Supplemental Information

Supplemental Figures

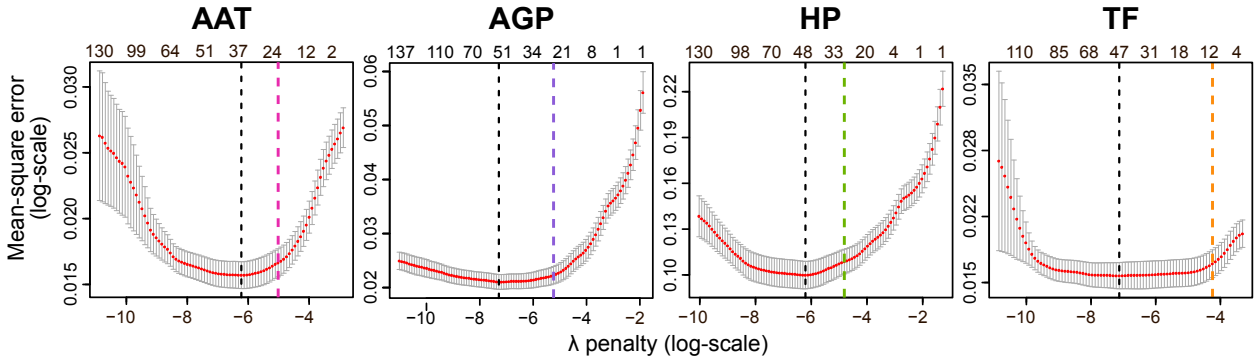

Figure S1: Glycoprotein imputation model training and selection in the 615 DILGOM07 participants with complete NMR metabolite measurements and matched glycoprotein data (N=611 for HP). Each plot shows the lasso model tuning in the 10-fold cross validation model training procedure (**Methods**). Grey bars show the range and red points show the average of the mean-square error (MSE) across the 10 test folds for each of the 100 lasso regression λ penalties (x-axes). Numbers on the top axes correspond to the number of features selected for inclusion given the corresponding λ penalty. Age, sex, BMI, and 149 metabolic measures by NMR (**Table S1**) were considered as candidate features for each imputation model. For each glycoprotein, the black dashed line indicates the imputation model with smallest average MSE across the 10 test folds during model training. The coloured dashed line indicates the selected model (detailed in **Table S2**); the simplest model within 1 standard error of the model with the smallest average MSE. Note the MSE cannot be compared between the different glycoproteins since their range of concentrations differ.

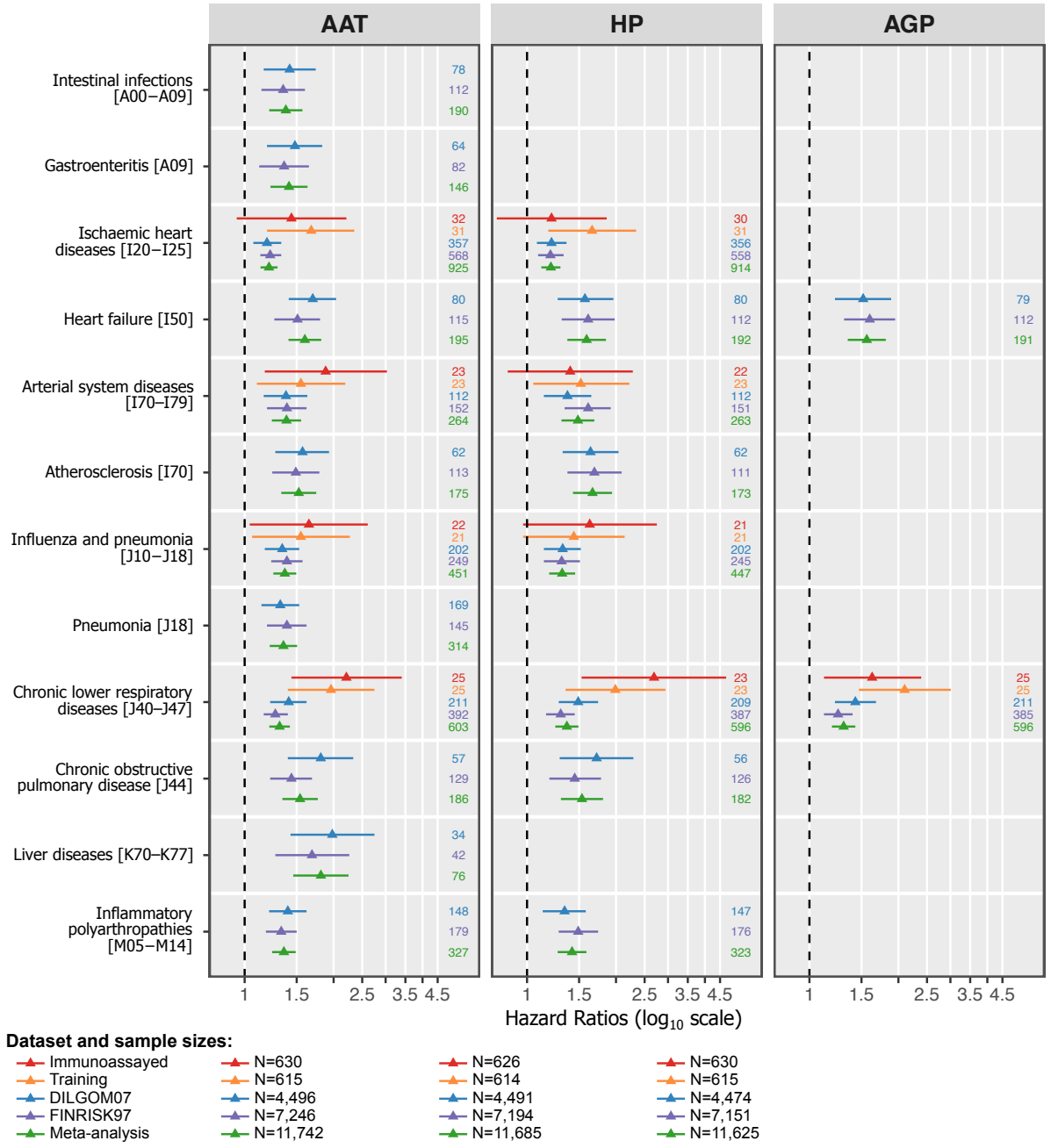

Figure S2: Forest plots for significant and replicable biomarker associations from **Figure 2** comparing hazard ratios from meta-analysis (green) to hazard ratios calculated in each of DILGOM07 (blue) and FINRISK97 (purple) (**Methods**). For comparison, hazard ratios calculated from the immunoassayed glycoprotein concentrations and calculated from the predicted glycoprotein concentrations in the 615 DILGOM07 participants used to train the imputation models (labelled "Training") are also shown in red and yellow respectively, provided there were >20 incident events in the diagnosis category (**Methods**). The number of incident events for each outcome in each cohort are shown to the right of each hazard ratio. The alphanumeric codes in the square brackets indicate the ICD10 codes or disease categories for each diagnosis. Exact hazard ratios are detailed in **Table S5**.

