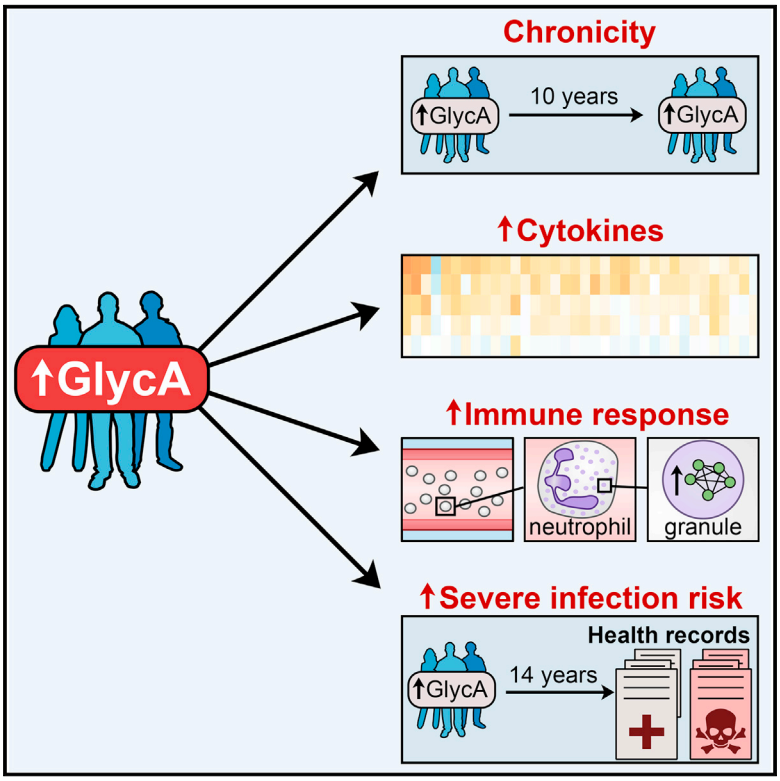
Cell Systems

Report

The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection

Graphical Abstract



Authors

Scott C. Ritchie, Peter Würtz, Artika P. Nath, ..., Mika Ala-Korpela, Johannes Kettunen, Michael Inouye

Correspondence

johannes.kettunen@computationalmedicine.fi (J.K.), minouye@unimelb.edu.au (M.I.)

In Brief

Ritchie et al. investigate the biology of GlycA, a known biomarker for short-term mortality. They reveal GlycA's long-term behavior in apparently healthy patients: it is stable for >10 years and associated with chronic low-grade inflammation. Accordingly, GlycA predicts death from infection up to 14 years in the future.

Highlights

- Elevated GlycA was stable within individuals for up to a decade
- GlycA marked the levels of myriad inflammatory cytokines in circulation
- A gene network enriched for neutrophil functions was associated with GlycA
- GlycA strongly predicted future risk of hospitalization and death from infection

Ritchie et al., 2015, Cell Systems *1*, 293–301
October 28, 2015
http://dx.doi.org/10.1016/j.cels.2015.09.007

CellPress

Cell Systems
Report

The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection

Scott C. Ritchie,[1,2] Peter Würtz,[4] Artika P. Nath,[1,3] Gad Abraham,[1,2] Aki S. Havulinna,[5,6] Liam G. Fearnley,[1,2] Antti-Pekka Sarin,[6] Antti J. Kangas,[4] Pasi Soininen,[4,7] Kristiina Aalto,[8] Ilkka Seppälä,[9] Emma Raitoharju,[9] Marko Salmi,[5,8] Mikael Maksimow,[5,8] Satu Männistö,[5] Mika Kähönen,[10] Markus Juonala,[11,12] Samuli Ripatti,[6,13] Terho Lehtimäki,[9] Sirpa Jalkanen,[8] Markus Perola,[5,6] Olli Raitakari,[14,15] Veikko Salomaa,[5] Mika Ala-Korpela,[4,7,16,17,18] Johannes Kettunen,[4,5,7,19,*] and Michael Inouye[1,2,3,19,*]

[1]Centre for Systems Genomics, School of BioSciences, The University of Melbourne, Parkville, 3010 Victoria, Australia
[2]Department of Pathology, The University of Melbourne, Parkville, 3010 Victoria, Australia
[3]Department of Microbiology and Immunology, The University of Melbourne, Parkville, 3010 Victoria, Australia
[4]Computational Medicine, Institute of Health Sciences, University of Oulu, Oulu 90014, Finland
[5]National Institute for Health and Welfare, Helsinki 00271, Finland
[6]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki 00014, Finland
[7]NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio 70211, Finland
[8]MediCity Research Laboratory, Department of Medical Microbiology and Immunology, University of Turku, Turku 20520, Finland
[9]Department of Clinical Chemistry, Fimlab Laboratories, School of Medicine, University of Tampere, Tampere 33520, Finland
[10]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, FI-33521 Tampere, Finland
[11]Department of Medicine, University of Turku and Division of Medicine, Turku University Hospital, FI-20520 Turku, Finland
[12]Murdoch Childrens Research Institute, Parkville, 3052 Victoria, Australia
[13]Department of Public Health, University of Helsinki, Helsinki 00014, Finland
[14]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland
[15]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland
[16]Oulu University Hospital, Oulu 90220, Finland
[17]Computational Medicine, School of Social and Community Medicine, University of Bristol, Bristol BS8 1TH, UK
[18]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK
[19]Co-senior author
*Correspondence: johannes.kettunen@computationalmedicine.fi (J.K.), minouye@unimelb.edu.au (M.I.)
http://dx.doi.org/10.1016/j.cels.2015.09.007

SUMMARY

The biomarker glycoprotein acetylation (GlycA) has been shown to predict risk of cardiovascular disease and all-cause mortality. Here, we characterize biological processes associated with GlycA by leveraging population-based omics data and health records from >10,000 individuals. Our analyses show that GlycA levels are chronic within individuals for up to a decade. In apparently healthy individuals, elevated GlycA corresponded to elevation of myriad inflammatory cytokines, as well as a gene coexpression network indicative of increased neutrophil activity, suggesting that individuals with high GlycA may be in a state of chronic inflammatory response. Accordingly, analysis of infection-related hospitalization and death records showed that increased GlycA increased long-term risk of severe non-localized and respiratory infections, particularly septicaemia and pneumonia. In total, our work demonstrates that GlycA is a biomarker for chronic inflammation, neutrophil activity, and risk of future severe infection. It also illustrates the utility of leveraging multi-layered omics data and health records to elucidate the molecular and cellular processes associated with biomarkers.

