A cross-omics integrative study of metabolic signatures of Chronic obstructive pulmonary disease

Ivana Nedeljković, Lies Lahousse, Maaike de Vries, Jun Liu, Marita Kalaoja, Emmi Tikkanen, Peter Würtz, Judith M. Vonk, Diana A. van der Plaat, Cleo C. van Diemen, Ashley van der Spek, Alexandra Zhernakova, Jingyuan Fu, Mohsen Ghanbari, Mika Ala-Korpela, Johannes Kettunen, Aki S. Havulinna, Markus Perola, Veikko Salomaa, Lars Lind, Johan Ärnlöv, Bruno H.C. Stricker, Guy G. Brusselle, H. Marike Boezen, Cornelia M. van Duijn*, Najaf Amin*.

**Both authors contributed equally.*

This chapter is submitted for publication.

The supplemental material for this paper is available at:

https://drive.google.com/file/d/1O9mbkPsUclOuOdR7IqxWxPqS3rzf2U8I/view?usp=sharing

ABSTRACT

Importance: Chronic obstructive pulmonary disease (COPD) is a disorder characterized by persistent and progressive airflow limitation. Beyond lung function impairments, metabolic changes in the circulation have been reported but their relation to the risk factors and prognosis of COPD has not been addressed. *Objective:* To identify metabolic signatures for COPD. *Design:* A comprehensive metabolic study of COPD and lung function was conducted in two large population-based studies in the Netherlands, the Rotterdam Study and the Erasmus Rucphen Family study. Significant findings were replicated in Lifelines-DEEP study, FINRISK and Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) studies. The data were integrated with publicly available data sets. *Setting:* Multicenter, population-based setting. *Participants:* A random sample of 5,557 individuals was included in the discovery cohort, whose lung function was characterized by spirometry. *Exposure:* Circulating levels of metabolites as measured by proton Nuclear Magnetic Resonance Spectroscopy. *Main outcomes and measures:* The primary outcome was COPD, defined as the ratio of Forced Expiratory Volume in 1 second (FEV_1) to Forced Vital Capacity (FVC) <0.7. Secondary outcomes were FEV_1/FVC (continuous variable), smoking status and pack-years of smoking. Tertiary outcome was all cause mortality. *Results:* There were 602 cases of COPD and 4955 controls used in the discovery meta-analysis. Our logistic regression results showed that higher levels of plasma Glycoprotein acetyls (GlycA) were significantly associated with COPD (OR=1.16, P=5.6×10^{-4} in the discovery and OR=1.30, P =1.8×10^{-6} in the replication sample). Smoking status (P=1.3×10^{-22}) and pack-years of smoking (P= 2.5×10^{-16}) were significantly associated with levels of GlycA. A bi-directional two-sample Mendelian randomization analysis has suggested that circulating blood GlycA is not causally related to COPD, but that COPD is causally associated with GlycA. Using the prospective data of the same sample of Rotterdam Study in Cox-regression, we show that circulating GlycA levels are predictive biomarker of COPD risk (HR=1.42, 95%CI 1.24-1.63, P=7.61×10^{-7}, comparing those in the highest and lowest quartile of GlycA) but are not significantly associated with mortality in patients (HR=1.06, SE=0.06, P=0.31). *Conclusions and Relevance:* Our study shows that circulating blood GlycA is a biomarker of preclinical COPD pathology.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a progressive inflammatory lung disease and currently the third leading cause of death worldwide.[1,2] COPD is characterised by chronic airway inflammation, airway remodelling and airflow limitation.[3] A reduced ratio of the Forced Expiratory Volume in 1 second (FEV_1) to Forced Vital Capacity (FVC) is a measure of obstruction and is used to diagnose COPD but also as an endophenotype for preclinical lung function.[3,4] Smoking is the most important risk factor for COPD and related impaired lung function.[2] COPD is a complex heterogeneous disease in which systemic features beyond airflow obstruction, including systemic inflammation, oxidative stress, muscle dysfunction, cachexia and vascular pathology occur.[5,6] Understanding these systemic effects may give new insights in the pathogenesis and progression of COPD but may alternatively yield important clues for preventive research.

Recent developments in metabolomics have made it possible to investigate the associations between circulating metabolites and the systemic effects in COPD. Glycoprotein acetyls (GlycA) were found to be predictive for several chronic diseases, among which COPD.[7] In a previous metabolomics study using proton Nuclear Magnetic Resonance (^{1}H-NMR), lower levels of lipoproteins, N,N-dimethylglycine and higher levels of glutamine, phenylalanine, 3-methylhistidine and ketone bodies were found in the circulation of ex-smoking COPD patients compared with ex-smoking controls.[8] In severe COPD patients, branched chain amino acids (BCAAs) were found to be lower, compared with controls.[8] Interestingly, BCAAs, 3-methylhistidine, ketone bodies, and triglycerides were negatively correlated with cachexia and positively correlated with systemic inflammation,[8] but these findings have not been replicated. Another question that remains to be answered is whether the metabolic changes are a cause or a consequence of COPD. If the latter is true, the metabolites may be relevant for the disease progression and prognosis.