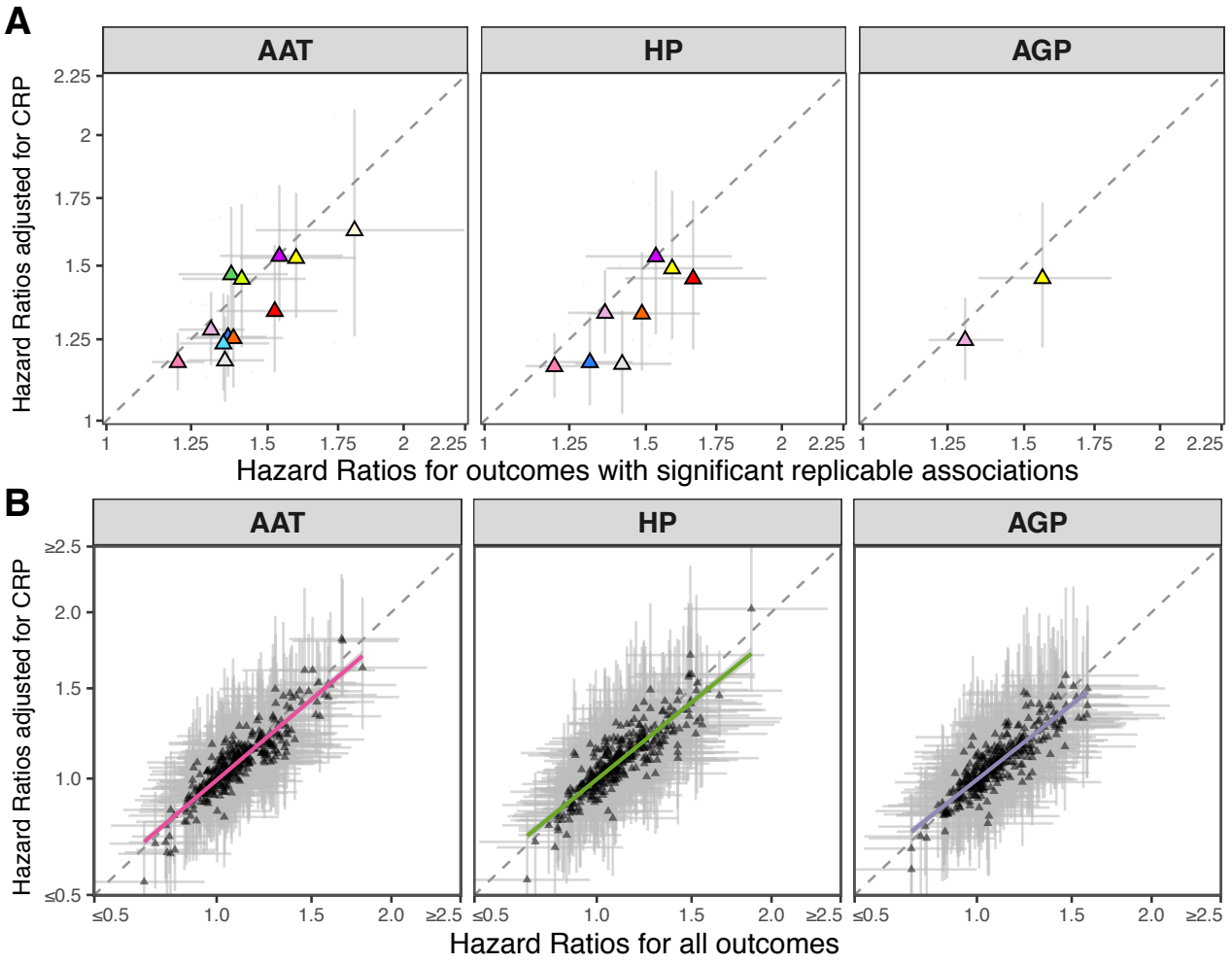

Figure S3: Sensitivity analysis of glycoprotein biomarker associations to CRP adjustment for **A)** outcomes with significant replicable and replicable associations for each biomarker, and **B)** across all outcomes with ≥ 20 events in both DILGOM07 and FINRISK97. In both **A)** and **B)** each plot compares the hazard ratios conferred per standard deviation increase of the glycoprotein (x-axes) to hazard ratios conferred per standard deviation increase of the glycoprotein adjusted for CRP (y-axes) in meta-analysis of DILGOM07 and FINRISK97 (**Methods**). Data are shown on a square root scale. The grey dashed diagonal line indicates the location where a hazard ratio would fall if it was unchanged after CRP adjustment. Light grey crosses centred on each hazard ratio represent the 95% confidence intervals for the hazard ratio (horizontal bars) and for the hazard ratio adjusted for CRP (vertical bars). In **A)** hazard ratios are coloured according to the legend. The coloured line in each plot in **B)** shows the line of best fit (linear regression) of the CRP-adjusted hazard ratios on the hazard ratios without CRP adjustment, indicating the overall attenuation by CRP across all outcomes. In **B)** 95% confidence intervals with an upper limit ≥ 2.5 or a lower limit ≤ 0.5 are truncated on the plot.

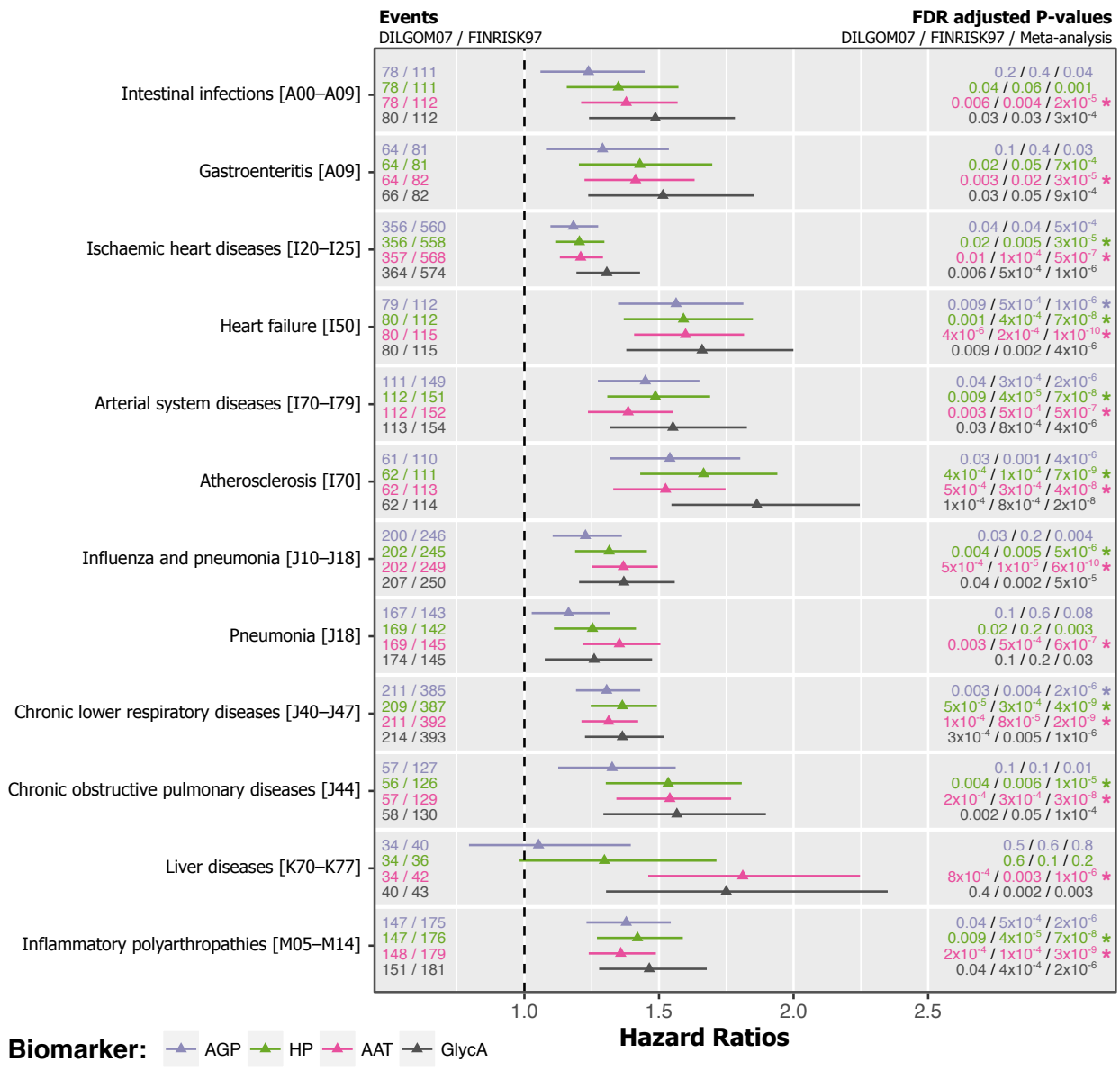

Figure S4: Comparison of glycoproteins in meta-analysis of DILGOM07 and FINRISK97 for all outcomes for which at least one of AAT, HP, or AGP was a significant and replicable biomarker in **Figure 2**. Triangles indicate Cox proportional hazard ratio estimates for each outcome conferred per standard deviation increase of each biomarker respectively in the meta-analysis (**Methods**). Bars around each hazard ratio indicate the 95% confidence interval. The number of events in DILGOM07 and FINRISK97 are shown to the left of each hazard ratio for each outcome. The Storey-Tibshirani FDR adjusted P-values in DILGOM07, FINRISK97, and meta-analysis are shown to the right of each hazard ratio. A '*' to the right of the P-values indicates the glycoprotein was a significant and replicable biomarker for that outcome in **Figure 2** (FDR < 0.05/3 in DILGOM07, FINRISK97, and meta-analysis). Alphanumeric codes in the square brackets indicate the ICD-10 codes for each outcome. Hazard ratios are detailed in **Table S7**.

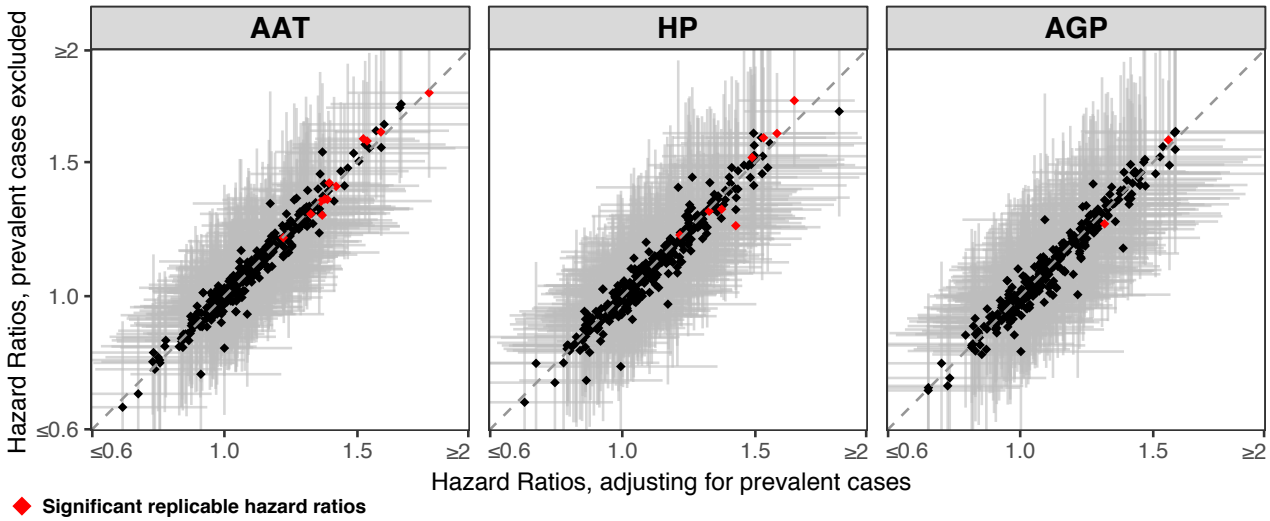

Figure S5: Sensitivity analysis of glycoprotein biomarker associations to prevalent disease. Each plot compares the hazard ratios (diamonds) calculated when adjusting for prevalent case status as a covariate (x-axes) to hazard ratios calculated excluding individuals with any prevalent cases for each outcome (y-axes) in meta-analysis of DILGOM07 and FINRISK97. Light grey crosses centred on each hazard ratio represent the 95% confidence intervals for the hazard ratio calculated when adjusting for prevalent cases as a covariate (horizontal bars) and for the hazard ratio calculated excluding prevalent cases of each outcome (vertical bars). The grey dashed diagonal line indicates the location where hazard ratios should fall if their estimates are identical in the different models. Red diamonds indicate outcomes significantly associated with each biomarker in **Figure 2**.