INTRODUCTION

The integration of large-scale, systems-wide biomolecular information with health records to identify statistical associations represents a foundation of future studies in "precision medicine." Early systems-wide studies have elucidated the etiology of complex diseases in natural human populations (Chen et al., 2012, 2008; Emilsson et al., 2008; Montoya et al., 2014; Stanberry et al., 2013; Wang et al., 2011; Zhang et al., 2013). In doing so, they have identified biomarkers of potential clinical utility. Exemplar are recent studies that used nuclear magnetic resonance (NMR) spectroscopy biomarker profiling to identify serum glycoprotein acetylation (GlycA) levels as a biomarker for risk of 5-year all-cause mortality and incident cardiovascular disease (CVD) events (Akinkuolie et al., 2014; Fischer et al., 2014). Fischer et al. showed that elevation of GlycA was predictive of death from all causes within 5 years in >17,000 generally healthy individuals from two independent population-based cohorts (Fischer et al., 2014). Each SD increase of GlycA above the population mean conferred a 67% and 55% increase in mortality risk in the 5-year follow-up period for Estonian and Finnish cohorts, respectively (Fischer et al., 2014). Similarly, Akinkuolie

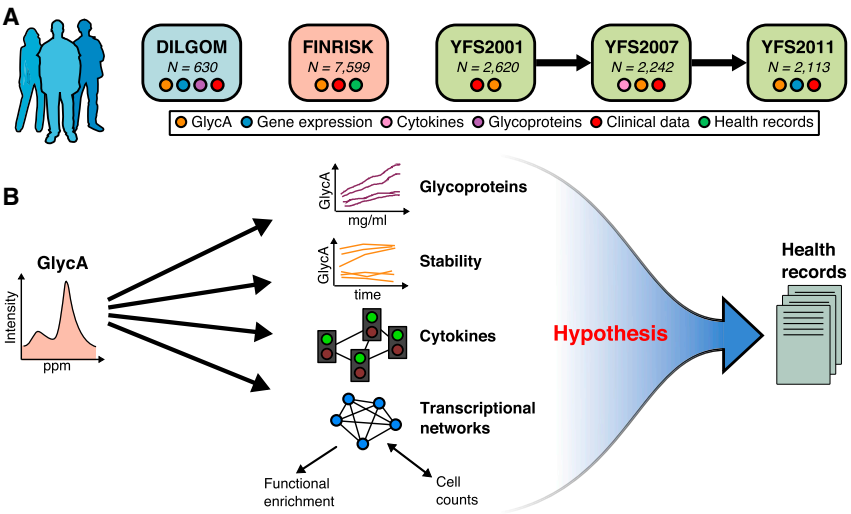


Figure 1. Study Guide
(A) Summary of available data for each cohort.
(B) Illustration of overall study flow.

et al. investigated GlycA in a recent study of 27,500 initially healthy women. They found that elevated GlycA was predictive of 15-year risk of CVD incidence and mortality (Akinkuolie et al., 2014). In both studies, the predictive capacity of GlycA was independent of age, sex, modifiable lifestyle risk factors, medication, and disease prevalence. Other studies on GlycA in relation to common chronic diseases are rapidly emerging (Akinkuolie et al., 2015; Ala-Korpela, 2015); however, the biological foundation of these associations remains unexplored.

GlycA itself is a complex heterogeneous NMR signal that reflects the abundance of mobile *N*-acetyl sugar groups found on glycoproteins in circulating blood (Bell et al., 1987; Otvos et al., 2015). Multiple circulating glycoproteins have been found to contribute to GlycA, including alpha-1-acid glycoprotein, alpha-1 antitrypsin, haptoglobin, transferrin (Bell et al., 1987), and, more recently, alpha-1 antichymotrypsin; however, their relative contribution to the GlycA signal remains unclear (Otvos et al., 2015). The glycoproteins contributing to GlycA are involved in the so-called "acute-phase response" to exogenous insults, including infection and physical injury. This is a broad set of systemic biochemical and physiological changes, primarily driven by cytokine production of inflammatory cells, which occur immediately following the insult (Gabay and Kushner, 1999). With the exception of transferrin, the circulating concentrations of the proteins that constitute the GlycA signal are known to increase during the acute-phase response (Aronsen et al., 1972; Gitlin and Colten, 1987). Accordingly, the GlycA signal itself is elevated during the acute phase in patient samples (Bell et al., 1987).

Although GlycA is understood to mark the acute-phase response, it has also been shown to be elevated in patients with a diverse range of inflammation-linked chronic conditions, including rheumatoid arthritis, hypertension, obesity, and metabolic disorders (Bell et al., 1987; Lauridsen et al., 2010; Otvos et al., 2015; Würtz et al., 2012, 2014). Clinical assessment of chronic inflammation is routinely used in patient monitoring, disease diagnosis, and risk assessment. Currently, the state of the art for assessing chronic inflammation in the clinic typically includes quantification of circulating C-reactive protein (CRP), a single molecular species for which highly sensitive immunoassays are widely available (Pearson et al., 2003). However, the efficacy of this single marker may be limited by false-negatives. Chronic inflammation is likely as heterogeneous as the acute-phase response, where biochemical and physiological changes are not uniform across inflammatory conditions (Ebersole and Cappelli, 2000; Morley and Kushner, 1982). Evidence is now emerging to suggest that elevated CRP is not a universal feature of chronic inflammation. Importantly, GlycA, a composite marker of multiple molecular species, has been shown to predict risk of all-cause mortality and CVD, even when adjusting for CRP (Akinkuolie et al., 2014; Fischer et al., 2014).

This study elucidates the molecular and cellular processes underlying circulating GlycA levels by utilizing a range of omics data and electronic health records from three large-scale population-based cohorts (n = 11,825 total). Our unbiased integrative analysis reveals that GlycA levels are persistent over time and positively associated with both diverse cytokines in individuals without recent infection and the neutrophil antimicrobial response. This suggests that apparently healthy individuals may be in a state of chronic inflammatory response. Accordingly, we demonstrate that GlycA is predictive of long-term risk of severe infection using a large prospective cohort with nearly 14 years of follow-up health records. The overall study flow is illustrated in Figure 1.

RESULTS

Study Data

In this study, we analyzed data collected by three population-based studies (Supplemental Experimental Procedures). The Dietary, Lifestyle, and Genetic Determinants of Obesity and Metabolic syndrome (DILGOM) Study is a cross-sectional cohort of 579 individuals (300 female; 52%) 25–74 years old (Inouye et al., 2010a, 2010b). The 1997 collection of the FINRISK Study is a cross-sectional cohort of 7,599 individuals (3,822 female; 50%) 24–74 years old with disease events followed prospectively via nation-wide electronic health records (Borodulin et al., 2015; Würtz et al., 2015). The Cardiovascular Risk in Young Finns Study (YFS) is a longitudinal population-based study of 3,596 individuals (1,832 female; 51%) recruited during childhood in 1980 (Raitakari et al., 2008). In this study, data were available from the 2001, 2007, and 2011 follow-up collections (YFS2001, YFS2007, and YFS2011, respectively) (Würtz et al., 2014). By the 2011 follow-up, individuals were 34–49 years old. Clinical and demographic characteristics for each study are summarized in Table 1.

