Multiomic blood correlates of genetic risk identify presymptomatic disease alterations

Michael Wainberg[a], Andrew T. Magis[a], John C. Earls[a], Jennifer C. Lovejoy[a], Nasa Sinnott-Armstrong[b], Gilbert S. Omenn[c], Leroy Hood[a,1], and Nathan D. Price[a,1]

[a]Institute for Systems Biology, Seattle, WA 98109; [b]Department of Genetics, Stanford University, Stanford, CA 94305; and [c]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109

Contributed by Leroy Hood, July 9, 2020 (sent for review March 9, 2020; reviewed by Ali Torkamani and Sheng Zhong)

Transitions from health to disease are characterized by dysregulation of biological networks under the influence of genetic and environmental factors, often over the course of years to decades before clinical symptoms appear. Understanding these dynamics has important implications for preventive medicine. However, progress has been hindered both by the difficulty of identifying individuals who will eventually go on to develop a particular disease and by the inaccessibility of most disease-relevant tissues in living individuals. Here we developed an alternative approach using polygenic risk scores (PRSs) based on genome-wide association studies (GWAS) for 54 diseases and complex traits coupled with multiomic profiling and found that these PRSs were associated with 766 detectable alterations in proteomic, metabolomic, and standard clinical laboratory measurements (clinical labs) from blood plasma across several thousand mostly healthy individuals. We recapitulated a variety of known relationships (e.g., glutamatergic neurotransmission and inflammation with depression, IL-33 with asthma) and found associations directly suggesting therapeutic strategies (e.g., Ω-6 supplementation and IL-13 inhibition for amyotrophic lateral sclerosis) and influences on longevity (leukemia inhibitory factor, ceramides). Analytes altered in high-genetic-risk individuals showed concordant changes in disease cases, supporting the notion that PRS-associated analytes represent presymptomatic disease alterations. Our results provide insights into the molecular pathophysiology of a range of traits and suggest avenues for the prevention of health-to-disease transitions.

polygenic risk scores | proteomics | metabolomics | presymptomatic disease

Inspired by a vision of medicine that is predictive, preventive, personalized, and participatory (P4) (1), our research group conducted a pilot 9-mo naturalistic study of 108 individuals combining whole-genome sequencing with longitudinal multiomic data collection (2–4) and targeted behavioral coaching (5). We furthered this vision by expanding data collection to 4,905 individuals, largely of self-identified white ancestry (3,752 of 4,789, 78.3%) and college-educated (2,543 of 2,812, 90.4%).

We reasoned that individuals at high genetic risk for a trait would show broad alterations in trait-related analytes (Fig. 1*A*), whether as causes or effects of biological changes leading up to the disease (6–10) or as noncausal proxies for other measured or unmeasured analytes (11, 12). To explore this hypothesis, we obtained genome-wide association studies (GWAS) results for 54 diseases/complex traits of broad interest (Fig. 1*B* and Dataset S1) from the GWAS catalog (13) and calculated polygenic risk scores (PRSs) (14) using a standard approach (Dataset S2): filtering variants based on their strength of association with the trait, then pruning to avoid high linkage disequilibrium between variants included in the PRS. We correlated all pairs of these PRSs across the individuals in our cohort and captured many expected disease relationships (*SI Appendix*, Fig. S1). We then performed outlier-robust correlations across self-identified white individuals (owing to the reduced applicability of PRSs across ancestries) between each of these 54 PRSs and 274 proteins (n = 2,114 individuals), 713 metabolites (n = 1,518), and 47 clinical laboratory tests (clinical labs) (n = 3,618), while correcting for age, sex, US state, weekday, month, season, technical factors, and outlier analyte levels (*Methods*).