Supplemental Tables

Table S1: Serum NMR measurements that were included as inputs to the lasso regression models used to identify the glycoprotein imputation models. Derived ratios were excluded from the analyses (**Methods**).

Name	Molecular species	Units
Lipoprotein particles and their constituents		
Chylomicrons and very low density lipoprotein (VLDL) particles		
VLDL-D	Mean diameter for VLDL particles	nm
VLDL-C	Total cholesterol in VLDL	mmol/l
VLDL-TG	Triglycerides in VLDL	mmol/l
Low density lipoprotein (LDL) particles		
LDL-D	Mean diameter for LDL particles	nm
LDL-C	Total cholesterol in LDL	mmol/l
LDL-TG	Triglycerides in LDL	mmol/l
High density lipoprotein (HDL) particles		
HDL-D	Mean diameter for HDL particles	nm
HDL-C	Total cholesterol in HDL	mmol/l
HDL-TG	Triglycerides in HDL	mmol/l
Lipoprotein subclasses and their constituents		
Chylomicrons and extremely large VLDL particles (particle diameters > 75 nm)		
XXL-VLDL-P	Total concentration of chylomicrons and extremely large VLDL particles	mol/l
XXL-VLDL-L	Total lipids in chylomicrons and extremely large VLDL	mmol/l
XXL-VLDL-PL	Phospholipids in chylomicrons and extremely large VLDL	mmol/l
XXL-VLDL-C	Total cholesterol in chylomicrons and extremely large VLDL	mmol/l
XXL-VLDL-CE	Cholesterol esters in chylomicrons and extremely large VLDL	mmol/l
XXL-VLDL-FC	Free cholesterol in chylomicrons and extremely large VLDL	mmol/l
XXL-VLDL-TG	Triglycerides in chylomicrons and extremely large VLDL	mmol/l
Very large VLDL particles (average particle diameter of 64 nm)		
XL-VLDL-P	Total concentration of very large VLDL particles	mol/l
XL-VLDL-L	Total lipids in very large VLDL	mmol/l
XL-VLDL-PL	Phospholipids in very large VLDL	mmol/l
XL-VLDL-C	Total cholesterol in very large VLDL	mmol/l
XL-VLDL-CE	Cholesterol esters in very large VLDL	mmol/l
XL-VLDL-FC	Free cholesterol in very large VLDL	mmol/l
XL-VLDL-TG	Triglycerides in very large VLDL	mmol/l
Large VLDL particles (average particle diameter of 53.6 nm)		
L-VLDL-P	Total concentration of large VLDL particles	mol/l
L-VLDL-L	Total lipids in large VLDL	mmol/l
L-VLDL-PL	Phospholipids in large VLDL	mmol/l
L-VLDL-C	Total cholesterol in large VLDL	mmol/l
L-VLDL-CE	Cholesterol esters in large VLDL	mmol/l
L-VLDL-FC	Free cholesterol in large VLDL	mmol/l
L-VLDL-TG	Triglycerides in large VLDL	mmol/l
Medium VLDL particles (average particle diameter of 44.5 nm)		
M-VLDL-P	Total concentration of medium VLDL particles	mol/l
M-VLDL-L	Total lipids in medium VLDL	mmol/l
M-VLDL-PL	Phospholipids in medium VLDL	mmol/l
M-VLDL-C	Total cholesterol in medium VLDL	mmol/l
M-VLDL-CE	Cholesterol esters in medium VLDL	mmol/l
M-VLDL-FC	Free cholesterol in medium VLDL	mmol/l
M-VLDL-TG	Triglycerides in medium VLDL	mmol/l
Small VLDL particles (average particle diameter of 36.8 nm)		
S-VLDL-P	Total concentration of small VLDL particles	mol/l
S-VLDL-L	Total lipids in small VLDL	mmol/l
S-VLDL-PL	Phospholipids in small VLDL	mmol/l
S-VLDL-C	Total cholesterol in small VLDL	mmol/l
S-VLDL-CE	Cholesterol esters in small VLDL	mmol/l
S-VLDL-FC	Free cholesterol in small VLDL	mmol/l

Name	Molecular species	Units
Lipoprotein subclasses and their constituents (continued)		
Small VLDL particles (average particle diameter of 36.8 nm) (continued)		
S-VLDL-TG	Triglycerides in small VLDL	mmol/l
Very small VLDL particles (average particle diameter of 31.3 nm)		
XS-VLDL-P	Total concentration of very small VLDL particles	mol/l
XS-VLDL-L	Total lipids in very small VLDL	mmol/l
XS-VLDL-PL	Phospholipids in very small VLDL	mmol/l
XS-VLDL-C	Total cholesterol in very small VLDL	mmol/l
XS-VLDL-CE	Cholesterol esters in very small VLDL	mmol/l
XS-VLDL-FC	Free cholesterol in very small VLDL	mmol/l
XS-VLDL-TG	Triglycerides in very small VLDL	mmol/l
Intermediate density lipoprotein particles (average particle diameter of 28.6 nm)		
IDL-P	Total concentration of IDL particles	mol/l
IDL-L	Total lipids in IDL	mmol/l
IDL-PL	Phospholipids in IDL	mmol/l
IDL-C	Total cholesterol in IDL	mmol/l
IDL-CE	Cholesterol esters in IDL	mmol/l
IDL-FC	Free cholesterol in IDL	mmol/l
IDL-TG	Triglycerides in IDL	mmol/l
Large LDL particles (average particle diameter of 25.5 nm)		
L-LDL-P	Total concentration of large LDL particles	mol/l
L-LDL-L	Total lipids in large LDL	mmol/l
L-LDL-PL	Phospholipids in large LDL	mmol/l
L-LDL-C	Total cholesterol in large LDL	mmol/l
L-LDL-CE	Cholesterol esters in large LDL	mmol/l
L-LDL-FC	Free cholesterol in large LDL	mmol/l
L-LDL-TG	Triglycerides in large LDL	mmol/l
Medium LDL particles (average particle diameter of 23.0 nm)		
M-LDL-P	Total concentration of medium LDL particles	mol/l
M-LDL-L	Total lipids in medium LDL	mmol/l
M-LDL-PL	Phospholipids in medium LDL	mmol/l
M-LDL-C	Total cholesterol in medium LDL	mmol/l
M-LDL-CE	Cholesterol esters in medium LDL	mmol/l
M-LDL-FC	Free cholesterol in medium LDL	mmol/l
M-LDL-TG	Triglycerides in medium LDL	mmol/l
Small LDL particles (average particle diameter of 18.7 nm)		
S-LDL-P	Total concentration of small LDL particles	mol/l
S-LDL-L	Total lipids in small LDL	mmol/l
S-LDL-PL	Phospholipids in small LDL	mmol/l
S-LDL-C	Total cholesterol in small LDL	mmol/l
S-LDL-CE	Cholesterol esters in small LDL	mmol/l
S-LDL-FC	Free cholesterol in small LDL	mmol/l
S-LDL-TG	Triglycerides in small LDL	mmol/l
Very large HDL particles (average particle diameter of 14.3 nm)		
XL-HDL-P	Total concentration of very large HDL particles	mol/l
XL-HDL-L	Total lipids in very large HDL	mmol/l
XL-HDL-PL	Phospholipids in very large HDL	mmol/l
XL-HDL-C	Total cholesterol in very large HDL	mmol/l
XL-HDL-CE	Cholesterol esters in very large HDL	mmol/l
XL-HDL-FC	Free cholesterol in very large HDL	mmol/l
XL-HDL-TG	Triglycerides in very large HDL	mmol/l
Large HDL particles (average particle diameter of 12.1 nm)		
L-HDL-P	Total concentration of large HDL particles	mol/l
L-HDL-L	Total lipids in large HDL	mmol/l
L-HDL-PL	Phospholipids in large HDL	mmol/l
L-HDL-C	Total cholesterol in large HDL	mmol/l
L-HDL-CE	Cholesterol esters in large HDL	mmol/l
L-HDL-FC	Free cholesterol in large HDL	mmol/l
L-HDL-TG	Triglycerides in large HDL	mmol/l

