed as a proportion; FEV1 percent predicted is calculated as the % of an individual's FEV1 to their expected FEV1 based on their age, height, sex and race, it was computed using the R package 'rspiro'
[a]*could not be computed for one subject*

Table E3: Pathways identified as significant in the Pathway analysis based on the ENT95% associated metabolites from EPIC-Norfolk for FEV_1 and FEV_1/FVC Ratio

	FEV_1 (n=156 metabolites)				FEV_1/FVC Ratio (n=65 metabolites)			
Pathway Name	**Match Status**	**P for enrichment**	**FDR**	**Impact**	**Match Status**	**P for enrichment**	**FDR**	**Impact**
Alanine, aspartate and glutamate metabolism	-	-	-	-	6/28	7.62E-05	0.006	0.334
Aminoacyl-tRNA biosynthesis	8/48	5.52E-04	0.023	0.167	5/48	9.15E-03	0.126	0.167
Arginine biosynthesis	-	-	-	-	4/14	4.21E-04	0.018	0.193
Ascorbate and aldarate metabolism	2/8	4.00E-02	0.420	0.500	2/8	1.89E-02	0.177	0.000
Butanoate metabolism	-	-	-	-	3/15	7.17E-03	0.126	0.000
Caffeine metabolism	3/10	6.55E-03	0.183	0.000	-	-	-	-
Citrate cycle (TCA cycle)	-	-	-	-	4/20	1.80E-03	0.050	0.165
Cysteine and methionine metabolism	3/33	9.93E-03	0.209	0.339	-	-	-	-
D-Glutamine and D-glutamate metabolism	-	-	-	-	2/6	1.05E-02	0.126	0.500
Galactose metabolism	-	-	-	-	3/27	3.66E-02	0.279	0.035
Glutathione metabolism	4/28	2.57E-02	0.420	0.156	3/28	4.02E-02	0.281	0.084
Glycine, serine and threonine metabolism	8/33	3.33E-05	0.003	0.586	-	-	-	-
Glyoxylate and dicarboxylate metabolism	-	-	-	-	4/32	1.05E-02	0.126	0.122
Pyruvate metabolism	-	-	-	-	3/22	2.12E-02	0.178	0.115
Sphingolipid metabolism	-	-	-	-	3/21	1.87E-02	0.177	0.069
Taurine and hypotaurine metabolism	2/8	4.00E-02	0.420	0.286	-	-	-	-
Valine, leucine and isoleucine biosynthesis	2/8	4.00E-02	0.420	0.000	-	-	-	-

Match status indicates the number of metabolites on which the enrichment is based the denominator is the total number of metabolites in pathway, and the numerator is the number of metabolites form that pathway identified as significant in the association analyses.
Impact is a metric based on the importance of the significant metabolites (the numerator) to the pathway topology. A higher impact indicates that the metabolites are more important to that pathway.

Table E4. Metabolites associated with FEV_1 at a threshold of ENT95% in the total sample, in men or in women (see excel file)

Beta can be interpreted as the change in FEV_1 liters with a one unit increase of a given metabolite

°Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters
Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Table E5. Metabolites associated with FEV_1/FVC at a threshold of ENT95% in the total sample, in men or in women (see excel file)

Beta can be interpreted as the change in FEV_1/FVC proportion with a one unit increase of a given metabolite

°Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters

Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Table E6: Comparison of results with the initial models and when including 1050 individuals from EPIC-Norfolk with self-reported emphysema or bronchitis at baseline

Model	*n* significant metabolites in EPIC-Norfolk when inc. emphysema/ bronchitis cases	*n* significant metabolites in EPIC-Norfolk when excl. emphysema/ bronchitis cases (primary analysis)	*n* common metabolites	*n* metabolites validated in NAS when inc. EPIC-Norfolk emphysema/ bronchitis cases	novel validated metabolites when inc. emphysema/bronchitis cases
FEV_1	173	156	147	40	4-acetamindobutanoate; N-acetylglutamate; pregnanediol sulfate; androstenediol (3alpha, 17alpha monosulfate) (2); erythronate°; X-21353
FEV1 (Men only)	104	79	76	24	Kynurenine; 7-methylguanine; X-21258; X-21353; X-15728; mannose;
FEV1/FVC	73	65	60	7	4-methyl-2-oxopentanoate
FEV1/FVC (Men only)	34	35	30	2	-