For all studies, NMR biomarker profiling data, including data for GlycA, were available (Soininen et al., 2009), as were CRP immunoassays. To fine-map the GlycA signal, alpha-1-acid glycoprotein, alpha-1 antitrypsin, haptoglobin, and transferrin

Table 1. Cohort Characteristics

Characteristic	YFS 2001	YFS 2007	YFS 2011	DILGOM	FINRISK
Collection year	2001	2007	2011	2007	1997
Number of individuals	2,232	2,159	2,041	579	7,599
Number (and %) of women	1,234 (55%)	1,185 (55%)	1,112 (54%)	300 (52%)	3,822 (50%)
Mean age in years (and range)	31.7 (24–39)	37.7 (30–45)	41.9 (34–49)	52.1 (25–74)	48.3 (24–74)
Body mass index (kg/m^2)	25.1 ± 4.4	26.0 ± 4.8	26.5 ± 5.1	26.8 ± 4.7	26.7 ± 4.5
GlycA (mmol/l)	1.4 ± 0.2	1.6 ± 0.3	1.6 ± 0.2	1.6 ± 0.2	1.4 ± 0.2
Triglycerides (mmol/l)	1.3 ± 0.8	1.4 ± 0.9	1.3 ± 0.8	1.3 ± 0.6	1.5 ± 1.1
CRP (mg/l)	1.9 ± 3.8	1.9 ± 3.8	1.7 ± 3.2	2.7 ± 5.7	2.5 ± 5.3
Number (and %):					
Recent febrile infection	112 (5.0%)	95 (4.4%)	93 (4.6%)	–	–
Use of antihypertensive therapy	58 (2.6%)	146 (6.8%)	197 (9.7%)	118 (20%)	1,015 (13%)
Use of lipid-lowering therapy	7 (0.3%)	45 (2.1%)	74 (3.6%)	86 (15%)	269 (0.4%)
Prevalent diabetes	–	–	35 (1.3%)	61 (10%)	440 (5.8%)
Incident diabetes	–	–	–	8 (1.4%)	605 (8.0%)
Prevalent CVD	–	–	20 (0.7%)	17 (2.9%)	263 (3.5%)
Incident CVD	–	–	–	7 (1.2%)	802 (11%)
Prevalent cancer	–	–	40 (1.4%)	18 (3.1%)	180 (2.4%)
Incident cancer	–	–	–	11 (1.9%)	12 (0.2%)
Prevalent immunodeficiency	–	–	10 (0.4%)	0 (0%)	8 (0.1%)
Incident immunodeficiency	–	–	–	0 (0%)	12 (0.2%)
Total number of deaths	–	–	–	15 (2.6%)	839 (11%)
Number of glycoprotein assays	–	–	–	579	–
Alpha-1-acid glycoprotein (g/l)	–	–	–	0.8 ± 0.2	–
Alpha-1 antitrypsin (g/l)	–	–	–	1.2 ± 0.2	–
Haptoglobin (g/l)	–	–	–	1.1 ± 0.5	–
Transferrin (g/l)	–	–	–	2.7 ± 0.4	–
Number of cytokine assays	–	2,159	–	–	–
Cell count data	–	–	2,021	–	–
Gene expression profiling: number	–	–	1,649	518	–

Data are reported as the mean ± SD unless otherwise indicated. Prevalent disease indicates events occurring prior to sample collection, while incident disease indicates events occurring after sample collection. For the YFS, a disease was considered prevalent if it occurred any time prior to the 2011 sample collection for any of the 2,766 individuals participating at any one time point. Cytokine characteristics are reported in Table S1. Dashes indicate that data are not applicable.

were quantified by immunoturbidimetric methods in the DILGOM Study. Multiplexed cytokine panels were performed for YFS2007. Cytokine characteristics are reported in Table S1. Fasting whole-blood transcriptomic data (Illumina HT-12 arrays) were available for the DILGOM Study and YFS2011. Electronic health records were collated for the FINRISK Study only. There was an insufficient number of endpoint events for GlycA survival analysis in the DILGOM Study due to small sample size and short follow-up time. Because of their relative youth and short follow-up time, there were also insufficient numbers of endpoint events in the YFS. Unless otherwise noted, all molecular measurements in all cohorts were performed on fasting serum samples. Protocols and processing details for each platform and study are described in the Supplemental Experimental Procedures.

Specific Glycoprotein Levels Contributing to GlycA in a Population-Based Study

In contrast to previous studies that used protein standards to deconvolute the GlycA signal (Bell et al., 1987; Otvos et al., 2015), we utilized a population-based approach through the DILGOM Study to determine the relative contributions of serum alpha-1-acid glycoprotein, alpha-1 antitrypsin, haptoglobin, and transferrin levels to GlycA (Supplemental Experimental Procedures). As expected, we found that all four glycoproteins significantly contributed to GlycA signal intensity. Univariable linear regression indicated that each SD increase of alpha-1-acid glycoprotein, haptoglobin, alpha-1 antitrypsin, and transferrin corresponded to a 0.54 ($p = 2 \times 10^{-53}$), 0.52 ($p = 9 \times 10^{-40}$), 0.29 ($p = 3 \times 10^{-12}$), and 0.20 ($p = 1 \times 10^{-6}$) SD increase of GlycA,

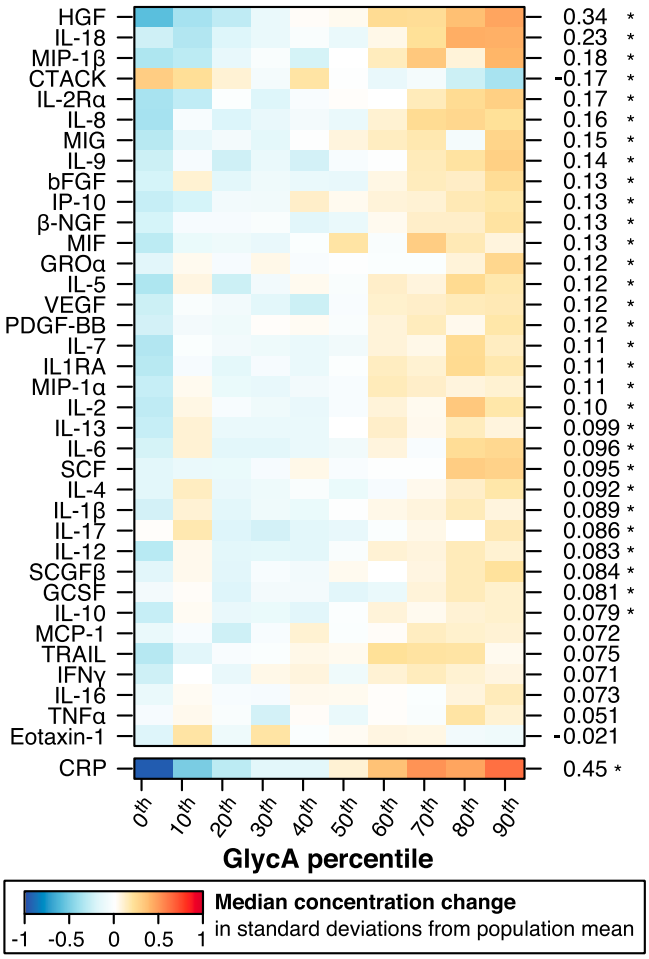


Figure 2. GlycA Summarizes Low-Grade Inflammation Cytokine Activity

Median difference in concentration (SDs) relative to the population mean for 36 assayed cytokines and CRP for ten GlycA-risk categories in the 2007 collection of the YFS cohort. The right axis denotes the change in SD units of GlycA per SD increase of each cytokine or CRP, adjusting for age and sex in the YFS cohort. *$p < 0.001$ (Bonferroni corrected for the number of cytokines). Table S1 gives the full name and Uniprot identifier for each cytokine.
See also Table S1.

respectively. Of the individuals in the top 10% for the GlycA levels, 40% were also in the top decile for alpha-1-acid glycoprotein levels, as were 43% for haptoglobin, 29% for alpha-1 antitrypsin, and 15% for transferrin. Pairwise correlations among the glycoproteins were either positive or, in the case of alpha-1-acid glycoprotein and transferrin, not significant. These analyses indicate that the GlycA signal is not dominated by any single glycoprotein; thus, GlycA is treated as a composite signal in all subsequent analyses.