Name	Molecular species	Units
Lipoprotein subclasses and their constituents (continued)		
Medium HDL particles (average particle diameter of 10.9 nm)		
M-HDL-P	Concentration of medium HDL particles	mol/l
M-HDL-L	Total lipids in medium HDL	mmol/l
M-HDL-PL	Phospholipids in medium HDL	mmol/l
M-HDL-C	Total cholesterol in medium HDL	mmol/l
M-HDL-CE	Cholesterol esters in medium HDL	mmol/l
M-HDL-FC	Free cholesterol in medium HDL	mmol/l
M-HDL-TG	Triglycerides in medium HDL	mmol/l
Small HDL particles (average particle diameter of 8.7 nm)		
S-HDL-P	Concentration of small HDL particles	mol/l
S-HDL-L	Total lipids in small HDL	mmol/l
S-HDL-PL	Phospholipids in small HDL	mmol/l
S-HDL-C	Total cholesterol in small HDL	mmol/l
S-HDL-CE	Cholesterol esters in small HDL	mmol/l
S-HDL-FC	Free cholesterol in small HDL	mmol/l
S-HDL-TG	Triglycerides in small HDL	mmol/l
HDL_2 fraction (1.063–1.125 g/mL)		
HDL2-C	Cholesterol in the HDL_2 fraction	mmol/l
HDL_3 fraction (1.125–1.210 g/mL)		
HDL3-C	Cholesterol in the HDL_3 fraction	mmol/l
Cholesterol		
Serum-C	Total cholesterol in serum	mmol/l
EstC	Esterified cholesterol	mmol/l
FreeC	Free cholesterol	mmol/l
Remnant-C	Remnant cholesterol (non-HDL, non-LDL -cholesterol)	mmol/l
Glycerides and phospholipids		
Serum-TG	Total triglycerides in serum	mmol/l
TotPG	Total phosphoglycerides	mmol/l
PC	Phosphatidylcholine and other cholines	mmol/l
SM	Sphingomyelins	mmol/l
TotCho	Total cholines	mmol/l
Apolipoproteins		
ApoA1	Apolipoprotein A-I	g/l
ApoB	Apolipoprotein B	g/l
Fatty acids and saturation		
TotFA	Total fatty acids	mmol/l
UnSat	Estimated degree of fatty acid unsaturation	%
FAw3	Omega-3 fatty acids	mmol/l
FAw6	Omega-6 fatty acids	mmol/l
PUFA	Polyunsaturated fatty acids	mmol/l
MUFA	Monounsaturated fatty acids; 16:1, 18:1	mmol/l
SFA	Saturated fatty acids	mmol/l
DHA	Docosahexaenoic acid; 22:6	mmol/l
LA	Linoleic acid; 18:2	mmol/l
Amino acids		
Ala	Alanine	mmol/l
Gln	Glutamine	mmol/l
Gly	Glycine	mmol/l
His	Histidine	mmol/l
Ile	Isoleucine	mmol/l
Leu	Leucine	mmol/l
Val	Valine	mmol/l
Phe	Phenylalanine	mmol/l
Tyr	Tyrosine	mmol/l
Glycolysis related metabolites		
Glc	Glucose	mmol/l
Lac	Lactate	mmol/l
Pyr	Pyruvate	mmol/l
Cit	Citrate	mmol/l

Name	Molecular species	Units
Glycolysis related metabolites (continued)		
Glol	Glycerol	mmol/l
Ketone bodies		
Ace	Acetate	mmol/l
AcAce	Acetoacetate	mmol/l
bOHBut	3-hydroxybutyrate	mmol/l
Fluid balance		
Crea	Creatinine	mmol/l
Alb	Albumin	signal area
Inflammation		
GlycA	Glycoprotein *N*-linked glycan methyl groups	mmol/l

Table S2: Summary of imputation models for AAT, AGP, HP, and TF. A) Here, coefficients are rounded to two significant figures. Full precision models are made available through the *imputegp* R package (**Methods**). Each model predicts the concentration of each glycoprotein on a natural logarithm scale. Concentrations in mg/L are obtained by exponentiating the result. Variables in the imputation models are also required on a natural logarithm scale, with the exceptions of participant Age and Sex. The coding for the Sex variable was 1 for men and 2 for women. The NMR-metabolite measures are described in **Table S1**. **B)** Model coefficients are given after log transformation of each variable and standardisation to their mean and standard deviation in DILGOM07, *i.e.* coefficients indicate relative contribution to determining the concentration of the glycoprotein. Variables are listed from left to right in descending order of their relative contribution in both **A)** and **B).**

A)

AAT =	− 11	+ 0.68 *GlycA* − 0.094 *FAw3* − 0.8 *VLDL-D* + 0.37 *HDL3-C* + 4.2 *LDL-D* + 0.14 *Phe* − 0.08 *Leu* − 0.064 *ApoB* − 0.19 *Alb* − 0.045 *Tyr* + 0.019 *bOHBut* − 0.045 *BMI* − 0.043 *Ala* + 0.011 *L-HDL-TG* − 0.022 *Ile* − 0.025 *Ace* − 0.038 *His* + 0.0024 *HDL-TG*
AGP =	5.7	+ 1.7 *GlycA* − 0.34 *TotFA* + 0.22 *IDL-FC* + 0.043 *L-HDL-FC* − 0.22 *His* − 0.092 *HDL-TG* + 0.11 *BMI* − 0.16 S-*HDL-FC* − 0.057 *S-LDL-TG* − 0.023 *bOHBut* − 0.062 *LA* + 0.066 *S-HDL-CE* − 0.055 *Lac* − 0.021 *S-VLDL-TG* − 0.042 *Ace* − 0.052 *Cit* − 0.038 *SFA* − 0.049 *Ala* + 0.0013 *XXL-VLDL-CE* + 0.014 *Glol* + 0.00017 *Age* + 0.0084 *Crea* + 0.0061 *Gly*
HP =	0.53	+ 3.0 *GlycA* − 0.82 *LA* + 0.36 *IDL-FC* + 0.48 *SM* − 0.26 *FAw3* − 0.29 *HDL-TG* − 0.068 *S-VLDL-CE* + 0.003 Age − 0.75 Alb − 0.13 Ile − 0.24 Cit − 0.89 *VLDL-D* − 0.13 *Leu* + 0.13 *Val* + 0.0066 *L-VLDL-CE* + 0.084 *Pyr* − 0.084 *Lac* − 0.094 *Gln* + 0.042 *M-HDL-FC* − 0.011 *XL-HDL-TG* + 0.0089 *XL-HDL-PL* − 0.048 *His* − 0.026 *Tyr* − 0.024 *BMI* − 0.0037 *L-HDL-TG* − 0.004 *PUFA* + 0.00056 *S-LDL-FC*
TF =	1.1	+ 0.14 *GlycA* + 0.032 *Sex* − 0.0010 *Age* + 0.090 *S-HDL-FC* − 0.037 *Ace* + 0.039 *Ala* + 0.024 *SFA* + 0.013 *His* − 0.0097 *Gln*

B)

AAT =	0	+ 0.54 *GlycA* − 0.17 *FAw3* − 0.16 *VLDL-D* + 0.14 *HDL3-C* + 0.11 *LDL-D* + 0.11 *Phe* − 0.096 *Leu* − 0.081 *ApoB* − 0.058 *Alb* − 0.057 *Tyr* + 0.052 *bOHBut* − 0.045 *BMI* − 0.036 *Ala* + 0.036 *L-HDL-TG* − 0.035 *Ile* − 0.033 *Ace* − 0.027 *His* + 0.0034 *HDL-TG*
AGP =	0	+ 0.93 *GlycA* − 0.27 *TotFA* + 0.26 *IDL-FC* + 0.12 *L-HDL-FC* − 0.11 *His* − 0.091 *HDL-TG* + 0.08 *BMI* − 0.079 *S-HDL-FC* − 0.073 *S-LDL-TG* − 0.05 *bOHBut* − 0.048 *LA* + 0.045 *S-HDL-CE* − 0.043 *Lac* − 0.043 *S-VLDL-TG* − 0.04 *Ace* − 0.037 *Cit* − 0.035 *SFA* − 0.032 *Ala* + 0.026 *XXL-VLDL-CE* + 0.022 *Glol* + 0.013 *Age* + 0.013 *Crea* + 0.0041 *Gly*
HP =	0	+ 0.84 *GlycA* − 0.29 *LA* + 0.21 *IDL-FC* + 0.18 *SM* − 0.16 *FAw3* − 0.14 *HDL-TG* − 0.088 *S-VLDL-CE* + 0.087 *Age* − 0.078 *Alb* − 0.076 *Ile* − 0.074 *Cit* − 0.062 *VLDL-D* − 0.055 *Leu* + 0.051 *Val* + 0.048 *L-VLDL-CE* + 0.04 *Pyr* − 0.032 *Lac* − 0.023 *Gln* + 0.022 *M-HDL-FC* − 0.014 *XL-HDL-TG* + 0.012 *XL-HDL-PL* − 0.012 *His* − 0.011 *Tyr* − 0.0086 *BMI* − 0.004 *L-HDL-TG* − 0.0014 *PUFA* + 0.00025 *S-LDL-FC*
TF =	0	+ 0.13 *GlycA* + 0.11 *Sex* − 0.097 *Age* + 0.08 *S-HDL-FC* − 0.057 *Ace* + 0.037 *Ala* + 0.031 *SFA* + 0.011 *His* − 0.0079 *Gln*

Table S3: **Glycoprotein imputation model accuracy** assessed in the 626 DILGOM07 participants with matched glycoprotein assay, serum NMR-metabolite measures, age, sex, and BMI. The first row shows the Spearman's rank correlation coefficient (ρ) calculated between the predicted and observed glycoprotein measurements in **Figure 1A**. The second row shows its mean ± standard deviation in the 10-fold cross-validation procedure used to train each imputation model (**Figure 1B**).