To answer these questions, we performed a comprehensive integrative metabolic analysis to identify plasma metabolic measures associated with COPD and lung function levels, defined as FEV_1/FVC, using the NMR approach in a set of large epidemiological studies, in depth characterized for genetic and environmental risk factors. The discovery phase of the study was conducted in two population-based studies in the Netherlands, the Rotterdam Study (RS)[9] and the Erasmus Rucphen Family study (ERF).[10,11] A replication meta-analysis was conducted in Lifelines-DEEP study (LLD),[12] two cohorts of FINRISK study[13,14] and Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study.[15,16]

METHODS

Study population

Studies included in the discovery sample

The RS is a population-based study of 14,926 people older than 45 years, from the Ommoord area of Rotterdam, incorporating three cohorts: RS-I (established in 1989), RS-II (2000) and RS-III (2006), with multiple subsequent visits.[9] Participants filled in questionnaires, underwent physical examination and provided fasting blood samples at each visit. For this analysis, three independent samples from different RS cohorts were enrolled: Sample 1) visit 4 of RS-I (RS-I-4); sample 2) a combined sample, which we collectively call RS-E5 in this manuscript, comprising of visit 5 of RS-I (RS-I-5), visit 3 of RS-II (RS-II-3), and visit 2 of RS-III (RS-III-2); and sample 3) another independent set from RS-III-2.

ERF is a population-based study from the south-west of the Netherlands. It is a genetically isolated population comprising 3,465 living descendants of 22 couples from the 19th century and their spouses.[10] The baseline data collection was performed in 2002-2005 when participants underwent physical examinations, provided blood samples and completed questionnaires. A follow-up of the participants was performed in 2015-2018, reviewing the medical records at the general practitioner's office.

Both RS and ERF were approved by the Medical Ethics committee of the Erasmus Medical Center and all participants gave informed consent for participation in the study and for evaluation of the available information from their physicians.

Studies included in the replication sample

LLD is a sub-cohort of the large general population-based cohort study LifeLines, which was initiated to study genes, exposures and their interactions in the etiology of complex multifactorial diseases and healthy ageing.[17,18] LLD consists of 1,500 LifeLines participants who registered at the LifeLines research site in Groningen between April and August 2013. These subjects gave additional biological materials, including blood samples for metabolite and inflammation profiling, and extensive phenotype information.[12] Metabolic and lung function data were available for 717 LLD individuals and these subjects are included in the current study. LLD was approved by the ethics committee of the University Medical Center Groningen and all participants signed an informed consent prior to enrolment.

The FINRISK cohorts comprise cross-sectional population surveys that are carried out every 5 years since 1972, to assess the risk factors of chronic diseases (e.g. CVD,

diabetes, obesity, cancer) and health behaviour in the working age population (25-74 years of age), in 3-5 large study areas of Finland. The FINRISK surveys are conducted by the National Institute for Health and Welfare, THL (previously National Public Health Institute, KTL). Extensive information from each participant was collected at baseline via questionnaire and health examination with blood collection. The cohorts were followed up by linking them to national health registers. The cohorts FINRISK 1997 (total of 6898 participants) and an extension of FINRISK 2007, known as DIetary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) study[19] (total of 4600 participants) are included in our replication sample for COPD analysis. The FINRISK 1997 study was approved by the Ethical Committee of the National Public Health Institute, while the DILGOM study was approved by the Coordinating Ethical Committee of the Hospital District of Helsinki and Uusimaa. All participants have signed an informed consent, allowing the use of their data and samples for studying environmental and genetic risk factors of chronic diseases.

The PIVUS study started in 2001 with the aim to investigate endothelial function as a prospective cardiovascular risk factor in elderly subjects. A random sample of Uppsala city residents were invited from the register of inhabitants within one month following their 70th birthday. No exclusion criteria were applied except that participants were required to have a Swedish identification number. In PIVUS, 1,016 agreed to participate, resulting in a participation rate of 50.1% of all invited, whereof 51.5 % were female. The participants have undergone a range of physical measurements, and given information about their medical history, lifestyle habits and regular medication. In addition, blood samples were drawn. The Ethics Committee of the University of Uppsala approved the study and the participants gave informed consent (approval number 00-419).

Assessment of COPD status and lung function measurements

COPD in the RS was defined as pre-bronchodilator $FEV_1/FVC<0.7$, assessed either by spirometry at the RS research center or by reviewing medical histories of the participants. Spirometry was performed in the RS by trained paramedical personnel, according to the guidelines of the American Thoracic Society/European Respiratory Society (ATS/ERS). When spirometry measurements were absent or uninterpretable, all files from specialists and general practitioners were reviewed to set a diagnosis of COPD. In total, this analysis included 541 COPD subjects and 4,407 subjects without COPD from all three RS cohorts.

For the ERF study, the doctor's diagnosis of COPD was confirmed by reviewing medical records based on $FEV_1/FVC<0.7$, with or without medication use. If the information on FVC was missing, the following criteria for COPD were used: $FEV_1<80\%$

of predicted, use of respiratory medication and a COPD diagnosis mentioned in the report of the respiratory specialist to the general practitioner. In total, 61 COPD subjects and 548 subjects without COPD were included from ERF. For ERF participants, we did not have lung function measurements at the time of the metabolomics measurements, so we did not include this cohort in the FEV_1/FVC analysis.

For LLD, COPD was defined as a FEV_1/FVC below 70%. Pre-bronchodilator spirometry was performed according to the ATS/ERS guidelines using a Welch Allyn Version 1.6.0.489, PC-based Spiroperfect with CA Workstation. Technical quality and results were assessed by well-trained assistants and abnormal results were re-evaluated by lung physicians.