GlycA Levels in Acute Phase, Chronicity over Time, and Relationship with Myriad Cytokines

All YFS collections contained samples from individuals reporting febrile infection within 2 weeks prior to sample collection. In these individuals, we observed modest elevation of GlycA in individuals from all YFS collections (mean increase of 0.41 SD; Figure S1), consistent with GlycA's role as a positive acute-phase reactant. Among YFS individuals not reporting a recent febrile infection, there was a strong autocorrelation of GlycA levels within individuals across all three time points (mean Pearson $r = 0.43$). Of the top 10% of individuals by GlycA level in 2001, 20% were also in the top decile in 2007, and 19% were also in the top decile in 2011. Individuals reporting a recent febrile infection were no more likely than others to have elevated GlycA levels at any subsequent time point.

Given that elevated GlycA levels correlated with both the acute-phase response and were stable over time, we asked whether GlycA was associated with inflammatory signatures in apparently healthy individuals. We utilized multiplex cytokine panels assayed in YFS2007 serum samples (Supplemental Experimental Procedures) from individuals reporting no recent febrile infections and observed a large number of robust associations (Bonferroni-adjusted $p < 0.001$): 30 of 36 tested cytokines were significantly associated with GlycA (Figure 2). The concentrations of 29 cytokines, with both anti- and pro- inflammatory roles, increased along with the GlycA signal (Figure 2). Only cutaneous T-cell-attracting chemokine (CTACK) showed a significant negative association with GlycA (Figure 2). Cytokine concentration changes associated with GlycA were relatively modest, with even the top decile of GlycA levels corresponding to cytokine elevations of 0.11–0.48 SD on average (Figure 2). Together, these results suggest an association between elevated GlycA and systemic low-grade inflammation in an apparently healthy population. It was not possible to test whether cytokine concentrations were stable over time, like those of GlycA, because cytokine panels were not available for YFS2001 or YFS2011.

A Transcriptional Subnetwork Associated with GlycA

When put in the context of previous studies, our results suggest that, in addition to marking the acute-phase response (Bell et al., 1987), high GlycA levels may also mark chronic inflammation in apparently healthy individuals. This suggests an intimate connection between the circulating, systemic concentration of GlycA and cellular-level changes in physiology. To gain insight into cellular processes associated with elevated GlycA, we utilized whole-blood transcriptome profiling in the DILGOM and YFS2011 cohorts. We inferred weighted gene coexpression network modules (Zhang and Horvath, 2005) in the DILGOM cohort and then tested each module for an association with the GlycA signal (Supplemental Experimental Procedures). Briefly, the full gene coexpression network was inferred as the pairwise Spearman correlation between all microarray probes. This was converted to a measure of network adjacency through a soft-threshold power transform. Modules were then detected through a dynamic tree cut algorithm on the topological overlap similarity of probe adjacencies. A summary expression vector (first principal component) was calculated for each module and used to test for the association with GlycA. Models were adjusted for age, sex, and, additionally, serum triglyceride levels, which have been shown to modify the multivariable effect of lipoprotein concentrations on the GlycA signal and associated mortality risk (Fischer et al., 2014). Significantly associated modules were subsequently tested for replication in the YFS cohort using NetRep (v0.10.4), a rapid permutation-based approach of

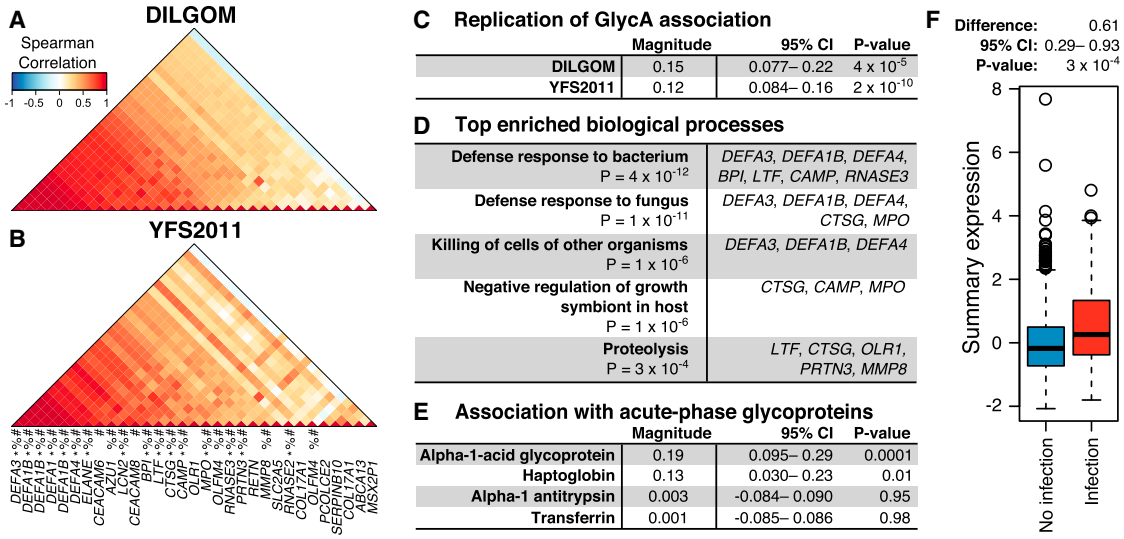


Figure 3. Neutrophil Gene Coexpression Module Associated with GlycA

(A and B) Probe coexpression (Spearman correlation) in the DILGOM cohort (A) and replication in the YFS2011 cohort (B). See Table S3 for module gene details. Symbols above gene symbol indicate expression in neutrophils (#), whether its product is located in neutrophil granules (%), and whether its product has antimicrobial functionality (*). See Table S5 for a summary of the literature for each gene.

(C) Replication of GlycA association in both cohorts. Associations were assessed using a linear regression of GlycA on the module summary expression adjusting for age, sex, and triglycerides. Magnitude corresponds to change in SD units of GlycA per SD increase of neutrophil module expression. CI, confidence interval.

(D) The top enriched GO biological process terms. A full list of significant GO terms and KEGG pathway enrichments can be found in Table S4.

(E) Association of neutrophil module expression with alpha-1-acid glycoprotein, haptoglobin, alpha-1 antitrypsin, and transferrin in the DILGOM cohort in a multivariable linear regression. Magnitude corresponds to change in SD units of GlycA per SD increase of each respective acute-phase glycoprotein.