	AAT	**AGP**	**HP**	**TF**
Predicted vs. observed	ρ = 0.66	ρ = 0.78	ρ = 0.74	ρ = 0.45
Cross-validation	ρ = 0.63 ± 0.077	ρ = 0.74 ± 0.078	ρ = 0.71 ± 0.043	ρ = 0.42 ± 0.10

Table S4: Listing of outcomes analysed for an association with each glycoprotein. Provided as a separate excel file. The number next to each glycoprotein indicates the number of outcomes with sufficient events to analyse in the subset of individuals in which each glycoprotein could be imputed. Numbers below the DILGOM07 and FINRISK97 subheadings indicated the number of samples in which each glycoprotein could be imputed.

Table S5: Significant and replicable glycoprotein associated risks of disease and mortality. Details of the cox proportional hazard ratios shown in **Figure S2**. Hazard ratios are listed only where associations were significant and replicable (Storey-Tibshirani FDR adjusted $P < 0.05/3$ in DILGOM07, FINRISK97, and meta-analysis). Diagnosis data were analysed for a matched 8-year follow-up period of DILGOM07 and FINRISK97. Models were fit using age as the time scale and adjusting for sex, smoking status, BMI, blood pressure, alcohol consumption, prevalent cases, and previously identified biomarkers for 5-year risk of all-cause mortality (citrate, albumin, and VLDL particle size). The alphanumeric codes in the square brackets indicate the ICD10 codes for each outcome. Dataset: "Meta": inverse-variance weighted fixed effects meta-analysis of DILGOM07 and FINRISK97. "DILGOM07": hazard ratios calculated using the full DILGOM07 cohort. "FINRISK97": hazard ratios calculated using the full FINRISK97 cohort. "Training": hazard ratios calculated from the imputed glycoprotein concentrations in the 615 DILGOM07 participants used to train the imputation models. "Assayed": hazard ratios calculated from the immunoassayed concentration in the 630 DILGOM07 participants with glycoprotein immunoassay data. #S: sample size of the dataset. #E: number of events that occurred in the 8-year follow-up in the corresponding dataset. #P: number of prevalent cases prior to baseline; adjustment was performed when there were ≥ 10 prevalent cases (**Methods**). HR: hazard ratio for the first diagnosis occurrence (fatal or non-fatal) conferred per standard deviation increase of AAT, HP, AGP, or GlycA. 95% CI: 95% confidence interval for the hazard ratio. P: hazard ratio P-value. Q: Storey-Tibshirani FDR adjusted P-value.

	Outcome	Dataset	#S	#E	#P	HR	95% CI	P	Q
AAT	Intestinal infections [A00–A09]	Meta	11,742	190	215	1.38	1.21–1.57	1×10^{-6}	2×10^{-5}
		FINRISK97	7,246	112	87	1.35	1.14–1.60	4×10^{-4}	0.004
		DILGOM07	4,496	78	128	1.42	1.16–1.74	7×10^{-4}	0.006
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
AAT	Gastroenteritis [A09]	Meta	11,742	146	137	1.41	1.22–1.63	3×10^{-6}	3×10^{-5}
		FINRISK97	7,246	82	51	1.36	1.12–1.65	0.002	0.02
		DILGOM07	4,496	64	86	1.48	1.19–1.83	3×10^{-4}	0.003
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
AAT	Ischaemic heart diseases [I20–I25]	Meta	11,742	925	502	1.21	1.13–1.29	2×10^{-8}	5×10^{-7}
		FINRISK97	7,246	568	268	1.22	1.13–1.33	2×10^{-6}	1×10^{-4}
		DILGOM07	4,496	357	234	1.19	1.07–1.33	0.002	0.01
		Training	615	31	26	1.68	1.19–2.35	0.003	0.03
		Assayed	630	32	27	1.44	0.94–2.21	0.09	0.3
AAT	Heart failure [I50]	Meta	11,742	195	35	1.60	1.41–1.82	5×10^{-13}	1×10^{-10}
		FINRISK97	7,246	115	12	1.51	1.26–1.80	5×10^{-6}	2×10^{-4}
		DILGOM07	4,496	80	23	1.70	1.41–2.04	2×10^{-8}	4×10^{-6}
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
AAT	Arterial system diseases [I70–190]	Meta	11,742	264	104	1.39	1.24–1.55	2×10^{-8}	5×10^{-7}
		FINRISK97	7,246	152	54	1.39	1.19–1.62	4×10^{-5}	5×10^{-4}
		DILGOM07	4,496	112	50	1.38	1.16–1.63	2×10^{-4}	0.003
		Training	615	23	≤ 10	1.55	1.10–2.19	0.01	0.07
		Assayed	630	23	≤ 10	1.88	1.17–3.03	0.009	0.2
AAT	Atherosclerosis [I70]	Meta	11,742	175	64	1.52	1.33–1.75	1×10^{-9}	4×10^{-8}
		FINRISK97	7,246	113	35	1.49	1.24–1.79	2×10^{-5}	3×10^{-4}
		DILGOM07	4,496	62	29	1.57	1.27–1.93	2×10^{-5}	5×10^{-4}
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-

	Outcome	Dataset	#S	#E	#P	HR	95% CI	P	Q
AAT	Influenza and pneumonia [J10–J18]	Meta	11,742	451	244	1.37	1.25–1.50	6×10^{-12}	6×10^{-10}
		FINRISK97	7,246	249	91	1.39	1.23–1.57	6×10^{-8}	1×10^{-5}
		DILGOM07	4,496	202	153	1.34	1.17–1.53	2×10^{-5}	5×10^{-4}
		Training	615	21	17	1.55	1.06–2.27	0.03	0.09
		Assayed	630	22	17	1.65	1.04–2.61	0.03	0.2
AAT	Pneumonia [J18]	Meta	11,742	314	156	1.35	1.22–1.50	3×10^{-8}	6×10^{-7}
		FINRISK97	7,246	145	66	1.39	1.19–1.62	3×10^{-5}	5×10^{-4}
		DILGOM07	4,496	169	90	1.32	1.14–1.53	2×10^{-4}	0.003
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
AAT	Chronic lower respiratory diseases [J40–J47]	Meta	11,742	603	365	1.31	1.21–1.42	3×10^{-11}	2×10^{-9}
		FINRISK97	7,246	392	171	1.27	1.16–1.40	9×10^{-7}	8×10^{-5}
		DILGOM07	4,496	211	194	1.41	1.22–1.62	2×10^{-6}	1×10^{-4}
		Training	615	25	19	1.96	1.40–2.75	1×10^{-4}	0.004
		Assayed	630	25	19	2.21	1.44–3.4	3×10^{-4}	0.01
AAT	Chronic obstructive pulmonary disease [J44]	Meta	11,742	186	51	1.54	1.34–1.77	9×10^{-10}	3×10^{-8}
		FINRISK97	7,246	129	28	1.44	1.22–1.69	2×10^{-5}	3×10^{-4}
		DILGOM07	4,496	57	23	1.81	1.40–2.33	5×10^{-6}	2×10^{-4}
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
AAT	Liver diseases [K70–K77]	Meta	11,742	76	48	1.81	1.46–2.25	7×10^{-8}	1×10^{-6}
		FINRISK97	7,246	42	15	1.69	1.27–2.26	3×10^{-4}	0.003
		DILGOM07	4,496	34	33	1.98	1.43–2.75	4×10^{-5}	8×10^{-4}
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
AAT	Inflammatory polyarthropathies [M05–M14]	Meta	11,742	327	213	1.36	1.24–1.49	6×10^{-11}	3×10^{-9}
		FINRISK97	7,246	179	81	1.33	1.18–1.50	3×10^{-6}	1×10^{-4}
		DILGOM07	4,496	148	132	1.40	1.21–1.62	4×10^{-6}	2×10^{-4}
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
HP	Ischaemic heart diseases [I20–I25]	Meta	11,685	914	497	1.20	1.12–1.30	1×10^{-6}	3×10^{-5}
		FINRISK97	7,194	558	263	1.20	1.09–1.33	2×10^{-4}	0.005
		DILGOM07	4,491	356	234	1.21	1.08–1.36	0.001	0.02
		Training	614	31	26	1.66	1.18–2.34	0.003	0.03
		Assayed	626	30	25	1.21	0.79–1.86	0.4	0.2
HP	Heart failure [I50]	Meta	11,685	192	35	1.59	1.37–1.85	1×10^{-9}	7×10^{-8}
		FINRISK97	7,194	112	12	1.61	1.31–1.98	7×10^{-6}	4×10^{-4}
		DILGOM07	4,491	80	23	1.57	1.27–1.96	5×10^{-5}	0.001
		Training	614	≤ 20	-	-	-	-	-
		Assayed	626	≤ 20	-	-	-	-	-
HP	Arterial system diseases [I70–190]	Meta	11,685	263	104	1.49	1.31–1.69	1×10^{-9}	7×10^{-8}
		FINRISK97	7,194	151	54	1.61	1.34–1.92	3×10^{-7}	4×10^{-5}
		DILGOM07	4,491	112	50	1.37	1.14–1.65	7×10^{-4}	0.009
		Training	614	23	≤ 10	1.52	1.05–2.22	0.03	0.08
		Assayed	626	22	≤ 10	1.40	0.86–2.28	0.2	0.1
HP	Atherosclerosis [I70]	Meta	11,685	173	64	1.67	1.43–1.94	5×10^{-11}	7×10^{-9}
		FINRISK97	7,194	111	35	1.69	1.37–2.09	1×10^{-6}	1×10^{-4}
		DILGOM07	4,491	62	29	1.64	1.32–2.04	1×10^{-5}	4×10^{-4}
		Training	614	≤ 20	-	-	-	-	-
		Assayed	626	≤ 20	-	-	-	-	-
HP	Influenza and pneumonia [J10–J18]	Meta	11,685	447	243	1.31	1.19–1.45	1×10^{-7}	5×10^{-6}
		FINRISK97	7,194	245	91	1.31	1.14–1.51	2×10^{-4}	0.005
		DILGOM07	4,491	202	152	1.32	1.14–1.52	2×10^{-4}	0.004
		Training	614	21	17	1.44	0.97–2.14	0.07	0.2
		Assayed	626	21	16	1.63	0.97–2.75	0.07	0.07