A Catalog of Analyte–PRS Correlations

We obtained 219 protein–PRS, 259 metabolite–PRS, and 288 clinical lab–PRS correlations at a false discovery rate (FDR) of 10% (Fig. 2*A* and Dataset S3), for a total of 766 significant analyte–PRS correlations. Seventy-six percent of these correlations remained at least nominally significant (uncorrected $P < 0.05$) after excluding self-reported disease cases (where ascertained in our cohort; Dataset S4), suggesting significant correlations are largely (although not exclusively) driven by dysregulation among healthy or undiagnosed individuals. The body mass index (BMI) PRS was correlated with the most analytes (84 proteins, 114 metabolites, and 32 clinical labs), consistent with the large-scale dysregulation induced by obesity (15). The next two most associated PRSs were for brain-related traits: educational attainment and depression. The number of variants in a trait's PRS did not significantly correlate with its number of associated analytes (Spearman $\rho = 0.03$, $P = 0.8$; *SI Appendix*, Fig. S2).

On the whole, plasma metabolites and proteins were associated with far fewer PRSs than plasma clinical labs (Fig. 2 *B* and *C* and Dataset S5), reflecting decades of development of clinical analytes for broad use. The most associated protein and metabolite, the cytokine IL12B and the diglyceride oleoyl-linoleoyl-glycerol (18:1/18:2), had only 6 and 5 significant PRS associations, respectively, while the most associated clinical lab, total Ω-3 abundance, was

Significance

Most diseases do not develop suddenly, but rather involve the gradual malfunction of physiological processes due to genetic and environmental influences. Here, we study blood measurements from 4,905 individuals and find that people at elevated genetic risk for various diseases have hundreds of detectable changes in blood protein and metabolite levels as well as clinical laboratory tests. These changes tend to also be present in people with the disease, suggesting that they represent early signs of physiological dysfunction occurring before the onset of disease symptoms.

Author contributions: M.W., L.H., and N.D.P. designed research; M.W., A.T.M., J.C.E., J.C.L., N.S.-A., G.S.O., L.H., and N.D.P. performed research; A.T.M. and J.C.E. contributed new reagents/analytic tools; M.W. analyzed data; and M.W., L.H., and N.D.P. wrote the paper.

Reviewers: A.T., Scripps Research Institute; and S.Z., University of California San Diego.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: lhood@systemsbiology.org or nathan.price@systemsbiology.org.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2001429117/-/DCSupplemental.

First published August 19, 2020.