In the FINRISK study the COPD information was extracted based on diagnoses and reimbursement information from the National health register, which include the Drug Reimbursement Register, the Care Register for Health Care, the Register for Prescribed Drug Purchases, the Causes-of-Death Register, and the Cancer Register. The maximum retrospective time period available for obtaining prevalent disease events was 20 years for DILGOM and 10 years for FINRISK97.

In the PIVUS study FEV_1 and FVC were assessed with spirometry using a Vitalograph Alpha spirometer (Vitalograph Ltd. Buckingham, UK) according to the American Thoracic Society recommendations.[20,21] The best value of three acceptable recordings was used. FEV_1 and FVC expressed as percent of predicted values, were adjusted for age, sex and height according to Hedenström's formula.[22,23]

Assessment of blood metabolites

Metabolic profiling in RS, ERF and LLD was done as part of the 4th Rainbow Project of the BioBanking for Medical Research Infrastructure of the Netherlands (BBMRI-NL) (https://www.bbmri.nl/omics-metabolomics/). To quantify the metabolite biomarkers from all samples fasting EDTA plasma samples were used for quantitative high-throughput ^{1}H-NMR metabolomics platform (Nightingale Ltd, Helsinki, Finland). Details and advantages of the NMR-based metabolomics analyses using plasma were described elsewhere.[24,25] Using this method, we were able to quantify a wide range of blood metabolite biomarkers such as lipoprotein fractions, amino-acids, cholesterol levels, glycerides, phospholipids, fatty acids, ketone bodies and metabolites related to inflammation and glycolysis. In total, 161 metabolites, overlapping between RS and ERF, were used in the discovery analysis.

Statistical analyses

Association of COPD and FEV_1/FVC with metabolites

The distributions of all metabolites were inspected for normality and natural logarithm or rank transformations were applied. Per cohort, we used transformed metabolite levels as independent variable and COPD status or FEV_1/FVC as dependent variables in logistic and linear regression models, respectively. The models were adjusted for age, sex, BMI (kg/m^2), lipid lowering medication use and smoking status (current, ex- or never smokers). For the discovery sample, the results from ERF, RS-I-4, RS-E5 and RS-III-2 were meta-analysed using fixed effect models in "*METAL*" software.[26] As the metabolites are known to be highly correlated, we applied the method by Li and Ji[27] to assess the number of independent metabolites. Using this method, we calculated that for the 161 metabolites, the number of independent tests was 45, which resulted in the Bonferroni significance threshold of P=0.001 (0.05/45). Significant metabolites were further tested for replication in the meta-analysis of LLD, FINRISK1997 and DILGOM studies for the COPD analysis and of LLD and PIVUS studies for the FEV_1/FVC analysis. Again, the same regression models were used for the fixed effect meta-analysis in "*METAL*" software.

For significant COPD metabolites, we investigated the odds ratios per quartile of the metabolite distribution in the discovery sample. To investigate the effects of smoking on this association, we used two logistic regression models, one adjusted for age, sex, BMI and lipid lowering medication use, and a second model additionally adjusted for smoking status (current, ex- and never smokers). Results from each cohort were combined using inverse-variance weighted fixed effects meta-analysis in "*rmeta*" package in R.

Association of smoking with metabolites

We further tested if the replicated metabolites from the COPD and FEV_1/FVC analyses (as dependent variables) were associated with smoking status (current, ex- and never smokers) and pack-years of smoking. We used models adjusted for age, sex, BMI and lipid lowering medication use in the discovery sample. Associations with smoking status were further tested for replication in the FINRISK1997 and DILGOM studies, using same models. All analyses were performed in R (version 3.2.1.). Replication studies did not have pack-years data to investigate further. Next, for comparison, we tested the same models in the discovery cohort after excluding the COPD cases.

Association of genetic variants with metabolites

We have used a bi-directional approach in which we examined whether: 1) the genetic determinants of the significant metabolic measures are associated with COPD and lung function, which would lead to the conclusion that the metabolites are most likely driving the disease; 2) the genetic determinants of COPD are associated with significant metabolites when the metabolites would most likely be altered as an integral part of the disease pathophysiology and may be biomarkers. In these analyses we use the genes as instrumental variable and a method which is referred to in genetics as a bi-directional Mendelian Randomization (MR) approach.[28] MR was conducted using "*gtx*" package in R.[29] To maximize the statistical power of the study[28] we used the genetic information from previously published genome-wide association studies (GWAS) on metabolites (Model 1)[25] and COPD (Model 2).[30] Genetic risk score (GRS), summarizing the effect of the SNPs genome-widely associated with either the significant metabolites or COPD, were used as instrumental variable. In GRS we included unique SNPs (mapped to human genome build hg19) in low linkage disequilibrium based on the data in ERF study (within 500Kb and R^2<0.05). MRs were performed with GRS explaining >1% of the variance, because the power of the MR using GRS that explains a lower proportion of the variance is too low to yield trustable results. To control for pleiotropic effects, we checked the heterogeneity of the SNPs included in the GRS and excluded the SNPs which were also genome-wide significantly associated with the outcome.

Association of metabolites with mortality

To investigate whether metabolites have a clinical utility in predicting COPD, we constructed classical receiver operating curves (ROC) and compared area's under the curve (AUC).[31] To further investigate whether the identified metabolites may act as biomarker of the disease prognosis, we performed a survival analysis in SPSS, similar to the previous study by Fischer and colleagues for all-cause mortality, ignoring any underlying morbidity.[32] To check whether the metabolites associated with mortality in COPD patients, we performed the Cox proportional hazards model in three RS cohorts. Analyses were adjusted for age at sampling, sex and smoking. We further performed a similar analysis using four quartiles of metabolite, testing in COPD cases and controls.