	Outcome	Dataset	#S	#E	#P	HR	95% CI	P	Q
HP	Chronic lower respiratory diseases [J40–J47]	Meta	11,685	596	364	1.36	1.25–1.49	1×10^{-11}	4×10^{-9}
		FINRISK97	7,194	387	170	1.30	1.16–1.45	4×10^{-6}	3×10^{-4}
		DILGOM07	4,491	209	194	1.49	1.28–1.74	2×10^{-7}	5×10^{-5}
		Training	614	23	19	1.99	1.35–2.94	5×10^{-4}	0.02
		Assayed	626	23	19	2.69	1.53–4.72	6×10^{-4}	0.009
HP	Chronic obstructive pulmonary disease [J44]	Meta	11,685	182	50	1.53	1.30–1.81	3×10^{-7}	1×10^{-5}
		FINRISK97	7,194	126	27	1.45	1.19–1.78	3×10^{-4}	0.006
		DILGOM07	4,491	56	23	1.72	1.29–2.29	2×10^{-4}	0.004
		Training	614	≤ 20	-	-	-	-	-
		Assayed	626	≤ 20	-	-	-	-	-
HP	Inflammatory polyarthropathies [M05–M14]	Meta	11,685	323	208	1.42	1.27–1.59	9×10^{-10}	7×10^{-8}
		FINRISK97	7,194	176	76	1.49	1.28–1.74	2×10^{-7}	4×10^{-5}
		DILGOM07	4,491	147	132	1.34	1.13–1.58	6×10^{-4}	0.009
		Training	614	≤ 20	-	-	-	-	-
		Assayed	626	≤ 20	-	-	-	-	-
AGP	Heart failure [I50]	Meta	11,625	191	35	1.56	1.35–1.81	4×10^{-9}	1×10^{-6}
		FINRISK97	7,151	112	12	1.60	1.31–1.95	4×10^{-6}	5×10^{-4}
		DILGOM07	4,474	79	23	1.52	1.22–1.89	2×10^{-4}	0.009
		Training	615	≤ 20	-	-	-	-	-
		Assayed	630	≤ 20	-	-	-	-	-
AGP	Chronic lower respiratory diseases [J40–J47]	Meta	11,625	596	364	1.31	1.19–1.43	1×10^{-8}	2×10^{-6}
		FINRISK97	7,151	385	170	1.25	1.12–1.40	7×10^{-5}	0.004
		DILGOM07	4,474	211	194	1.43	1.22–1.68	1×10^{-5}	0.003
		Training	615	25	19	2.10	1.47–3.01	5×10^{-5}	4×10^{-4}
		Assayed	630	25	19	1.63	1.12–2.39	0.01	0.07

Table S6: Spearman correlation between GlycA and each glycoprotein. Immunoassayed: Spearman correlation between GlycA and each immunoassayed glycoprotein in the 626 DILGOM07 participants with matched immunoassay and NMR-metabolite measures. Training: Spearman correlation between GlycA and the imputed glycoprotein levels in the 615 DILGOM07 participants used for model training. DILGOM07: Spearman correlation between GlycA and the imputed glycoprotein levels in all 4,540 DILGOM07 participants. FINRISK97: Spearman correlation between GlycA and the imputed glycoprotein levels in all 7,321 FINRISK97 participants.

Glycoprotein	Immunoassayed	Training	DILGOM07	FINRISK97
AAT	0.31	0.54	0.44	0.71
HP	0.59	0.84	0.82	0.82
AGP	0.65	0.85	0.83	0.83
TF	0.28	-	-	-

Table S7: Comparison of biomarkers in meta-analysis of DILGOM07 and FINRISK97 for all outcomes for which at least one of AAT, HP, or AGP was a significant and replicable biomarker in **Figure 2**. Details of Cox proportional hazard ratios shown in **Figure S4**. The alphanumeric codes in the square brackets indicate the ICD10 codes or disease categories for each outcome. A "*" next to AAT, HP, or AGP indicates that its association was significant and replicable (Storey Tibshirani FDR adjusted P < 0.05/3 in DILGOM07, FINRISK97, and meta-analysis). #S: total number of samples in the meta-analysis. #E: total number of events that occurred in the meta-analysis. #P: number of prevalent cases prior to baseline in all samples in the meta-analysis. HR: hazard ratio for combined incidence and mortality conferred per standard deviation increase of AAT, HP, AGP, or GlycA. 95% CI: 95% confidence interval for the hazard ratio. P: hazard ratio P-value. Q: Storey-Tibshirani FDR adjusted P-value (**Methods**).

Outcome	Biomarker	#S	#E	#P	HR	95% CI	P	Q
Intestinal infections [A00-A09]	*AAT	11,742	190	215	1.38	1.21–1.57	1×10^{-6}	2×10^{-5}
	HP	11,685	189	214	1.35	1.16–1.57	1×10^{-4}	0.001
	AGP	11,625	189	213	1.24	1.06–1.45	0.007	0.04
	GlycA	11,861	192	217	1.49	1.24–1.78	2×10^{-5}	2×10^{-4}
Gastroenteritis [A09]	*AAT	11,742	146	137	1.41	1.22–1.63	3×10^{-6}	3×10^{-5}
	HP	11,685	145	136	1.43	1.20–1.70	5×10^{-5}	7×10^{-4}
	AGP	11,625	145	137	1.29	1.08–1.54	0.004	0.03
	GlycA	11,861	148	138	1.51	1.24–1.85	6×10^{-5}	7×10^{-4}
Ischaemic heart diseases [I20-I25]	*AAT	11,742	925	502	1.21	1.13–1.29	2×10^{-8}	5×10^{-7}
	*HP	11,685	914	497	1.20	1.12–1.30	1×10^{-6}	3×10^{-5}
	AGP	11,625	916	496	1.18	1.10–1.27	1×10^{-5}	5×10^{-4}
	GlycA	11,861	938	509	1.31	1.19–1.43	7×10^{-9}	7×10^{-7}
Heart failure [I50]	*AAT	11,742	195	35	1.60	1.41–1.82	5×10^{-13}	1×10^{-10}
	*HP	11,685	192	35	1.59	1.37–1.85	1×10^{-9}	7×10^{-8}
	*AGP	11,625	191	35	1.56	1.35–1.81	4×10^{-9}	1×10^{-6}
	GlycA	11,861	195	35	1.66	1.38–2.00	9×10^{-8}	3×10^{-6}
Arterial system diseases [I70-I79]	*AAT	11,742	264	104	1.39	1.24–1.55	2×10^{-8}	5×10^{-7}
	*HP	11,685	263	104	1.49	1.31–1.69	1×10^{-9}	7×10^{-8}
	AGP	11,625	260	105	1.45	1.27–1.65	2×10^{-8}	2×10^{-6}
	GlycA	11,861	267	106	1.55	1.32–1.83	1×10^{-7}	3×10^{-6}
Atherosclerosis [I70]	*AAT	11,742	175	64	1.52	1.33–1.75	1×10^{-9}	4×10^{-8}
	*HP	11,685	173	64	1.67	1.43–1.94	5×10^{-11}	7×10^{-9}
	AGP	11,625	171	64	1.54	1.32–1.80	7×10^{-8}	4×10^{-6}
	GlycA	11,861	176	65	1.86	1.55–2.25	6×10^{-11}	9×10^{-9}
Influenza and pneumonia [J10-J18]	*AAT	11,742	451	244	1.37	1.25–1.50	6×10^{-12}	6×10^{-10}
	*HP	11,685	447	243	1.31	1.19–1.45	1×10^{-7}	5×10^{-6}
	AGP	11,625	446	242	1.23	1.10–1.36	1×10^{-4}	0.004
	GlycA	11,861	457	247	1.37	1.20–1.56	2×10^{-6}	4×10^{-5}
Pneumonia [J18]	*AAT	11,742	314	156	1.35	1.22–1.51	3×10^{-8}	6×10^{-7}
	HP	11,685	311	155	1.25	1.11–1.41	3×10^{-4}	0.003
	AGP	11,625	310	155	1.16	1.03–1.32	0.02	0.08
	GlycA	11,861	319	159	1.26	1.08–1.47	0.004	0.02
Chronic lower respiratory diseases [J40-J47]	*AAT	11,742	603	365	1.31	1.21–1.42	3×10^{-11}	2×10^{-9}
	*HP	11,685	596	364	1.36	1.25–1.49	1×10^{-11}	4×10^{-9}
	*AGP	11,625	596	364	1.31	1.19–1.43	1×10^{-8}	2×10^{-6}
	GlycA	11,861	607	368	1.36	1.22–1.52	2×10^{-8}	1×10^{-6}
Chronic obstructive pulmonary diseases [J44]	*AAT	11,742	186	51	1.54	1.34–1.77	9×10^{-10}	3×10^{-8}
	*HP	11,685	182	50	1.53	1.30–1.81	3×10^{-7}	1×10^{-5}
	AGP	11,625	184	51	1.33	1.13–1.56	7×10^{-4}	0.01
	GlycA	11,861	188	55	1.57	1.29–1.90	4×10^{-6}	8×10^{-5}