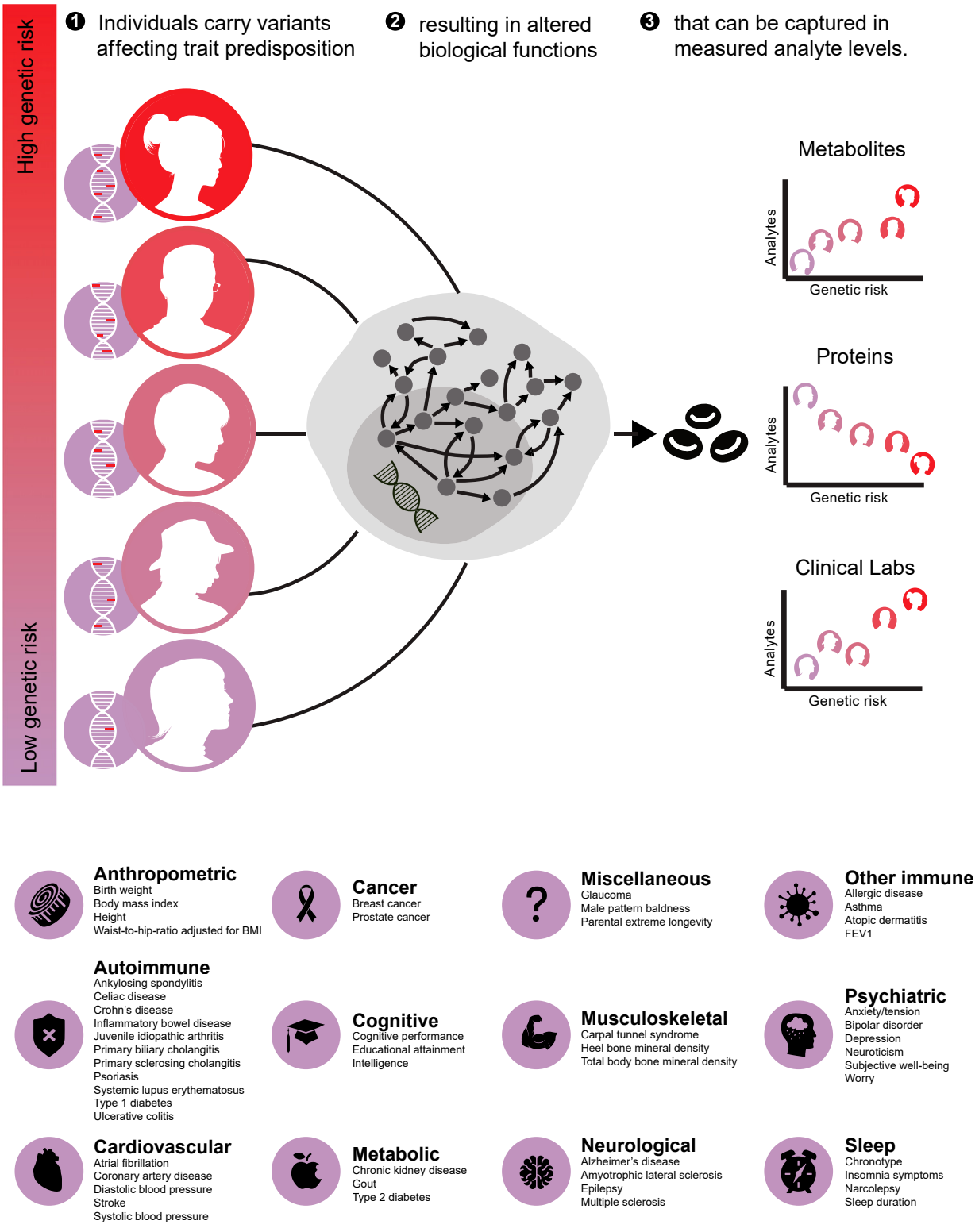

Fig. 1. Study overview. (*Top*) Conceptual overview. (*Bottom*) The 54 traits with polygenic risk scores.

correlated with 21 PRSs (14 positively and 7 negatively), nearly 40% of those tested. While 98% of clinical labs were significantly associated with at least 1 of the 54 PRSs, the same was true for only 50% of proteins and 27% of metabolites. Though potentially attributable in part to lesser technical variation among clinical labs or the larger number of individuals with clinical laboratory data, these differences suggest clinical labs tend to come closer to providing broad-brushed portraits of overall health (as one would expect), whereas plasma metabolites and proteins tend to only be associated with a small subset of traits.