RESULTS

Descriptive characteristics of the samples

Descriptive characteristics of all cohorts used in the analysis are presented in **Table 1**.

Table 1. Discovery population characteristics per cohort

	Discovery cohorts				Replication cohorts			
Study	**ERF**	**RS-I-4**	**RS-E5**	**RS-III-2**	**LLD**	**FINRISK97**	**DILGOM**	**PIVUS**
N	609	2777	686	1485	717	6898	4600	854
Age, mean (sd)	49.0 (13.3)	74.8 (6.5)	68.4 (5.7)	62.8 (5.8)	46.0 (14.3)	48.0 (13.1)	52.3 (13.5)	70 (0)
Women, % (n)	55.8 (340)	58.2 (1615)	57.6 (395)	57.8 (859)	56.3 (404)	51.6 (3561)	53.4 (2458)	48.2 (412)
COPD cases, % (n)	10.0 (61)	12.1 (336)	10.3 (71)	9.0 (134)	13.8 (99)	0.6 (43)	0.8 (35)	NA
FEV_1/FVC, mean (sd), % of all	NA	0.73 (0.08), 48.8	0.76 (0.07), 91.3	0.77 (0.07), 91.9	0.77 (0.08), 100	NA	NA	0.76 (0.11), 100
BMI, mean (sd)	27.2 (4.85)	27.4 (4.1)	27.8 (4.3)	27.4 (4.5)	25.4 (4.1)	26.6 (4.5)	27.2 (4.8)	27.1 (4.26)
Current smokers, % (n)	43.3 (264)	12.6 (349)	9.5 (65)	13.7 (203)	20.5 (147)	23.9 (1648)	17.6 (810)	10.2 (87)
Ex-smokers, % (n)	30.0 (183)	56.1 (1559)	57.0 (391)	50.2 (746)	NA	22.9 (1577)	26.3 (1210)	41.5 (354)
Never smokers, % (n)	26.6 (162)	31.3 (869)	33.5 (230)	36.1 (536)	79.4 (570)	53.2 (3673)	56.1 (2580)	48.2 (412)
Pack-years of smoking, mean (sd), % of all[a]	24.9 (20.4) 72.7	24.2 (23.4), 64.7	22.0 (20.8) 66.3	19.5 (20.3) 63.8	NA	NA	NA	NA
Lipid lowering medication users, % (n)	12.3 (75)	22.4 (621)	32.5 (223)	22.2 (329)	3.9 (28)	3.4 (237)	15.7 (721)	16.5 (141)

RS-E5: consists of RS-I-5, RS-II-3 and RS-III-2; [a] Pack-years calculated in current and ex-smokers only, so "% of all" excludes never smokers; NA - not applicable;

Comparing the discovery cohorts, ERF participants were younger (mean age 49.0±13.3) and had a higher percentage of current smokers compared to the participants of the three RS cohorts (RS-I-4 mean age 74.8±6.5; RS-E5 mean age 68.4±5.7; RS-III-2 mean age 62.8±5.8). The RS cohorts had a higher percentage of lipid lowering medication users, compared to ERF (**Table 1**). The mean FEV_1/FVC and BMI were comparable across the studies. Descriptive characteristics for COPD cases and

subjects without COPD separately in the discovery cohorts are provided in **eTable 1** in the Supplement**.** In general, COPD subjects were older and more often smokers compared with subjects without COPD.

Association of COPD and FEV_1/FVC with metabolites

In the discovery sample, six plasma metabolites were associated with COPD at a significance level of 5% (**Table 2, Figure 1**).

Table 2. Metabolites associated with COPD in the discovery and replication studies

	Discovery meta-analysis						Replication meta-analysis					
Metabolite	**B**	**SE**	**OR**	**P-value**	**Direction[a]**	**N**	**β**	**SE**	**OR**	**P-value**	**Direction[b]**	**N**
GlycA	0.152	0.044	1.16	**5.6×10^{-4}**	++++	5557	0.266	0.053	1.30	**1.8×10^{-6}**	+++	12205
3-hydroxybutyrate	0.122	0.041	1.13	0.003	++++	5002	-0.031	0.057	0.97	0.662	+--	12173
Histidine	-0.097	0.047	0.91	0.037	----	5534	-0.153	0.063	0.86	0.020	---	12200
Free cholesterol in med. HDL	0.099	0.049	1.10	0.045	+-++	5557	0.004	0.063	1.00	0.867	+--	12208
Acetoacetate	0.084	0.042	1.09	0.047	++-+	5551	-0.061	0.059	0.94	0.360	---	12204
18:2, linoleic acid	-0.095	0.048	0.91	0.049	+---	5546	-0.036	0.057	0.96	0.238	+-+	12167

Model adjusted for age, sex, BMI, lipid lowering medication and smoking status; GlycA – Glycoprotein acetyls; HDL - high density lipoprotein; β - effect size; SE - standard error; OR - odds ratio; Direction - direction of the effect in individual studies; N - meta-analysis sample size; [a] Direction of the effect in the discovery studies in order: ERF, RS-III-2, RS-E5, RS-I-4; [b] Direction of the effect in the replication studies in order: DILGOM, FINRISK 1997, LLD; In bold: significant results (P<0.001).