Significance in EPIC-Norfolk defined as ENT95%;
Validation determined by $p<0.05$ and concordant direction of effect in NAS
primary analysis refers to that presented in the main body of text
for FEV1/FVC (men only) both validated metabolites were the same as in the initial model

°*Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters*
Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Table E7. Metabolites associated with FEV_1 at a threshold of ENT95% in participants with a healthy weight BMI, participants who are overweight and those who are obese, and metabolites that validated in NAS (see excel file)

Beta can be interpreted as the change in FEV_1 liters with a one unit increase of a given metabolite

°Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters

Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Table E8. Metabolites associated with FEV_1/FVC at a threshold of ENT95% in participants with a healthy weight BMI, participants who are overweight and those who are obese, and metabolites that validated in NAS (see excel file)

Beta can be interpreted as the change in FEV_1/FVC proportion with a one unit increase of a given metabolite

°Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters

Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Table E9. Metabolites associated with FEV_1 at a threshold of ENT95% in never smokers, former smokers and current smokers, and metabolites that validated in NAS (see excel file)

Beta can be interpreted as the change in FEV_1 liters with a one unit increase of a given metabolite

°Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters

Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Table E10. Metabolites associated with FEV_1/FVC at a threshold of ENT95% in never smokers, former smokers and current smokers, and metabolites that validated in NAS (see excel file)

Beta can be interpreted as the change in FEV_1/FVC proportion with a one unit increase of a given metabolite

°Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters

Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Supplementary Figures

Figure E1: Venn Diagram displaying the cross-over in significant metabolites between the Total EPIC-Norfolk population, Men only and Women only at a significance threshold of p<ENT95%