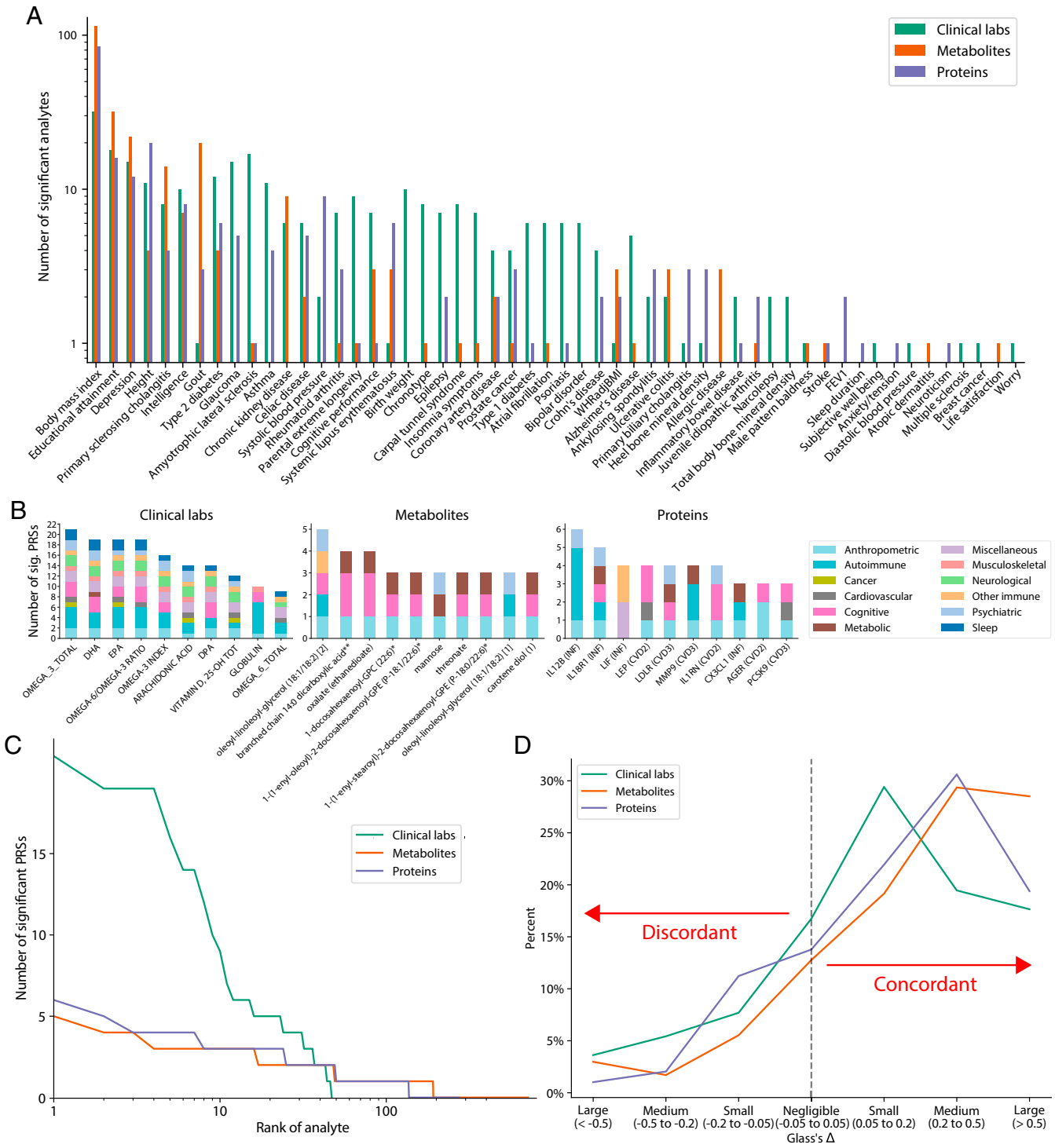

Fig. 2. Multiomic correlates of genetic risk. (*A*) Number of significant analytes associated with each PRS, stratified by analyte type. (*B*) Number of significant PRSs associated with each analyte, stratified by analyte type. For proteins, INF, CVD2, and CVD3 denote Olink panels. (*C*) Breakdown of PRS associations across trait categories for the 10 most associated analytes of each type. (*D*) Distribution of normalized effect sizes (Glass's Δ) of analyte abundances between disease cases and controls, for the 652 significant analyte–PRS correlations for diseases with case–control status ascertained in our cohort, stratified by analyte type. By convention, effect sizes are multiplied by the sign (+1 or −1) of the analyte–PRS correlation, so that positive effect sizes signify that analyte abundances are increased in both high-PRS individuals and disease cases, or decreased in both (concordant); negative effect sizes signify increased abundance in high-PRS individuals and decreased abundance in disease cases, or vice versa (discordant).