(F) Boxplots showing association with recent febrile infection (self-reported) (n = 72 with febrile infection, and n = 1,550 without). T-test difference, 95% confidence interval, and p value are reported. GlycA and triglyceride levels were log-transformed. All continuous measurements were standardized.

See also Tables S3, S4, and S5.

network topology features (Supplemental Experimental Procedures) (Langfelder et al., 2011). NetRep is available at https://github.com/InouyeLab/NetRep/.

We identified 40 coexpression modules within the DILGOM whole-blood expression data, of which 3 were significantly associated with GlycA after adjusting for multiple testing ($p < 0.001$). One module's network topology and GlycA association was replicated in the YFS2011 whole-blood expression data (permutation test p value < 0.01 for all network-based statistics and GlycA association p value < 0.001; Table S2). We subsequently performed in-depth analysis of the module (Figure 3) and its 27 genes (Table S3). For this module, single-gene associations with GlycA were consistent with the whole-module association with GlycA.

The enriched GO processes and KEGG pathways of the module largely implicated innate immune response to microbes (Table S4). Consistent with this, module expression was elevated by 0.61 SDs ($p = 3 \times 10^{-4}$) in individuals reporting recent febrile infection (Figure 3). Most genes in the preserved module were primarily reported in the literature as coding for antimicrobial peptides synthesized in and secreted from neutrophil granules (14 genes, 52%; Figure 3; Table S5). Genes with antimicrobial function constituted 11 of the top 14 genes when ranked by mean intramodular connectivity across datasets (Table S3), a topological measure known to correspond to relative biological importance (Langfelder et al., 2013). A further three genes coded for proteins serving a non-antimicrobial role in neutrophil-granule-associated processes (Table S5), and two genes coded for proteins expressed on the surface of neutrophils (Table S5). Since a majority of these genes have well-characterized antimicrobial roles in neutrophil granules (Faurschou and Borregaard, 2003), we refer hereinafter to this module as the neutrophil module.

Neutrophils are the dominant white blood cell type in adult humans, but their abundance varies in populations and within individuals over time (Summers et al., 2010). While neutrophil cell counts were not available, we observed a strong positive association between overall leukocyte abundance and elevated neutrophil module summary expression ($\beta = 0.24$ SD leukocyte abundance increase per SD increase of neutrophil module expression, $p = 2 \times 10^{-17}$) in the YFS2011, suggesting that elevated neutrophil module expression may be associated, to some extent, with increased circulating neutrophil abundance. Adjustment of the GlycA model for leukocyte abundance showed that neutrophil module expression retained a strong association with GlycA; however, both the module and leukocyte abundance were highly significant ($\beta = 0.09$, $p = 2 \times 10^{-6}$, and $\beta = 0.15$, $p = 2 \times 10^{-15}$, respectively), indicating that increased GlycA corresponds to increased production of both antimicrobial peptides and circulating leukocytes. Neutrophil module expression was significantly associated with alpha-1-acid glycoprotein and haptoglobin ($p = 1 \times 10^{-4}$ and $p = 0.01$, respectively; Figure 3E). Both proteins are synthesized in and secreted from neutrophil granules (Theilgaard-Mönch et al., 2005, 2006),

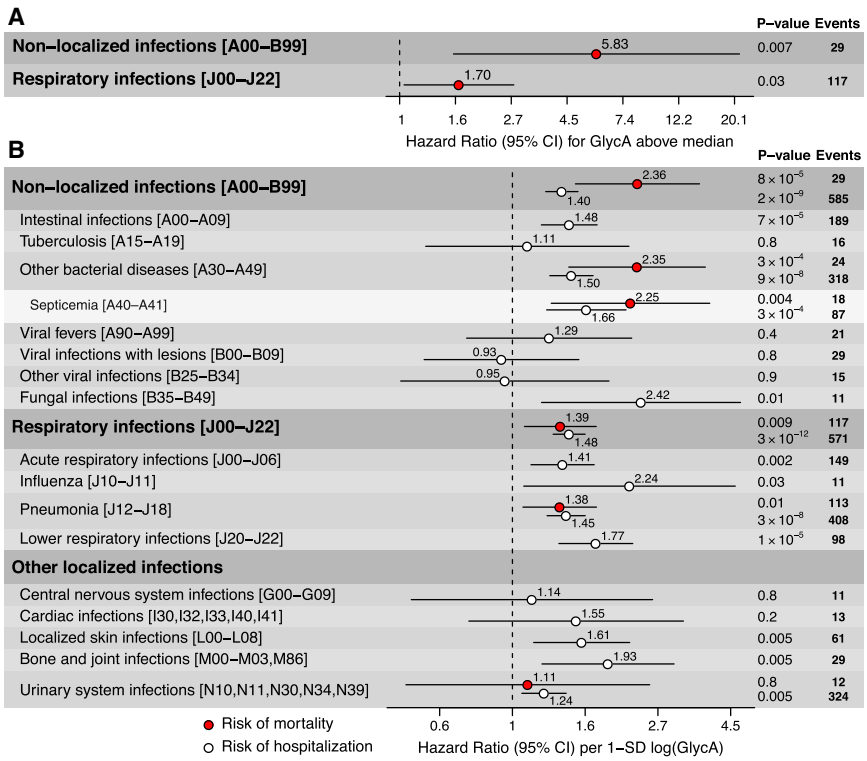


Figure 4. GlycA Predicts Hospitalization and Death from Infectious Diseases

(A and B) Hazard ratios (HRs) for fatal events from broad infection-related categories for (A) individuals with GlycA concentrations above the population median, and (B) risk of fatal events (red) or hospitalization (white) conferred per SD increment of log-transformed GlycA for infection-related diagnoses with more than ten events in the FINRISK 1997 cohort. 7,599 apparently healthy individuals from the general population were prospectively observed over a 13.8-year follow-up period. Cox models were adjusted for age, sex, triglycerides, and incidence of the same diagnosis in the 10 year prior to sample collection. Numbers in square brackets indicate the International Classification of Diseases (tenth edition; ICD-10) code blocks for the corresponding diagnosis. Figure S2 gives an extended breakdown into individual ICD-10 diagnosis codes, and Figure S3 shows sensitivity analyses of HRs when adjusted for prevalent chronic diseases, compared to CRP, and adjusted for CRP. CI, confidence interval. See also Figures S2 and S3.

suggesting an overabundance of neutrophils and that their granule content may directly contribute to elevated GlycA.

GlycA Is Associated with Risk of Hospitalization and Death from Microbial Infection

Because GlycA and the microbial-associated inflammatory response were statistically associated, and it is known that the inflammatory response correlates with adverse outcomes in the clinic, we hypothesized that GlycA level may also be reflective of future infection severity risk. We asked whether there was a relationship between risk of future hospitalization or death and elevated GlycA levels. To do this, we assessed infection-related causes of hospitalization or death in a population-based collection of 7,599 individuals from FINRISK 1997 who had baseline GlycA profiles and electronic health records with 13.8 years of follow-up (Supplemental Experimental Procedures). Cox proportional hazards models were fitted to each infectious disease cause, with adjustments for age, sex, triglycerides, and hospitalization due to the same diagnosis category in up to 10 years prior to the GlycA measurement.