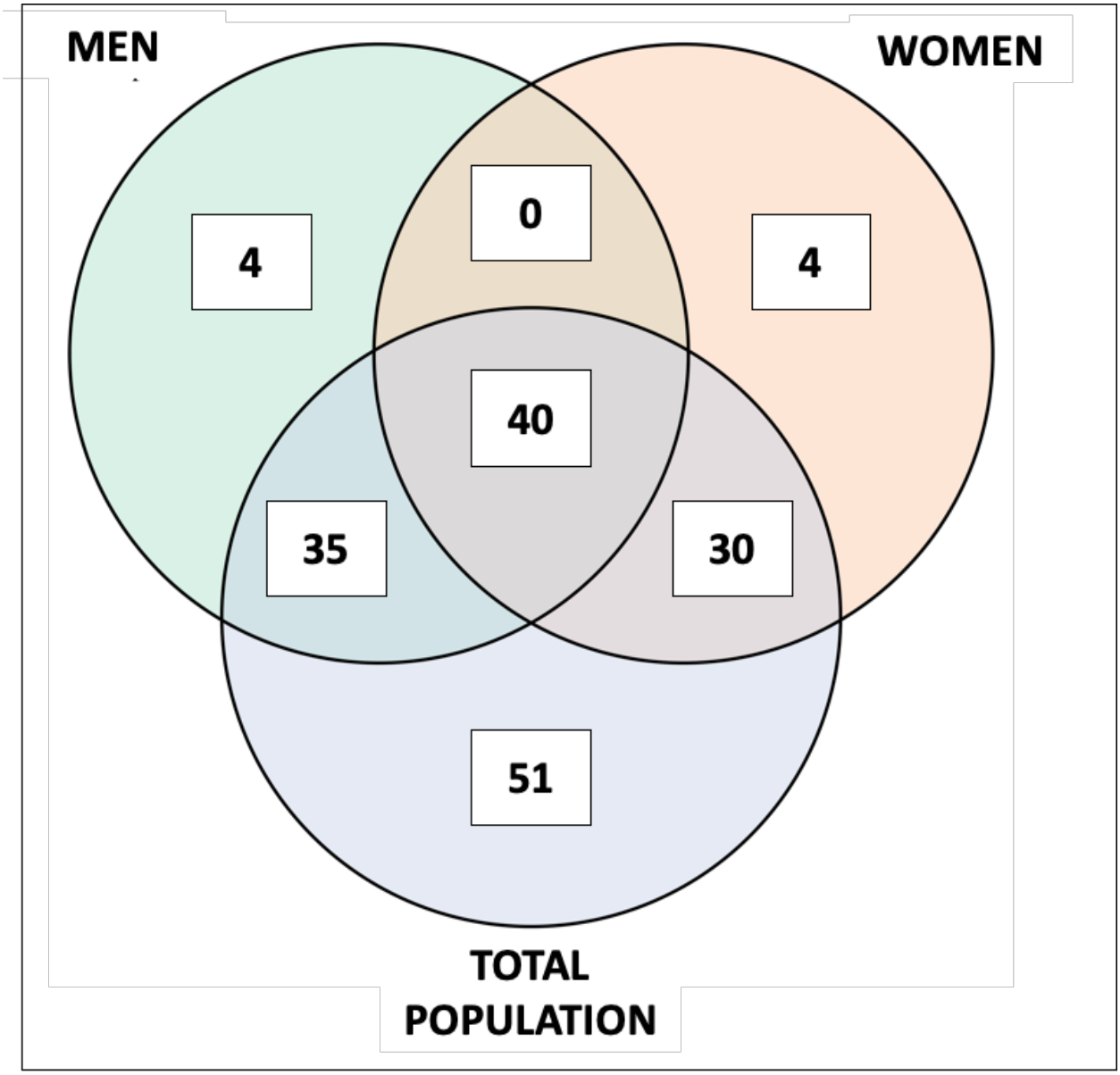
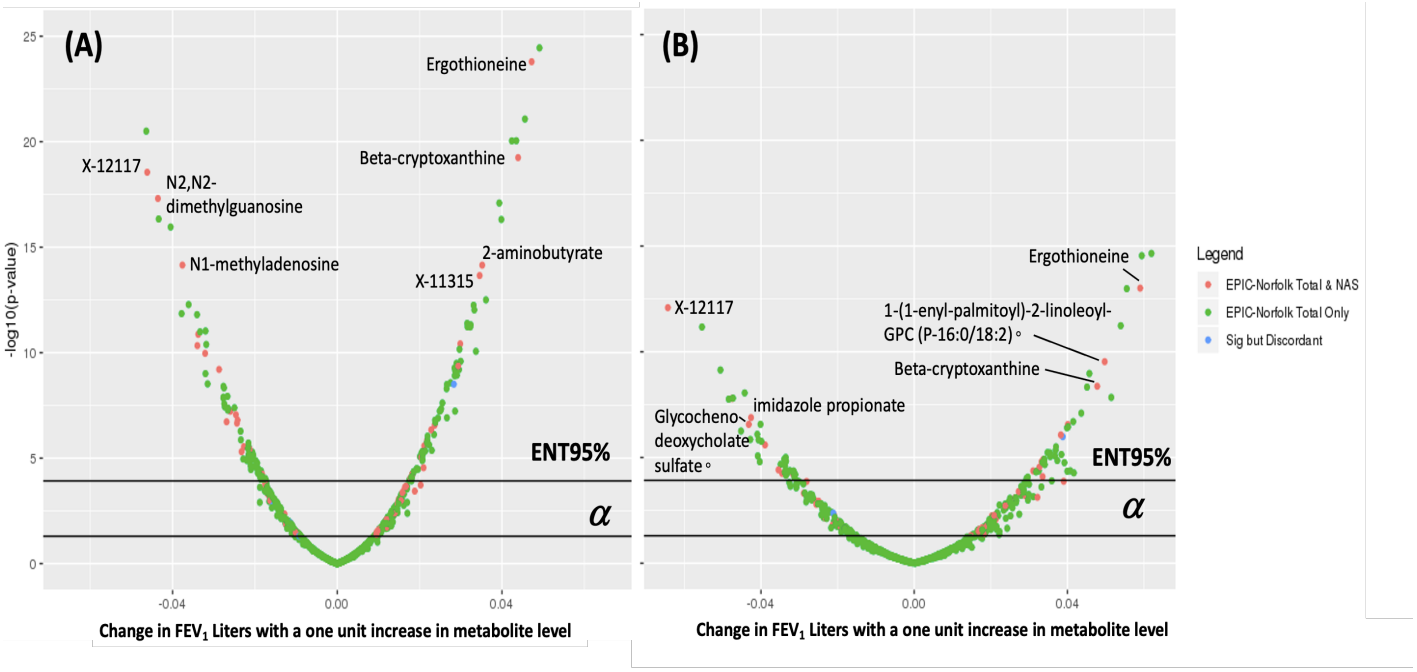