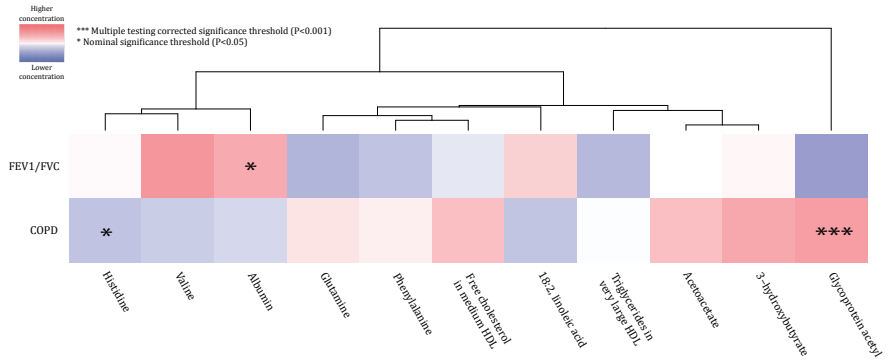

Figure 1. Top metabolites associated with COPD and/or FEV_1/FVC. Colors represent standardized effect estimates of the metabolite association with corresponding trait (COPD, FEV_1/FVC). Red color represents the trait associated with an increase in metabolite concentration, while blue represents a decrease. For replicated metabolites, replication significance threshold is shown with stars: *P<0.05 and ***P<0.001. HDL – high-density lipoprotein.

At nominal significance, higher levels of GlycA (OR=1.16; P=5.6×10^{-4}), 3-hydroxybutyrate (OR=1.13; P=0.003), free cholesterol in medium HDL (OR=1.10; P=0.045) and acetoacetate (OR=1.09; P=0.047) were associated with a higher prevalence of COPD. Higher levels of histidine and 18:2 linoleic acid (OR=0.91 for both, P=0.04 and P=0.05 respectively) were associated with a lower prevalence of COPD. When taking into account the multiple testing correction threshold, only GlycA was significantly associated with COPD (P= 5.6×10^{-4}).

We tested all six metabolites for replication in the independent samples. The association of higher levels of GlycA with COPD was significantly replicated (OR=1.30, P=1.75×10^{-4}) in the 12,205 participants of the replication sample, after multiple testing correction.

Findings for the FEV_1/FVC ratio were not consistent over the discovery and replication studies. Adjusting for multiple testing, we found in the discovery cohorts that lower levels of valine (β=0.005, P=2.5×10^{-4}) and higher levels of GlycA (β=-0.005, P=4.5×10^{-4}) were associated with a lower FEV_1/FVC ratio (**Table 3, Figure 1**). Other metabolites that reached nominal significance in the discovery included albumin which was positively associated with FEV_1/FVC, and glutamine, triglycerides in very large HDL and phenylalanine which were negatively associated with FEV_1/FVC (**Table 3, Figure 1**).

Table 3. Top metabolites associated with FEV_1/FVC - Results of the discovery and replication studies

	Discovery meta-analysis					Replication meta-analysis				
Metabolite	**B**	**SE**	**P-value**	**Direction[a]**	**N**	**β**	**SE**	**P-value**	**Direction[b]**	**N**
Valine	0.005	0.001	**2.5×10-4**	+++	3324	-0.0015	0.0023	0.5314	-+	1460
GlycA	-0.005	0.001	**4.5×10-4**	---	3324	-0.0010	0.0022	0.6438	--	1463
Albumin	0.004	0.001	0.0047	+++	3324	0.0045	0.0021	0.0353	++	1463
Glutamine	-0.003	0.001	0.0097	---	3323	0.0029	0.0023	0.1923	++	1393
Triglycerides in very large HDL	-0.003	0.001	0.0160	---	3324	0.0031	0.0022	0.1491	++	1469
Phenylalanine	-0.003	0.001	0.0334	---	3324	-0.0012	0.0023	0.5899	+-	1450

Model adjusted for age, sex, BMI, lipid lowering medication and smoking status; HDL - high density lipoprotein; β - effect size; SE - standard error; Direction - direction of the effect in individual studies; N - meta-analysis sample size; [a] Direction of the effect in the discovery studies in order: RS-III-2, RS-E5, RS-I-4; [b] Direction of the effect in the replication studies in order: LLD, PIVUS; In bold: significant results (P<0.001).

Only the association of FEV_1/FVC to albumin showed nominal significance in the replication samples (β=0.005, P=0.03), but none were significantly associated when considering multiple testing correction.

Association of smoking with metabolites

Although the above analyses were adjusted for smoking, metabolite levels may have changed as a consequence of smoking and may be an intermediary in the relation of smoking to COPD. We tested the association of smoking status and pack-years with the metabolite levels that were significantly associated with COPD. The results are presented in the **eTable 2** in the Supplement. GlycA was significantly positively associated with both smoking status (current, ex- or never smoker) and pack-years of smoking (β=0.15, P=1.31×10^{-22} and β=0.006, P=2.52×10^{-16}, respectively). **eTable 2** in the Supplement also shows the replication of the association between the metabolites and smoking status in the FINRISK1997 and DILGOM studies. Data on pack-years was not available in these replication studies. When excluding COPD cases from the discovery sample, the identified associations of GlycA with smoking status and pack-years attenuated, both in the effect size and p-value, yet remained significant (**eTable 3** in the Supplement).