Outcome	Biomarker	#S	#E	#P	HR	95% CI	P	Q
Liver diseases [K70-K77]	*AAT	11,742	76	48	1.81	1.46–2.25	7×10^{-8}	1×10^{-6}
	HP	11,685	70	46	1.30	0.98–1.71	0.07	0.2
	AGP	11,625	74	47	1.05	0.79–1.40	0.7	0.8
	GlycA	11,861	83	50	1.75	1.30–2.35	2×10^{-4}	0.002
Inflammatory polyarthropathies [M05-M14]	*AAT	11,742	327	213	1.36	1.24–1.49	6×10^{-11}	3×10^{-9}
	*HP	11,685	323	208	1.42	1.27–1.59	9×10^{-10}	7×10^{-8}
	AGP	11,625	322	210	1.38	1.23–1.54	3×10^{-8}	2×10^{-6}
	GlycA	11,861	332	215	1.46	1.28–1.68	4×10^{-8}	1×10^{-6}

Table S8: Association between AAT and replicable gene coexpression network modules in DILGOM07. The tested gene coexpression network modules are those that were previously identified in DILGOM07 and found to topologically replicate in an independent cohort study (**Methods**). Linear regression models were fit between log-transformed AAT and each modules summary expression profile (first principal component) adjusting for participant age and sex. Modules were given numeric labels in descending order of module size (40 modules were identified, 20 were found to be topologically replicable). Module names are given based on GO term enrichment performed in previous studies, with the exception of module 5 whose enrichment (**Methods**) is first reported in this study (**Table S9**). Effect size indicates the change in standard deviations of immunoassayed AAT conferred per standard deviation increase of each module's coordinated expression adjusting for participant age and sex. CI 95% indicates the 95% confidence interval for the effect size. An association was considered significant (top three modules) where its P-value < 0.0025 (0.05/20, Bonferroni correcting for the total number of modules tested).

Module	Name	Genes	Effect size	CI 95%	P-value
6	General immune module A	1,019	0.17	[0.084, 0.25]	1×10^{-4}
10	General immune module B	339	0.15	[0.066, 0.24]	7×10^{-4}
5	RNA processing module	1,734	-0.14	[-0.23, -0.056]	0.001
22	B cell activity module	67	-0.13	[-0.22, -0.046]	0.003
29	Neutrophil module	31	0.11	[0.019, 0.20]	0.02
14	-	225	0.10	[0.013, 0.19]	0.02
16	Cytotoxic cell-like module	179	-0.098	[-0.19, -0.0073]	0.03
18	Viral response module	112	0.089	[0.0026, 0.17]	0.04
7	-	604	-0.083	[-0.17, 0.0059]	0.07
26	-	40	-0.071	[-0.16, 0.015]	0.1
12	-	255	-0.065	[-0.15, 0.024]	0.2
27	-	39	0.064	[-0.024, 0.15]	0.2
9	-	545	-0.049	[-0.14, 0.037]	0.3
33	-	28	0.045	[-0.042, 0.13]	0.3
38	Lipid leukocyte module	15	0.031	[-0.055, 0.12]	0.5
23	-	64	-0.024	[-0.11, 0.062]	0.6
2	-	5,403	0.011	[-0.078, 0.099]	0.8
4	-	1,775	-0.0080	[-0.096, 0.080]	0.9
17	Platelet module	138	-0.0050	[-0.092, 0.082]	0.9
13	-	239	-0.0038	[-0.091, 0.083]	0.9

Table S9: Enrichment analysis for the RNA processing gene coexpression network module. A) Gene Ontology (GO) biological process terms significantly over-represented in the core gene set for the RNA processing module (**Methods**). #AG: number of genes on the Illumina HT-12 array annotated with the corresponding GO term (17,519 genes on the array with any GO annotation). #MG: number of core module genes annotated for the corresponding GO terms (183 core module genes). P-value: P-value from a hypergeometric test. FDR: Benjamini-Hochberg FDR adjusted P-value. GO terms are displayed where FDR < 0.05. **B)** REVIGO ranking of significant GO terms by information content. Frequency: "percentage" of array annotated for the corresponding GO term. Uniqueness: REVIGO measure of semantic uniqueness when compared to all other GO terms. Dispensability: REVIGO measure of semantic redundancy when compared to its most similar GO term.

A)

GO ID	GO Term	#AG	#MG	P-value	FDR
GO:0006396	RNA processing	721	27	8×10^{-9}	1×10^{-4}
GO:0010467	gene expression	941	30	4×10^{-8}	3×10^{-4}
GO:0034641	cellular nitrogen compound metabolic process	4,948	85	1×10^{-7}	5×10^{-4}
GO:0006807	nitrogen compound metabolic process	5,282	86	1×10^{-6}	4×10^{-3}
GO:0046483	heterocycle metabolic process	4,371	74	3×10^{-6}	8×10^{-3}
GO:0006139	nucleobase-containing compound metabolic process	4,184	71	5×10^{-6}	0.01
GO:0034470	ncRNA processing	313	14	5×10^{-6}	0.01
GO:0006725	cellular aromatic compound metabolic process	4,386	73	6×10^{-6}	0.01
GO:0007005	mitochondrion organization	496	17	2×10^{-5}	0.03
GO:0016071	mRNA metabolic process	545	18	2×10^{-5}	0.03
GO:1901360	organic cyclic compound metabolic process	4,616	74	2×10^{-5}	0.03
GO:0071826	ribonucleoprotein complex subunit organization	192	10	3×10^{-5}	0.04

B)

GO ID	GO Term	Frequency	Uniqueness	Dispensability
GO:0006396	RNA processing	1.63%	0.467	0
GO:0007005	mitochondrion organization	0.65%	0.711	0.046
GO:0006807	nitrogen compound metabolic process	26.57%	0.642	0.100
GO:0010467	gene expression	15.38%	0.542	0.283
GO:1901360	organic cyclic compound metabolic process	25.43%	0.572	0.299
GO:0071826	ribonucleoprotein complex subunit organization	0.33%	0.772	0.341
GO:0034641	cellular nitrogen compound metabolic process	25.26%	0.430	0.354
GO:0006139	nucleobase-containing compound metabolic process	24.26%	0.391	0.697
GO:0046483	heterocycle metabolic process	24.80%	0.534	0.361
GO:0006725	cellular aromatic compound metabolic process	24.91%	0.534	0.362
GO:0034470	ncRNA processing	0.55%	0.506	0.384
GO:0016071	mRNA metabolic process	1.10%	0.482	0.415

Table S10: Curated gene sets significantly enriched for genes associated with AAT. Hallmark pathways, KEGG pathways, Reactome pathways, GO biological process (GO:BP) terms, GO molecular function (GO:MF) terms, and GO cellular compartment (GO:CC) terms significantly enriched for AAT-associated differential expression (**Methods**). Size: number of genes on the Illumina HT-12 array annotated for the corresponding gene set. NES: enrichment score normalised by the average enrichment score for the respective gene set in a permutation procedure with 1,000 permutations (**Methods**). A positive NES indicates the gene set is enriched for genes upregulated with elevated AAT while a negative NES indicates the gene set is enriched for genes downregulated with elevated AAT. FDR: Benjamini-Hochberg FDR adjusted permutation test P-value for enrichment. FDR correction was performed within each collection separately. Gene sets with FDR<0.05 are shown.