Figure E2. Volcano plot demonstrating the FEV_1-metabolite associations in EPIC Norfolk in (A) the total population and (B) Men only, indicating those which replicated in the NAS

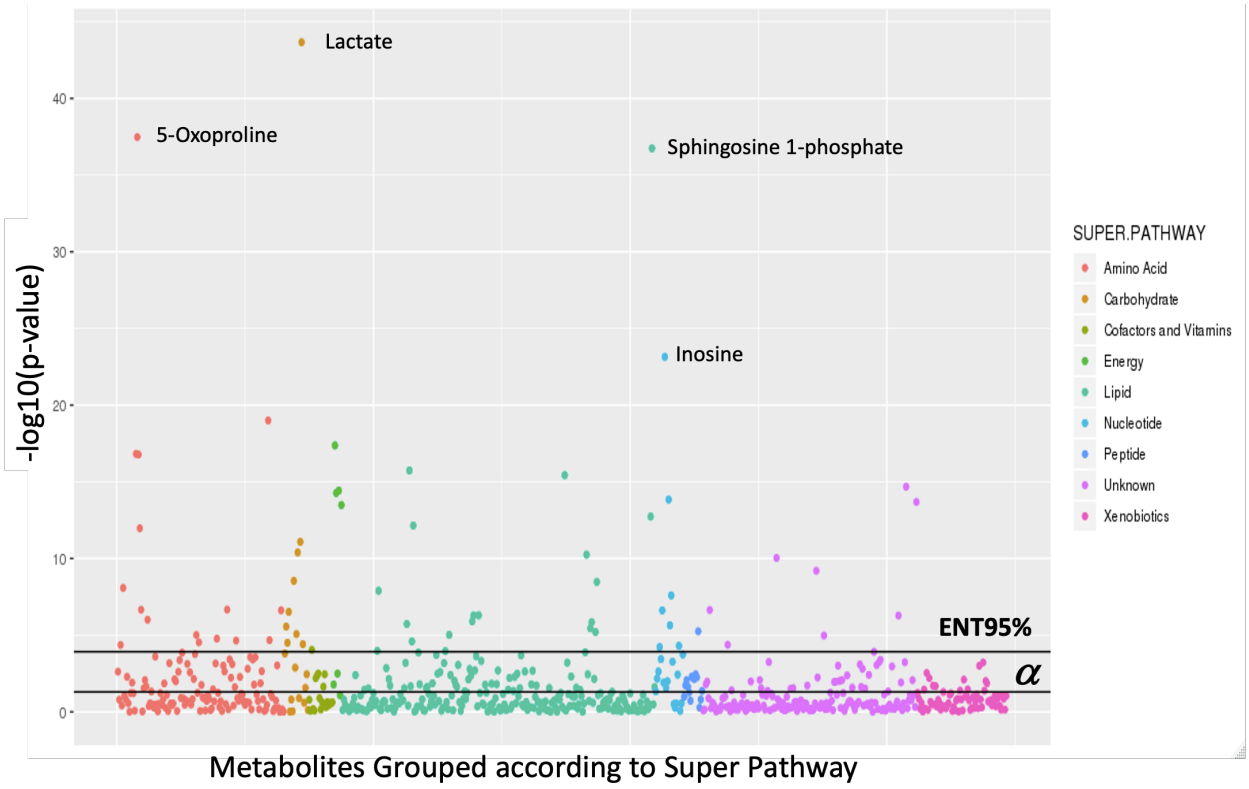


A metabolite is considered replicated (pink marker) if the p-value in the NAS is <0.05 and the direction of effect is concordant with EPIC-Norfolk. The top replicated metabolites are indicated by name.

°Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters

Metabolites with the format Xnnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

Figure E3: Manhattan plot demonstrating the strength of association between FEV_1/FVC ratio and 693 metabolites in EPIC-Norfolk



Top metabolite hits are identified according to their common name.
The black vertical lines indicate significance thresholds

Figure E4: Venn Diagram displaying the cross-over in significant metabolites for FEV_1/FVC between the Total EPIC-Norfolk population, Men only and Women only at a significance threshold of p<ENT95%

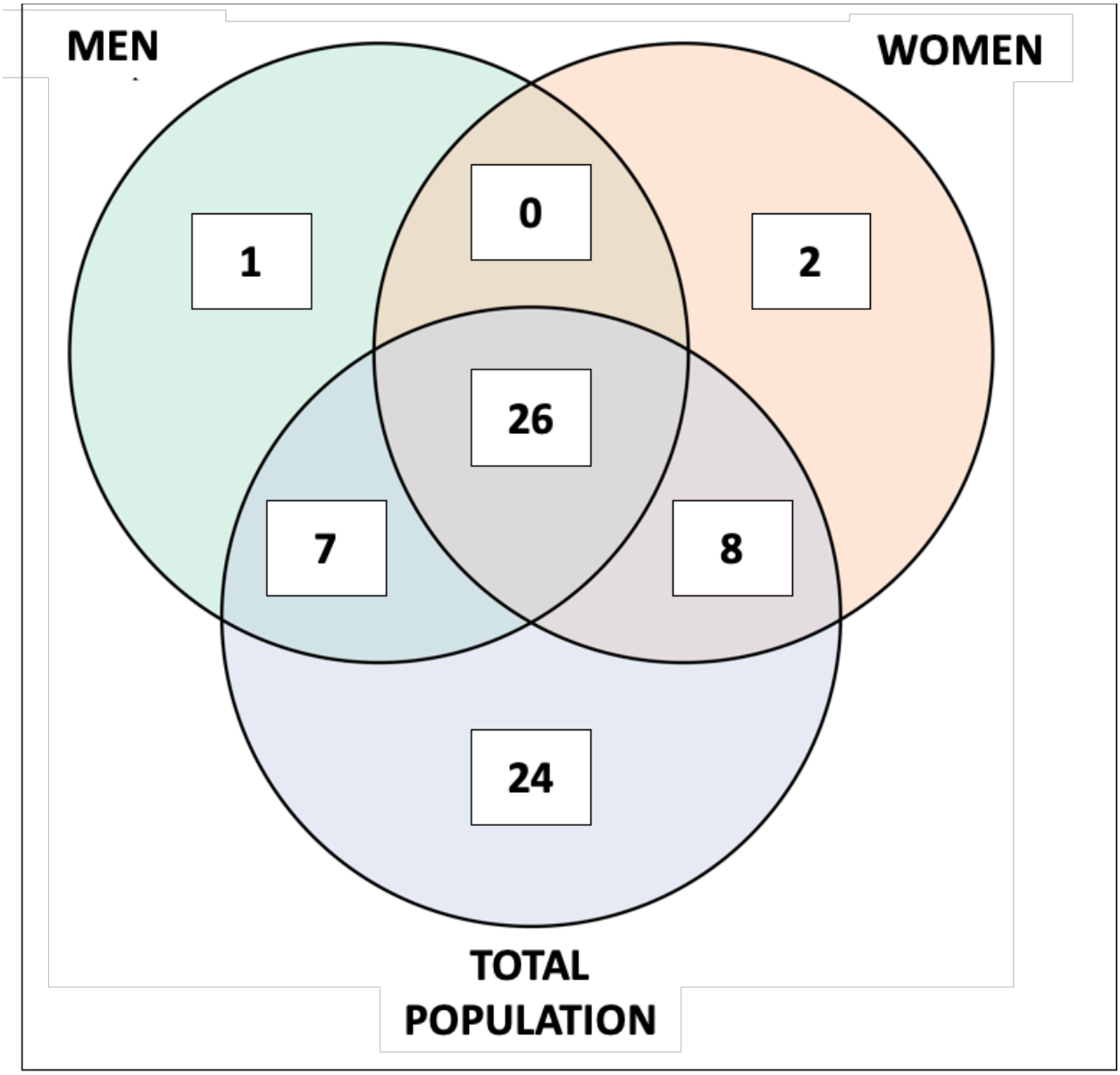
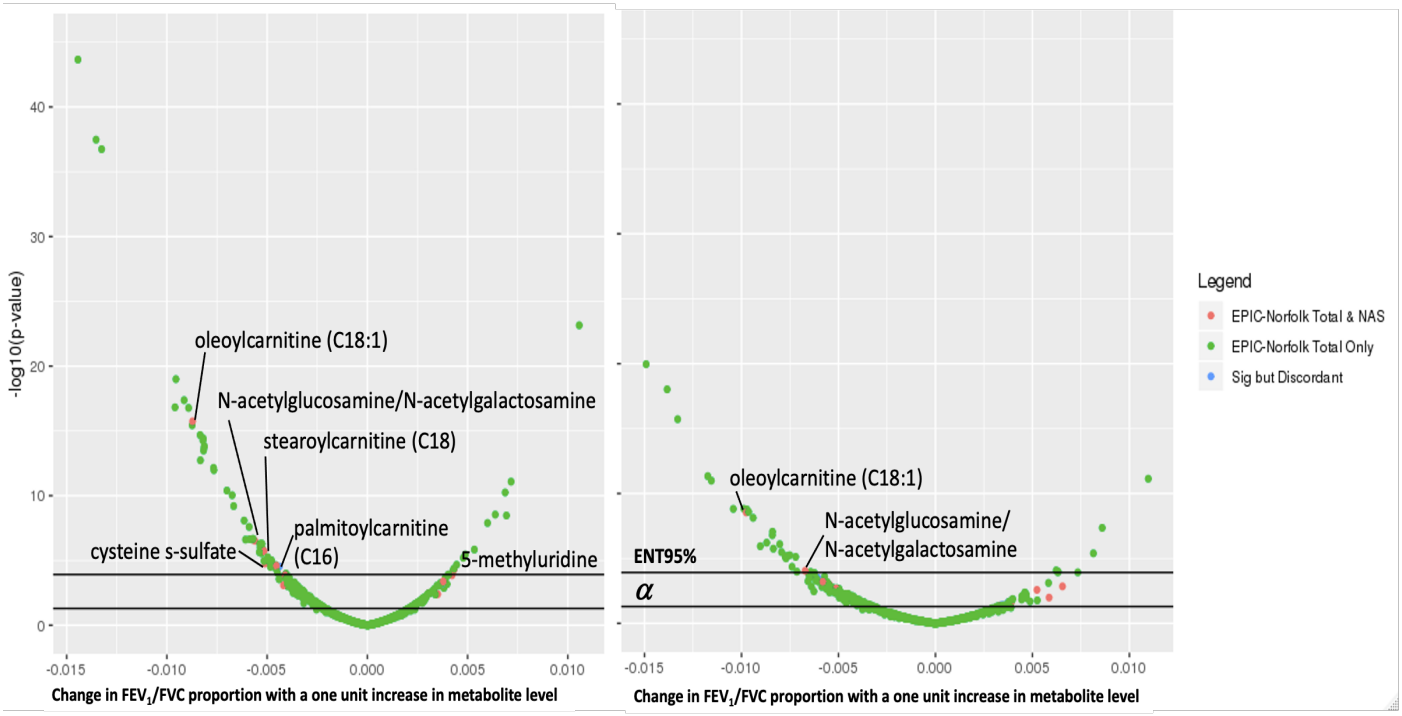