Association of genetic variants of circulating GlycA and COPD with metabolites

Next, we performed a Mendelian Randomisation experiment investigating the hypothesis that: 1) GlycA is increasing the risk of COPD and therefore the genetic determinants of GlycA (used as instrumental variables) are also associated with COPD and 2) the opposite scenario is true in which (pre)clinical COPD pathology increases GlycA levels. The results of both models are presented in **Table 4.**

Table 4. Results of the bidirectional Mendelian randomization approach on GlycA and COPD.

Model	Exposure	Outcome	R^2	nSNP	β	SE	P-value	Ors	Phet
1	GlycA	COPD	0.023	9	-0.001	0.047	**0.988**	13.6	0.09
2	COPD	GlycA	0.03	7	0.165	0.053	**0.0018**	9.6	0.14

R^2 - the explained variance in the exposure by applied genetic risk score; nSNP - number of SNPs used to construct the genetic risk score; β - the weighted effect of the genetic risk score of exposure on outcome; SE - standard error; Significance threshold = P-value < 0.05; Ors: heterogeneity test statistic; Phet: heterogeneity test P-value

The GRS for Model 1 included nine independent SNPs (R^2=0.023, **eTable 4** in the Supplement) and yielded no significant evidence for association (P=0.99). In Model 2, we found that genes associated with a higher risk of COPD are also associated with higher levels of GlycA (**Table 4**, P=0.002), suggesting that COPD pathology increased GlycA levels. This analysis is based on seven independent SNPs in the GRS (R^2=0.03, **eTable 5** in the Supplement). No heterogeneity effect or potential pleiotropic SNPs were found in either model.

Is circulating GlycA predictive biomarker for COPD?

The question to answer next is whether GlycA in the circulation is a biomarker of early pathology thus can be used as a predictive or diagnostic biomarker or rather a prognostic biomarker for mortality in COPD patients. To this end, we performed an analysis in the Rotterdam Study in which we associated GlycA to the future risk of COPD. We determined the relative risk by quartile of the GlycA concentrations in the circulation, using the lowest quartile as a reference (**eTable 6** in the Supplement). Only incident patients are included in these analyses; prevalent COPD patients are excluded. Compared to the lowest quartile, those subjects in the highest quartile of GlycA had a 1.99-fold (95% Confidence interval: 1.52-2.60) higher risk of COPD, after adjustment for age, sex, BMI and lipid lowering medication. Smoking accounted for a part of the observed association between plasma GlycA and COPD attenuating the OR for those in the highest quartile of GlycA to 1.74, while the association remained significant (95% Confidence interval: 1.32-2.28). To test whether circulating GlycA adds to the predictive value, we compared the AUC curves for the models including: 1) age and sex (AUC=0.601); 2) age, sex and smoking (AUC=0.670) and 3) age, sex, smoking and circulating GlycA levels in blood (AUC=0.675). The AUC comparing model 2 and 1 shows that smoking is associated with an increase in AUC by 0.069. Adding circulating GlycA increased the AUC further by only 0.005 (**eFigure 1**).

Is circulating GlycA a prognostic biomarker for mortality in COPD?

Previous study has shown that GlycA is a predictor of mortality.[32] We confirm this in current study, after adjustment for age, sex and smoking (HR=1.16, $P=4.93\times10^{-9}$) (**eTable 7**). We therefore tested the hypothesis that GlycA is a marker of COPD related to future mortality. We first compared mortality across the quartiles of GlycA and found that those in the highest quartile have 1.42-fold (95% Confidence interval: 1.24-1.63, $P=7.61\times10^{-7}$) increased risk of mortality during follow-up compared to those in the lowest quartile (**eTable 7**). However, when stratifying these analyses by COPD status, we observed that this association is driven by controls (**eTable 7, eFigure 2**). In COPD patients, circulating GlycA levels were not significantly associated with mortality when studying GlycA as a continuous variable (HR=1.06, $P=0.31$) nor for those in the highest quartile (HR=1.02, $P=0.93$ in COPD cases). In those without COPD, the association of GlycA to mortality was stronger and significant (HR=1.18, $P=1.32\times10^{-9}$).

DISCUSSION

In our metabolome-wide discovery analysis we identified 11 plasma metabolites associated with COPD or lung function levels (FEV_1/FVC) at marginal significance. Of the 11, only higher levels of GlycA were significantly associated with COPD when adjusting for multiple testing and this is the only metabolite we could replicate in an independent sample. The association of GlycA with COPD remained significant when adjusting for smoking. GlycA levels in the circulation were significantly associated with smoking. Our MR analysis showed that the genetic predisposition to COPD associates with GlycA. Although GlycA was found to be a predictor of mortality in the general population,[33] the metabolite did not predict mortality in COPD patients.