We observed strong associations between baseline GlycA levels in FINRISK 1997 and increased risk of hospitalization and death from non-localized infections (hospitalization HR = 1.40 per SD of GlycA, $p = 2 \times 10^{-9}$, cases = 585; death HR = 2.36, $p = 8 \times 10^{-5}$, cases = 29) and respiratory infections (hospitalization hazard ratio [HR] = 1.48, $p = 3 \times 10^{-12}$, cases = 571; death HR = 1.39, $p = 9 \times 10^{-3}$, cases = 113) during the 13.8 years of follow-up (Figure 4). Individuals with GlycA levels above the population median were at more than 5-fold increased risk of death from non-localized infections in subsequent years ($p = 7 \times 10^{-3}$, cases = 29; Figure 4). The increased risk of death from non-localized infections was largely attributable to septicemia (HR = 2.25, $p = 4 \times 10^{-3}$, cases = 18), while death from respiratory infection was largely attributable to pneumonia (HR = 1.38, $p = 1 \times 10^{-2}$, cases = 113) (Figure 4B). GlycA was also broadly predictive of hospitalization from infection, including fungal infections and localized infections of the skin, bones, joints, intestinal, and urinary system (Figure 4B). Adjustment of GlycA models for prevalent CVD, diabetes, cancer, immunodeficiencies, and obesity made little difference to GlycA HRs for future risk of hospitalization and death from infection (Figure S3A). GlycA displayed stronger prediction on average than CRP across all infection-related events (Figure S3B), and adjustment of the models for CRP only slightly attenuated the GlycA HRs (Figure S3C). Taken together, these results suggest that apparently healthy individuals with elevated baseline levels of GlycA are at increased risk for severe infection events for up to 14 years in the future.

DISCUSSION

GlycA is a composite NMR-based signal related to changes in multiple circulating glycoproteins (Bell et al., 1987; Otvos et al., 2015), which are, themselves, responsive to myriad inflammatory stimuli (Gabay and Kushner, 1999). Previous studies have shown that GlycA is predictive of long-term risk for CVD events and all-cause mortality (Akinkuolie et al., 2014; Fischer et al., 2014). In this study, we characterized the molecular and cellular processes underlying GlycA variation in circulation. In apparently healthy individuals, we have shown that GlycA can be chronically elevated for periods of up to a decade; in individuals with high GlycA, widespread, modest elevation of numerous cytokines are suggestive of a prolonged low-grade inflammatory state. Accordingly, we identified a robust transcriptional module in whole blood that was strongly positively

associated with GlycA levels and enriched for innate immune response genes—specifically, those with neutrophil antimicrobial function. Elevated GlycA levels had long-term ramifications for risk of infection: high GlycA levels correlated with an increased risk of hospitalization and death from diverse infections—particularly septicemia and pneumonia—that persists for over a decade.

The question remains as to the cause of GlycA-associated chronic inflammation. Taken together, our results suggest multiple potential scenarios where apparently healthy individuals with high GlycA levels have (1) persistent but clinically silent, low-grade microbial infection(s), (2) a past severe infection event that results in modest chronic inflammation, (3) low-grade, chronic inflammation that mimics an anti-microbial response, or (4) a combination of these. This GlycA-related immune response was characterized by a gene coexpression network that harbored a clear neutrophil transcriptional signature indicative of antimicrobial function (Faurschou and Borregaard, 2003). Once recruited to the site of infection, neutrophils secrete antimicrobial peptides from their granules into the extracellular matrix and into phagosomes that surround microbes. They also form neutrophil extracellular traps (NETs): extracellular fibers composed of neutrophil DNA and granule antimicrobial peptides (Brinkmann et al., 2004; Urban et al., 2009). Neutrophil antimicrobial peptides are generally cytotoxic, damaging and degrading proximal tissue. Aberrantly increased neutrophil activity has been observed in a range of chronic inflammatory conditions that are generally thought to be independent of microbial infection, including rheumatoid arthritis (Branzk and Papayannopoulos, 2013; Wright et al., 2010), and increased neutrophil abundance itself has been shown to have predictive capacity for long-term mortality and myocardial infarction event risk (Horne et al., 2005).

Consistent with the diverse ramifications of neutrophil-driven inflammation, Warnatsch et al. recently reported a causal link between cholesterol, NET formation, and pro-inflammatory response driving atherosclerotic plaque growth (Warnatsch et al., 2015). Notably, two of the genes we identify as part of the neutrophil module, *ELANE* and *MPO*, have key roles in NET formation (Papayannopoulos et al., 2010). Furthermore, the long-term risk of severe infection events in apparently healthy individuals with elevated GlycA levels was consistent with overactivity of immune response, though GlycA may not necessarily play a causal role. Overactivity of immune response from severe infection events, such as sepsis and influenza, is known to cause tissue damage and organ dysfunction through hypercytokinemia (Tisoncik et al., 2012), and baseline elevation of circulating cytokines correlated with GlycA may contribute to its association with hospitalization and mortality risk. While our findings are largely observational, the prospect of systematic integration of prior medical interventions may yield causal relationships. In addition, an immediately fruitful avenue of research may be in vivo assays of neutrophil function, using neutrophils derived from individuals with variable GlycA levels.

This study demonstrates the utility of an approach that integrates diverse omics data and long-term follow-up of health records to uncover putative cellular-level processes associated with a recently uncovered biomarker. As multi-omics studies become increasingly common, biomarker associations with medical histories and health outcomes are likely to improve both patient care and our knowledge of the basic biological mechanisms underlying human disease. This study represents a potentially useful strategy for leveraging omics information for future studies in precision medicine.

EXPERIMENTAL PROCEDURES

Full experimental procedures can be found in the Supplemental Experimental Procedures. For YFS, ethics were approved by the Joint Commission on Ethics of the Turku University and the Turku University Central Hospital; for DILGOM, ethics were approved by the Coordinating Ethical Committee of the Helsinki and Uusimaa Hospital District; and for FINRISK, ethics were approved by the ethical committee of the National Public Health Institute, Finland (Supplemental Experimental Procedures).

Data Collection and Quantification

Concentrations of GlycA and triglycerides were quantified from serum samples using a proton NMR platform (Soininen et al., 2009). Concentrations of high-sensitivity CRP and the four glycoproteins were quantified from serum using immunoturbidimetric assays, and the 36 cytokines were quantified using multiplex cytokine panels. Genome-wide gene expression profiling was carried out on whole blood using the Illumina HT12 platform, and blood cell count analysis was carried out using flow cytometry. Incident event data were obtained from the Finnish National Hospital Discharge Register and the National Causes of Death Register (Lim et al., 2014). Clinical characteristics were also obtained from survey questions. Data type availability, sample sizes, and cohort characteristics are provided in Table 1.