Figure E5. Volcano plot demonstrating the FEV_1/FVC ratio-metabolite associations in EPIC Norfolk in (A) the total population and (B) Men only, indicating those which replicated in the NAS



A metabolite is considered replicated (pink marker) if the p-value in the NAS is <0.05 and the direction of effect is concordant with EPIC-Norfolk. The top replicated metabolites are indicated by name.

Figure E6: Direction of Effect and P-value in the validated lung function associated metabolites in the EPIC-Norfolk Males analysis: (A) EPIC-Norfolk Males and NAS for 18 FEV1 validated metabolites, (B) EPIC-Norfolk Males and NAS for 2 FEV_1 validated metabolites

(A)

Metabolite	EPIC: FEV1	NAS: FEV1	EPIC: FEV1/FVC	NAS: FEV1/FVC
X - 12117	8.2E-13	7.3E-04	1.1E-02	2.6E-01
imidazole propionate	1.6E-08	3.0E-02	1.7E-01	6.8E-01
glycochenodeoxycholate sulfate	1.3E-07	1.6E-02	6.2E-01	2.9E-01
N2,N2-dimethylguanosine	2.7E-07	6.6E-05	3.2E-04	2.3E-01
N1-methyladenosine	2.5E-06	2.5E-02	7.5E-02	6.4E-01
pseudouridine	3.9E-05	3.6E-03	2.2E-02	9.7E-01
vanillylmandelate (VMA)	5.6E-05	2.0E-04	1.0E-04	6.0E-02
N-acetylputrescine	6.0E-05	5.3E-04	5.7E-03	6.4E-01
ergothioneine	9.8E-14	2.4E-02	8.5E-01	4.2E-01
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)◦	2.9E-10	2.9E-03	4.4E-02	1.9E-01
beta-cryptoxanthin	4.3E-09	2.4E-04	3.8E-01	2.2E-01
4-allylphenol sulfate	2.8E-07	2.7E-02	9.5E-01	1.1E-02
X - 18901	8.5E-07	3.5E-02	5.8E-03	3.8E-01
linoleoylcarnitine (C18:2)◦	1.3E-05	2.3E-02	5.7E-06	1.9E-01
2-aminobutyrate	1.6E-05	1.8E-02	2.3E-01	2.9E-01
X - 11315	2.9E-05	5.9E-03	7.0E-01	4.0E-01
linoleoyl ethanolamide	4.3E-05	4.7E-03	1.0E-04	9.6E-01
threonine	8.0E-05	1.1E-02	7.0E-03	2.4E-01

(B)

Metabolite	EPIC: FEV1/FVC	NAS: FEV1/FVC	EPIC: FEV1	NAS: FEV1
oleoylcarnitine (C18:1)	2.9E-09	3.2E-02	3.5E-01	6.5E-02
N-acetylglucosamine/N-acetylgalactosamine	8.9E-05	4.0E-02	2.5E-05	2.5E-01

Legend
Inverse Association with Lung Metric
Positive Association with Lung Metric

To allow comparison between FEV_1 and FEV_1/FVC, the direction of effect and p-value for both are given for each set of metabolites in each cohort

◦ Indicates a metabolite has not been confirmed using an analytical standard, but Metabolon are confident in its identify based on its analytical parameters

Metabolites with the format X-nnnnn are of unknown identity, but can be tracked and quantified and therefore Metabolon are confident they represent biologically relevant molecules and not analytical artifacts.

1. Jakes RW, Day NE, Patel B, Khaw KT, Oakes S, Luben R, Welch A, Bingham S, Wareham NJ: **Physical inactivity is associated with lower forced expiratory volume in 1 second : European Prospective Investigation into Cancer-Norfolk Prospective Population Study**. *Am J Epidemiol* 2002, **156**(2):139-147.
2. Sparrow D, O'Connor G, Colton T, Barry CL, Weiss ST: **The relationship of nonspecific bronchial responsiveness to the occurrence of respiratory symptoms and decreased levels of pulmonary function. The Normative Aging Study**. *Am Rev Respir Dis* 1987, **135**(6):1255-1260.
3. **Standardization of Spirometry**. *American Review of Respiratory Disease* 1987, **136**:1285-1298.
4. **Standardization of Spirometry**. *American Review of Respiratory Disease* 1994, **152**:1107-1136.
5. Kelly RS, Bayne H, Spiro A, 2nd, Vokonas P, Sparrow D, Weiss ST, Schwartz J, Nassan FL, Lee-Sarwar K, Huang M *et al*: **Metabolomic Signatures of Lead Exposure in the VA Normative Aging Study**. *Environ Res* 2020:110022.
6. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL *et al*: **Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)**. *Metabolomics* 2007, **3**(3):211-221.