The most convincing and interesting finding of our study is that of GlycA. We recently associated this metabolite with the incidence of a variety of disorders, including COPD in our study based on record linkage.[7] The record linkage study focussed specifically on the relation of GlycA with a wide variety of disorders. Using two population-based cohorts, we identified new associations with GlycA including incident COPD, alcoholic liver disease, chronic renal failure, glomerular diseases and inflammatory polyarthropathies. The GlycA associations were for a large part independent of that of high-sensitivity C-reactive protein (hsCRP), but GlycA and hsCRP also share contributions to mortality risk, suggesting chronic inflammation as the common pathway. GlycA is shown to be a biomarker for chronic inflammation, neutrophil activity and risk of future severe infection, even superior compared with CRP.[34,35]

The present study extends our findings published previously in that we have increased the number of NMR metabolites studied and found that GlycA is the only metabolite significantly associated with COPD when adjusting for multiple testing. In the present study we also have studied effects of GlycA beyond COPD, and found GlycA is consistently associated with smoking status and quantity (pack-years of smoking). Smoking is related to GlycA levels in the circulation but it does not explain the association between GlycA and COPD. This is compatible with the view that smoking, the major driver of COPD risk in the population, is associated with GlycA which in turn is associated with COPD risk. In the present paper we used data integration approach (MR) to test the hypothesis that GlycA increases the risk of COPD causally or rather is a biomarker that is part of the disease pathogenesis. The findings of the present paper suggest that the latter is more likely, as the genes associated with COPD also associate with GlycA levels. No marginally significant support was found for the hypothesis that GlycA is a determinant of COPD: the genes that are known to determine GlycA levels are not associated with the risk of COPD.

In the present paper we do not find evidence that GlycA is associated with COPD mortality. Such a relationship was seen in our findings for cardiovascular disease. GlycA not only increased the risk of incident cardiovascular disease[7,36] but was also associated with a 5-fold increased 12-year risk of mortality in those with the highest GlycA levels.[7]

GlycA, also called orosomucoid,[37] is a positive acute phase protein, and its concentration increases in response to systemic tissue injury, inflammation or infection.[38] GlycA is mainly produced by the liver, but it is also synthesized in myelocytes and released by activated neutrophils.[39] Being a type I acute phase protein, GlycA is induced by cytokines, interleukins and tumor necrosis factor alpha (TNFα),[40,41] which among others stimulate a systemic inflammatory response in COPD patients who lose weight.[42] GlycA is one of the main drug binding proteins, carrying basic and neutral lipophilic drugs such as steroid hormones or medications in blood.[43]

A strength of our study is that it is the largest and most comprehensive metabolic study of COPD and lung function. Another strength is the use of the NMR platform, which is valued for being non-invasive, non-destructive, fast and for providing highly reproducible results.[44] Our MR approach allowed us to gain more insight into the direction of the effects, yielding a new interpretation of our data suggesting that GlycA is an independent risk factor of COPD. Yet we have to acknowledge that a limitation of MR is that our knowledge of the genetic determinants of both COPD and GlycA is very limited. In addition, we acknowledge possible limitations of MR due to pleiotropy, the lack of trans-ethnic studies and remaining bias due to canalization.

Altogether, combining the epidemiological data with our MR analyses suggests that GlycA is a predictor of COPD and may be a mediator in the causal pathway linking smoking to COPD. Further functional studies investigating the role of GlycA in COPD will provide more insight into the pathogenesis, prognosis and treatment response of COPD and lung function decline. Our study highlights the power of cross-omics and epidemiological data integration.

REFERENCES

1. World Health Organizaon - WHO. *World Health Statistics 2008*. World Health Organization; 2008.
2. Lozano R, Naghavi M FK. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(380):2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0
3. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP
4. Wain L V., Shrine N, Artigas MS, et al. Supplementary: Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49(3):416-425. doi:10.1038/ng.3787
5. Vestbo J, Anderson W, Coxson HO, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J*. 2008;31(4):869-873. doi:10.1183/09031936.00111707
6. Agusti A, Calverley PMA, Celli B, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res*. 2010;11(1):122.
7. Kettunen J, Ritchie SC, Anufrieva O, et al. Biomarker Glycoprotein Acetyls Is Associated With the Risk of a Wide Spectrum of Incident Diseases and Stratifies Mortality Risk in Angiography Patients. *Circ Genomic Precis Med*. 2018;11(11):e002234. doi:10.1161/CIRCGEN.118.002234
8. Ubhi BK, Riley JH, Shaw PA, et al. Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur Respir J*. 2012;40(2):345-355. doi:10.1183/09031936.00112411
9. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol*. 2017;32(9):807-850. doi:10.1007/s10654-017-0321-4
10. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet*. 2005;69(3):288-295. doi:10.1046/j.1529-8817.2005.00162.x
11. Henneman P, Aulchenko YS, Frants RR, van Dijk KW, Oostra BA, van Duijn CM. Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: The Erasmus Rucphen Family study. *J Med Genet*. 2008;45(9):572-577. doi:10.1136/jmg.2008.058388
12. Tigchelaar EF, Zhernakova A, Dekens JAM, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open*. 2015;5(8):e006772. doi:10.1136/bmjopen-2014-006772
13. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort Profile: The National FINRISK Study. *Int J Epidemiol*. 2018;47(3):696-696i. doi:10.1093/ije/dyx239
14. Borodulin K, Vartiainen E, Peltonen M, et al. Forty-year trends in cardiovascular risk factors in Finland. *Eur J Public Health*. 2015;25(3):539-546. doi:10.1093/eurpub/cku174
15. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler Thromb Vasc Biol*. 2005;25(11):2368-2375. doi:10.1161/01.ATV.0000184769.22061.da
16. Salihovic S, Ganna A, Fall T, et al. The metabolic fi ngerprint of p , p ′-DDE and HCB exposure in humans. *Environ Int*. 2016. doi:10.1016/j.envint.2015.12.015

17. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015;44(4):1172-1180. doi:10.1093/ije/dyu229
18. Stolk RP, Rosmalen JGM, Postma DS, et al. Universal risk factors for multifactorial diseases. *Eur J Epidemiol*. 2008;23(1):67-74. doi:10.1007/s10654-007-9204-4
19. Konttinen H, Llewellyn C, Silventoinen K, et al. Genetic predisposition to obesity, restrained eating and changes in body weight: a population-based prospective study. *Int J Obes (Lond)*. 2018;42(4):858-865. doi:10.1038/ijo.2017.278
20. Engström G, Gerhardsson de Verdier M, Dahlbäck M, Janson C, Lind L. BP variability and cardiovascular autonomic function in relation to forced expiratory volume: a population-based study. *Chest*. 2009;136(1):177-183. doi:10.1378/chest.08-2529
21. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med*. 1995;152(3):1107-1136. doi:10.1164/ajrccm.152.3.7663792
22. Hedenström H, Malmberg P, Fridriksson H V. Reference values for lung function tests in men: regression equations with smoking variables. *Ups J Med Sci*. 1986;91(3):299-310. http://www.ncbi.nlm.nih.gov/pubmed/3811032. Accessed November 15, 2018.
23. Hedenström H, Malmberg P, Agarwal K. Reference values for lung function tests in females. Regression equations with smoking variables. *Bull Eur Physiopathol Respir*. 21(6):551-557. http://www.ncbi.nlm.nih.gov/pubmed/4074961. Accessed November 15, 2018.
24. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circ Cardiovasc Genet*. 2015;8(1):192-206. doi:10.1161/Circgenetics.114.000216
25. Kettunen J, Demirkan AA, Würtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun*. 2016;7:11122. doi:10.1038/ncomms11122
26. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340
27. Li J, Ji L. Adjusting multiple testing in multilocus analysis using the eigenvalues of a correlaton matrix. *Heredity (Edinb)*. 2005;95(3):1-7.
28. Liu J, van Klinken JB, Semiz S, et al. A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. *Diabetes*. 2017;66(11):2915-2926. doi:10.2337/db17-0199
29. Dastani Z, Hivert MF, Timpson NJ, et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet*. 2012;8(3):e1002607. doi:10.1371/journal.pgen.1002607PGENETICS-D-11-02099 [pii]
30. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
31. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr*. 2007;96(5):644-647. doi:10.1111/j.1651-2227.2006.00178.x
32. Fischer K, Kettunen J, Würtz P, et al. Biomarker Profiling by Nuclear Magnetic Resonance Spectroscopy for the Prediction of All-Cause Mortality: An Observational Study of 17,345 Persons. Minelli C, ed. *PLoS Med*. 2014;11(2):e1001606. doi:10.1371/journal.pmed.1001606
33. Lawler PR, Akinkuolie AO, Chandler PD, et al. Circulating N-Linked Glycoprotein Acetyls and Longitudinal Mortality Risk. *Circ Res*. 2016;118(7):1106-1115. doi:10.1161/CIRCRESAHA.115.308078

34. Ritchie SC, Würtz P, Nath AP, et al. The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection. *Cell Syst*. 2015;1(4):293-301. doi: 10.1016/j.cels.2015.09.007
35. Otvos JD, Shalaurova I, Wolak-Dinsmore J, et al. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clin Chem*. 2015;61(5):714-723. doi:10.1373/clinchem.2014.232918
36. Gruppen EG, Riphagen IJ, Connelly MA, Otvos JD, Bakker SJL, Dullaart RPF. GlycA, a Pro-Inflammatory Glycoprotein Biomarker, and Incident Cardiovascular Disease: Relationship with C-Reactive Protein and Renal Function. Shimosawa T, ed. *PLoS One*. 2015;10(9):e0139057. doi:10.1371/journal.pone.0139057
37. Fournier T, Medjoubi-N N, Porquet D. Alpha-1-acid glycoprotein. *Biochim Biophys Acta - Protein Struct Mol Enzymol*. 2000;1482(1):157-171. doi:https://doi.org/10.1016/S0167-4838(00)00153-9
38. Gabay C, Kushner I. Acute-Phase Proteins and Other Systemic Responses to Inflammation. *N Engl J Med*. 1999;340(6):448-454. doi:10.1056/NEJM199902113400607
39. Theilgaard-Mönch K, Jacobsen LC, Rasmussen T, et al. Highly glycosylated alpha1-acid glycoprotein is synthesized in myelocytes, stored in secondary granules, and released by activated neutrophils. *J Leukoc Biol*. 2005;78(2):462-470. doi:10.1189/jlb.0105042
40. Moshage H. REVIEW ARTICLECytokines and the Hepatic Acute Phase Response. *J Pathol*. 1997;266(August 1996):257-266.
41. Bistrian BR. Acute phase proteins and the systemic inflammatory response. *Crit Care Med*. 1999;27(3):452-453. doi:10.1097/00003246-199903000-00004
42. de Godoy I, Donahoe M, Calhoun WJ, Mancino J, Rogers RM. Elevated TNF-alpha production by peripheral blood monocytes of weight-losing COPD patients. *Am J Respir Crit Care Med*. 1996;153(2):633-637. doi:10.1164/ajrccm.153.2.8564110
43. Israili ZH, Dayton PG. Human alpha 1-acid glycoprotein and its interactions with drugs. *Drug Metab Rev*. 2001;33(2):161-235.
44. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem*. 2007;387(2):525-527. doi:10.1007/s00216-006-0687-8