Network Analysis

Network inference and module detection were performed in the DILGOM cohort using weighted gene coexpression network analysis (WGCNA) (Zhang and Horvath, 2005). Network inference was adjusted for age and sex. The soft-threshold power used for network construction was 5 (Figure S4A). Replication of the three GlycA-associated modules was assessed in the YFS2011 cohort through permutation testing of seven network-based statistics (Langfelder et al., 2011) (Table S6), using a pre-release of the NetRep R package (version 0.10.4, available at https://github.com/InouyeLab/NetRep/releases/tag/v0.10.4). For this purpose, the network inference component of WGCNA was performed on the YFS2011 gene expression. The soft threshold power was 4 (Figure S4b). Summary expression profiles were calculated as the eigenvector of the first principal component of each module's gene expression matrix. We use the shorthand "neutrophil module expression" to refer to the neutrophil module's summary expression profile throughout the text. Over-representation analysis of Gene Ontology (GO) terms and KEGG pathways was performed using the web tool, GeneCoDis (Carmona-Saez et al., 2007).

Statistical Analyses

A natural log-transform was applied to GlycA, each glycoprotein, triglycerides, and leukocyte abundance. A rank-based inverse normal transform was applied to each cytokine. Each measurement, including each module's summary expression, was standardized to SD units. All regression models were adjusted for age and sex, and models testing association between GlycA and module expression were additionally adjusted for triglycerides. HRs of GlycA for hospitalization or death were assessed by Cox proportional-hazard models using age as the timescale and adjusted for sex, triglycerides, and prior occurrence of the diagnosis category within ten years prior to sample collection. Sensitivity analyses were performed by (1) additionally adjusting for prevalent chronic diseases or inflammatory disorders, (2) replacing GlycA for CRP and, (3) adjusting GlycA for CRP.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2015.09.007.

ACKNOWLEDGMENTS

This study was supported by funding from National Health and Medical Research Council of Australia (NHMRC) grant APP1062227. M.I. was supported by an NHMRC and Australian Heart Foundation Career Development Fellowship (#1061435). S.R. was supported by an Australian Postgraduate Award and a PhD student top-up award from Victorian Life Sciences Computation Initiative (VLSCI). This study was further supported by the Strategic Research Funding from the University of Oulu, the Sigrid Juselius Foundation, the Academy of Finland (grants 141136), the Yrjö Jahnsson Foundation, the Emil Aaltonen Foundation, and the Finnish Diabetes Research Foundation. The Cardiovascular Risk in Young Finns Study is supported by the Finnish Foundation for Cardiovascular Research; Oulu, Helsinki, Kuopio, Tampere, and Turku University Central Hospital Medical Funds; the Paavo Nurmi Foundation; the Juho Vainio Foundation; the Finnish Cultural Foundation; and the Finnish Funding Agency for Technology and Innovation. M.J. was supported by the Paulo Foundation, Maud Kuistila Foundation, and Finnish Medical Foundation. The DILGOM Study and the National FINRISK Study are supported by the Academy of Finland (grants 139 and 635). The quantitative serum NMR metabolomics platform and its development has been supported by the Academy of Finland, TEKES-the Finnish Funding Agency for Technology and Innovation, the Sigrid Juselius Foundation, the Novo Nordisk Foundation, the Finnish Diabetes Research Foundation, the Paavo Nurmi Foundation, and the strategic and infrastructural research funding from the University of Oulu, as well as by the British Heart Foundation, the Wellcome Trust, and the Medical Research Council, UK. P.W., A.J.K., P.S., and M.A.-K. are shareholders of Brainshake Ltd., a company offering NMR-based metabolite profiling. P.W., A.J.K., P.S., and J.K. report employment relation for Brainshake Ltd.

Received: August 19, 2015
Revised: September 24, 2015
Accepted: September 28, 2015
Published: October 22, 2015

REFERENCES

Akinkuolie, A.O., Buring, J.E., Ridker, P.M., and Mora, S. (2014). A novel protein glycan biomarker and future cardiovascular disease events. J. Am. Heart Assoc. *3*, e001221.

Akinkuolie, A.O., Pradhan, A.D., Buring, J.E., Ridker, P.M., and Mora, S. (2015). Novel protein glycan side-chain biomarker and risk of incident type 2 diabetes mellitus. Arterioscler. Thromb. Vasc. Biol. *35*, 1544–1550.

Ala-Korpela, M. (2015). Serum nuclear magnetic resonance spectroscopy: one more step toward clinical utility. Clin. Chem. *61*, 681–683.

Aronsen, K.F., Ekelund, G., Kindmark, C.O., and Laurell, C.B. (1972). Sequential changes of plasma proteins after surgical trauma. Scand. J. Clin. Lab. Invest. Suppl. *124*, 127–136.

Bell, J.D., Brown, J.C., Nicholson, J.K., and Sadler, P.J. (1987). Assignment of resonances for 'acute-phase' glycoproteins in high resolution proton NMR spectra of human blood plasma. FEBS Lett. *215*, 311–315.

Borodulin, K., Vartiainen, E., Peltonen, M., Jousilahti, P., Juolevi, A., Laatikainen, T., Männistö, S., Salomaa, V., Sundvall, J., and Puska, P. (2015). Forty-year trends in cardiovascular risk factors in Finland. Eur. J. Public Health *25*, 539–546.

Branzk, N., and Papayannopoulos, V. (2013). Molecular mechanisms regulating NETosis in infection and disease. Semin. Immunopathol. *35*, 513–530.

Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y., and Zychlinsky, A. (2004). Neutrophil extracellular traps kill bacteria. Science *303*, 1532–1535.

Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J.M., and Pascual-Montano, A. (2007). GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol. *8*, R3.

Chen, Y., Zhu, J., Lum, P.Y., Yang, X., Pinto, S., MacNeil, D.J., Zhang, C., Lamb, J., Edwards, S., Sieberts, S.K., et al. (2008). Variations in DNA elucidate molecular networks that cause disease. Nature *452*, 429–435.

Chen, R., Mias, G.I., Li-Pook-Than, J., Jiang, L., Lam, H.Y.K., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E., et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell *148*, 1293–1307.

Ebersole, J.L., and Cappelli, D. (2000). Acute-phase reactants in infections and inflammatory diseases. Periodontol. 2000 *23*, 19–49.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. Nature *452*, 423–428.

Faurschou, M., and Borregaard, N. (2003). Neutrophil granules and secretory vesicles in inflammation. Microbes Infect. *5*, 1317–1327.

Fischer, K., Kettunen, J., Würtz, P., Haller, T., Havulinna, A.S., Kangas, A.J., Soininen, P., Esko, T., Tammesoo, M.L., Mägi, R., et al. (2014). Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. PLoS Med. *11*, e1001606.

Gabay, C., and Kushner, I. (1999). Acute-phase proteins and other systemic responses to inflammation. N. Engl. J. Med. *340*, 448–454.

Gitlin, J.D., and Colten, H.R. (1987). Molecular biology of the acute phase plasma proteins. Lymphokines *4*, 123–153.

Horne, B.D., Anderson, J.L., John, J.M., Weaver, A., Bair, T.L., Jensen, K.R., Renlund, D.G., and Muhlestein, J.B.; Intermountain Heart Collaborative Study Group (2005). Which white blood cell subtypes predict increased cardiovascular risk? J. Am. Coll. Cardiol. *45*, 1638–1643.