Collection	Gene set	Size	NES	FDR
Hallmark	Reactive oxygen species pathway	43	2.21	0.002
Hallmark	Inflammatory response	193	2.09	0.004
Hallmark	TNFa signaling via NFkB	194	1.92	0.01
Hallmark	PI3K/AKT/mTOR signaling	101	1.88	0.02
Hallmark	IL6/JAK/STAT3 signaling	81	1.94	0.02
Hallmark	Apoptosis	154	1.85	0.02
Hallmark	Complement	188	1.80	0.02
Hallmark	Hypoxia	186	1.83	0.02
Hallmark	Interferon gamma response	195	1.81	0.02
Hallmark	Interferon alpha response	92	1.74	0.03
KEGG	Leishmania infection	65	1.95	0.04
KEGG	Toll-like receptor signaling pathway	96	2.05	0.04
KEGG	Type II diabetes mellitus	43	1.98	0.04
KEGG	Fc gamma R-mediated phagocytosis	92	1.89	0.05
KEGG	NOD-like receptor signaling pathway	57	1.82	0.05
KEGG	Systemic lupus erythematosus	123	1.84	0.05
KEGG	Insulin signaling pathway	127	1.78	0.05
KEGG	Amino sugar and nucleotide sugar metabolism	42	1.76	0.05
KEGG	ErbB signaling pathway	81	1.85	0.05
KEGG	Progesterone mediated oocyte maturation	77	1.77	0.05
KEGG	GnRH signaling pathway	94	1.79	0.05
KEGG	Nicotinate and nicotinamide metabolism	22	1.80	0.05
Reactome	Activated TLR4 signalling	82	1.95	0.04
Reactome	Meiotic synapsis	66	1.93	0.04
Reactome	MYD88:Mal cascade initiated on plasma membrane	72	1.98	0.04
Reactome	MAPK targets nuclear events mediated by MAP kinases	28	1.96	0.04
Reactome	GABA(B) receptor activation	38	1.94	0.04
Reactome	Toll receptor cascades	102	1.98	0.04
Reactome	Packaging of telomere ends	46	1.90	0.05
Reactome	Latent infection of homo sapiens with mycobacterium tuberculosis	28	1.90	0.05
Reactome	Amyloids	73	1.99	0.05
GO:BP	Cell redox homeostasis	59	2.22	0.005
GO:BP	Cytokine production involved in immune response	17	2.24	0.006
GO:BP	Membrane invagination	27	2.22	0.006
GO:BP	Regulation of interleukin 8 production	55	2.14	0.008
GO:BP	Cellular response to mechanical stimulus	73	2.17	0.008
GO:BP	Positive regulation of interleukin 1 secretion	24	2.15	0.009
GO:BP	Defense response to bacterium	188	2.15	0.01
GO:BP	Regulation of interleukin 1 secretion	32	2.25	0.01
GO:BP	CAMP biosynthetic process	16	2.10	0.02
GO:BP	Response to bacterium	464	2.05	0.02
GO:BP	Acute phase response	38	2.05	0.02
GO:BP	Positive regulation of interleukin 1 production	36	2.05	0.02
GO:BP	Defense response to gram positive bacterium	65	2.07	0.02

Collection	Gene set	Size	NES	FDR
GO:BP	Defense response to gram negative bacterium	39	2.06	0.02
GO:BP	Cytokine production	111	2.06	0.02
GO:BP	Positive regulation of interleukin 8 production	42	2.02	0.02
GO:BP	Fc gamma receptor signaling pathway	72	2.02	0.02
GO:BP	Organ or tissue specific immune response	27	2.03	0.02
GO:BP	Regulation of myoblast fusion	18	2.01	0.02
GO:BP	Positive regulation of interleukin 1 beta production	30	2.03	0.02
GO:BP	Defense response to other organism	432	1.99	0.03
GO:BP	Response to muscle stretch	18	2.00	0.03
GO:BP	Cytokine secretion	35	1.99	0.03
GO:BP	Phagocytosis	156	1.98	0.03
GO:BP	Dendritic cell differentiation	32	1.98	0.03
GO:BP	Regulation of tumor necrosis factor superfamily cytokine production	95	1.98	0.03
GO:BP	NADP metabolic process	24	1.98	0.03
GO:BP	Detection of biotic stimulus	26	1.96	0.03
GO:BP	Regulation of cytokine secretion	139	1.96	0.03
GO:BP	T cell differentiation involved in immune response	28	1.96	0.03
GO:BP	Regulation of NF-kappaB import into nucleus	43	1.96	0.03
GO:BP	Antimicrobial humoral response	43	1.97	0.03
GO:BP	Cellular response to external stimulus	244	1.95	0.03
GO:BP	Interleukin 1 production	15	1.95	0.03
GO:BP	Regulation of Toll-like receptor signaling pathway	44	1.95	0.04
GO:BP	positive regulation of Toll-like receptor signaling pathway	19	1.92	0.04
GO:BP	Regulation of syncytium formation by plasma membrane fusion	23	1.92	0.04
GO:BP	Cellular response to starvation	107	1.92	0.04
GO:BP	Defense response to fungus	35	1.93	0.04
GO:BP	Positive regulation of mitotic nuclear division	46	1.93	0.04
GO:BP	Ammonium transport	57	1.93	0.04
GO:BP	Regulation of tumor necrosis factor biosynthetic process	18	1.92	0.04
GO:BP	Positive regulation of reactive oxygen species metabolic process	75	1.93	0.04
GO:BP	MyD88 dependent Toll-like receptor signaling pathway	25	1.93	0.04
GO:BP	Negative regulation of biomineral tissue development	17	1.93	0.04
GO:BP	Response to molecule of bacterial origin	304	1.93	0.04
GO:BP	Regulation of translation in response to stress	17	1.89	0.04
GO:BP	Positive regulation of nuclear division	57	1.89	0.04
GO:BP	Myeloid leukocyte activation	92	1.90	0.04
GO:BP	Acute inflammatory response	68	1.89	0.04
GO:BP	Protein secretion	103	1.90	0.04
GO:BP	Detection of other organism	18	1.89	0.04
GO:BP	Ectoderm development	19	1.89	0.04
GO:BP	Regulation of T helper cell differentiation	25	1.90	0.04
GO:BP	Response to starvation	141	1.89	0.04
GO:BP	Positive regulation of neuron apoptotic process	44	1.90	0.04
GO:BP	Regulation of interleukin 1 production	57	1.89	0.04
GO:BP	Inflammatory response	426	1.90	0.04
GO:BP	DNA replication dependent nucleosome organization	30	1.88	0.04
GO:BP	Negative regulation of blood circulation	31	1.90	0.04
GO:BP	Response to fungus	48	1.88	0.04
GO:BP	Antigen processing and presentation of exogenous peptide antigen via MHC class I	63	1.87	0.05
GO:BP	Cellular response to growth hormone stimulus	20	1.88	0.05
GO:BP	Regulation of viral entry into host cell	27	1.87	0.05
GO:BP	Negative regulation of oxidative stress induced intrinsic apoptotic signaling pathway	19	1.86	0.05
GO:BP	Trachea development	19	1.87	0.05
GO:BP	Regulation of innate immune response	325	1.86	0.05
GO:BP	Response to mechanical stimulus	196	1.87	0.05
GO:BP	Regulation of interleukin 1 beta production	47	1.86	0.05
GO:BP	Phagocytosis engulfment	17	1.87	0.05
GO:BP	Antigen processing and presentation of peptide antigen via MHC class I	86	1.86	0.05
GO:BP	Peptidyl cysteine modification	18	1.86	0.05

Collection	Gene set	Size	NES	FDR
GO:BP	Regulation of the force of heart contraction	27	1.85	0.05
GO:BP	Innate immune response in mucosa	18	1.85	0.05
GO:BP	Immune effector process	426	1.86	0.05
GO:BP	Regulation of interleukin 6 production	95	1.85	0.05
GO:BP	Embryonic placenta development	76	1.85	0.05
GO:BP	T-cell activation involved in immune response	58	1.86	0.05
GO:BP	STAT cascade	50	1.85	0.05
GO:BP	Negative regulation of viral process	84	1.85	0.05
GO:BP	Negative regulation of immune response	113	1.85	0.05
GO:BP	Receptor catabolic process	15	1.83	0.05
GO:BP	Regulation of type 2 immune response	25	1.82	0.05
GO:BP	Positive regulation of response to extracellular stimulus	46	1.83	0.05
GO:BP	Activation of immune response	366	1.82	0.05
GO:BP	Myeloid dendritic cell differentiation	20	1.83	0.05
GO:MF	Solute cation antiporter activity	26	2.11	0.008
GO:MF	Ion antiporter activity	42	2.14	0.009
GO:MF	Cation-cation antiporter activity	20	2.19	0.009
GO:MF	Cysteine type endopeptidase activity involved in apoptotic process	15	2.03	0.02
GO:MF	Lysophospholipid acyltransferase activity	18	2.00	0.03
GO:MF	Signaling pattern recognition receptor activity	17	1.98	0.03
GO:MF	Acylglycerol O-acyltransferase activity	25	1.94	0.03
GO:MF	Immunoglobulin binding	22	1.95	0.04
GO:MF	G protein beta gamma subunit complex binding	19	1.93	0.04
GO:MF	O-acyltransferase activity	43	1.91	0.04
GO:CC	Phagocytic vesicle	74	2.11	0.008
GO:CC	Phagocytic vesicle membrane	48	1.95	0.03
GO:CC	Primary lysosome	16	1.95	0.04
KEGG	Homologous recombination	24	-1.88	0.03
KEGG	Base excision repair	31	-1.84	0.04
KEGG	Lysine degradation	42	-1.81	0.04
KEGG	DNA replication	33	-1.83	0.04
KEGG	Cysteine and methionine metabolism	33	-1.89	0.04
KEGG	Nucleotide excision repair	39	-1.77	0.05
KEGG	Histidine metabolism	27	-1.74	0.05
GO:BP	RNA secondary structure unwinding	38	-2.25	0.002
GO:MF	Lysine acetylated histone binding	15	-1.93	0.05