Inouye, M., Silander, K., Hamalainen, E., Salomaa, V., Harald, K., Jousilahti, P., Männistö, S., Eriksson, J.G., Saarela, J., Ripatti, S., et al. (2010a). An immune response network associated with blood lipid levels. PLoS Genet. *6*, e1001113.

Inouye, M., Kettunen, J., Soininen, P., Silander, K., Ripatti, S., Kumpula, L.S., Hämäläinen, E., Jousilahti, P., Kangas, A.J., Männistö, S., et al. (2010b). Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol. Syst. Biol. *6*, 441.

Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is my network module preserved and reproducible? PLoS Comput. Biol. *7*, e1001057.

Langfelder, P., Mischel, P.S., and Horvath, S. (2013). When is hub gene selection better than standard meta-analysis? PLoS ONE *8*, e61505.

Lauridsen, M.B., Bliddal, H., Christensen, R., Danneskiold-Samsøe, B., Bennett, R., Keun, H., Lindon, J.C., Nicholson, J.K., Dorff, M.H., Jaroszewski, J.W., et al. (2010). 1H NMR spectroscopy-based interventional metabolic phenotyping: a cohort study of rheumatoid arthritis patients. J. Proteome Res. *9*, 4545–4553.

Lim, E.T., Würtz, P., Havulinna, A.S., Palta, P., Tukiainen, T., Rehnström, K., Esko, T., Mägi, R., Inouye, M., Lappalainen, T., et al.; Sequencing Initiative Suomi (SISu) Project (2014). Distribution and medical impact of loss-of-function variants in the Finnish founder population. PLoS Genet. *10*, e1004494.

Montoya, D., Inkeles, M.S., Liu, P.T., Realegeno, S., Teles, R.M.B., Vaidya, P., Munoz, M.A., Schenk, M., Swindell, W.R., Chun, R., et al. (2014). IL-32 is a molecular marker of a host defense network in human tuberculosis. Sci. Transl. Med. *6*, 250ra114.

Morley, J.J., and Kushner, I. (1982). Serum C-reactive protein levels in disease. Ann. N Y Acad. Sci. *389*, 406–418.

Otvos, J.D., Shalaurova, I., Wolak-Dinsmore, J., Connelly, M.A., Mackey, R.H., Stein, J.H., and Tracy, R.P. (2015). GlycA: A composite nuclear magnetic resonance biomarker of systemic inflammation. Clin. Chem. *61*, 714–723.

Papayannopoulos, V., Metzler, K.D., Hakkim, A., and Zychlinsky, A. (2010). Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. J. Cell Biol. *191*, 677–691.

Pearson, T.A., Mensah, G.A., Alexander, R.W., Anderson, J.L., Cannon, R.O., 3rd, Criqui, M., Fadl, Y.Y., Fortmann, S.P., Hong, Y., Myers, G.L., et al.; Centers for Disease Control and Prevention; American Heart Association (2003). Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation *107*, 499–511.

Raitakari, O.T., Juonala, M., Rönnemaa, T., Keltikangas-Järvinen, L., Räsänen, L., Pietikäinen, M., Hutri-Kähönen, N., Taittonen, L., Jokinen, E., Marniemi, J., et al. (2008). Cohort profile: the cardiovascular risk in Young Finns Study. Int. J. Epidemiol. *37*, 1220–1226.

Soininen, P., Kangas, A.J., Würtz, P., Tukiainen, T., Tynkkynen, T., Laatikainen, R., Järvelin, M.-R., Kähönen, M., Lehtimäki, T., Viikari, J., et al. (2009). High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst (Lond.) *134*, 1781–1785.

Stanberry, L., Mias, G.I., Haynes, W., Higdon, R., Snyder, M., and Kolker, E. (2013). Integrative analysis of longitudinal metabolomics data from a personal multi-omics profile. Metabolites *3*, 741–760.

Summers, C., Rankin, S.M., Condliffe, A.M., Singh, N., Peters, A.M., and Chilvers, E.R. (2010). Neutrophil kinetics in health and disease. Trends Immunol. *31*, 318–324.

Theilgaard-Mönch, K., Jacobsen, L.C., Rasmussen, T., Niemann, C.U., Udby, L., Borup, R., Gharib, M., Arkwright, P.D., Gombart, A.F., Calafat, J., et al. (2005). Highly glycosylated alpha1-acid glycoprotein is synthesized in myelocytes, stored in secondary granules, and released by activated neutrophils. J. Leukoc. Biol. *78*, 462–470.

Theilgaard-Mönch, K., Jacobsen, L.C., Nielsen, M.J., Rasmussen, T., Udby, L., Gharib, M., Arkwright, P.D., Gombart, A.F., Calafat, J., Moestrup, S.K., et al. (2006). Haptoglobin is synthesized during granulocyte differentiation, stored in specific granules, and released by neutrophils in response to activation. Blood *108*, 353–361.

Tisoncik, J.R., Korth, M.J., Simmons, C.P., Farrar, J., Martin, T.R., and Katze, M.G. (2012). Into the eye of the cytokine storm. Microbiol. Mol. Biol. Rev. *76*, 16–32.

Urban, C.F., Ermert, D., Schmid, M., Abu-Abed, U., Goosmann, C., Nacken, W., Brinkmann, V., Jungblut, P.R., and Zychlinsky, A. (2009). Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against Candida albicans. PLoS Pathog. *5*, e1000639.

Wang, T.J., Larson, M.G., Vasan, R.S., Cheng, S., Rhee, E.P., McCabe, E., Lewis, G.D., Fox, C.S., Jacques, P.F., Fernandez, C., et al. (2011). Metabolite profiles and the risk of developing diabetes. Nat. Med. *17*, 448–453.

Warnatsch, A., Ioannou, M., Wang, Q., and Papayannopoulos, V. (2015). Inflammation. Neutrophil extracellular traps license macrophages for cytokine production in atherosclerosis. Science *349*, 316–320.

Wright, H.L., Moots, R.J., Bucknall, R.C., and Edwards, S.W. (2010). Neutrophil function in inflammation and inflammatory diseases. Rheumatology (Oxford) *49*, 1618–1631.

Würtz, P., Mäkinen, V.P., Soininen, P., Kangas, A.J., Tukiainen, T., Kettunen, J., Savolainen, M.J., Tammelin, T., Viikari, J.S., Rönnemaa, T., et al. (2012). Metabolic signatures of insulin resistance in 7,098 young adults. Diabetes *61*, 1372–1380.

Würtz, P., Wang, Q., Kangas, A.J., Richmond, R.C., Skarp, J., Tiainen, M., Tynkkynen, T., Soininen, P., Havulinna, A.S., Kaakinen, M., et al. (2014). Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. PLoS Med. *11*, e1001765.

Würtz, P., Havulinna, A.S., Soininen, P., Tynkkynen, T., Prieto-Merino, D., Tillin, T., Ghorbani, A., Artati, A., Wang, Q., Tiainen, M., et al. (2015). Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation *131*, 774–785.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article17.

Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell *153*, 707–720.