Crucially, analytes that positively correlated with genetic risk were highly likely to also be elevated in disease cases, and vice versa. For 558 (86%) of the 652 significant correlations for diseases with available case–control status data in our cohort (Dataset S4), the analyte showed at least small alterations (Glass's Δ > 0.05 in magnitude) in disease cases, and for 353

(54%) the analyte showed moderate or large alterations (Glass's $\Delta > 0.2$ in magnitude). Of these 353, 90% showed a concordant effect direction between PRS and disease cases (positively correlated with genetic risk and elevated in disease cases, or negatively correlated with genetic risk and diminished in disease cases), while only 10% were discordant (binomial $P = 2 \times 10^{-56}$; Fig. 2*D*). This broad concordance suggests that even individuals who have not yet transitioned to frank disease, or never will, still harbor disease-risk–associated signatures that are broadly detectable in the blood. The concept of PRS-associated analytes as presymptomatic disease alterations suggests avenues that could prevent high-risk individuals from transitioning to disease in the first place.

Our catalog of PRS–analyte associations includes many analyte–disease relationships with strong literature support (Table 1). These include the positive correlations of BMI with leptin (the strongest correlation across all traits and analytes), systemic inflammation (C-reactive protein, IL-6), and insulin resistance (insulin, HOMA-IR, HbA1C); chronic kidney disease with creatinine and potassium; coronary artery disease with low-density lipoprotein (LDL) cholesterol; depression with inflammation and excitatory neurotransmission; and type 2 diabetes with glucose, insulin, and HbA1C.

Several PRS–analyte associations in Table 1 point to known therapeutic strategies. The strongest proteomic association with coronary artery disease was PCSK9, the target of the LDL-lowering drugs alirocumab and evolocumab. The sole association of the cytokine IL-33 was a positive correlation with genetic risk for asthma. The *IL33* locus is a GWAS hit for asthma (29), and a loss-of-function splice variant in *IL33* is associated with halved asthma risk (17); multiple monoclonal antibodies targeting IL-33 are under development for asthma (30).

A particularly interesting class of associations consists of analytes known to be dysregulated among individuals prior to disease onset. Perhaps the best-known example is HbA1C, one measurement defining prediabetes and a strong risk factor for diabetes (31). A more subtle example is the metabolite 1-stearoyl-2-docosapentaenoyl-GPC, also known as phosphatidylcholine (18:0/22:5), which is negatively correlated with celiac disease polygenic risk. In a longitudinal cohort, plasma levels of phosphatidylcholines were lower by age 3 mo among children who eventually developed celiac disease at a median age of 4.8 y (23). These examples further support the notion that correlations with genetic risk may capture dysregulation that occurs prior to disease onset.

Noncanonical Disease–Analyte Associations

A large fraction of PRS–analyte associations are not canonically known; many are mechanistically plausible (Table 2). Several classes of these associations are of particular interest.

First, cognitive traits had a particular abundance of PRS–analyte associations. Intelligence, cognitive performance, and educational attainment were variously associated with total Ω-3s (as well as 10 esters of the Ω-3 fatty acid DHA, mainly phospholipids), leptin, platelet-derived growth factor A (PDGFA), PCSK9, the LDL receptor LDLR, IL-6, the adipocytokine PAI-1 (encoded by *SERPINE1*), and adenosine 5′-monophosphate (AMP). Notably, all correlations were negative (higher analyte levels were associated with impaired cognition), with the

Table 1. Selected literature-supported PRS–analyte associations

Trait	Analyte	Type	Sign	Notes
Ankylosing spondylitis	Globulin	P	+	Levels elevated in ankylosing spondylitis (16) and other autoimmune diseases
Asthma	IL-33	P	+	IL-33's only association; loss-of-function variant in *IL33* associated with halved asthma risk (17)
BMI	CRP, IL-6	L/P	+	Obesity associated with systemic inflammation and elevated CRP and IL-6 (18, 19)
	Insulin, HOMA-IR, HbA1C	L	+	Obesity associated with insulin resistance (20)
	Leptin	P	+	Strongest proteomic hit for BMI; obesity characterized by elevated leptin levels, leptin resistance (21)
	SERPINE1	P	+	a.k.a. PAI-1; adipocytokine overexpressed in obesity (22)
Celiac disease	1-stearoyl-2-docosapentaenoyl-GPC	M	–	a.k.a. phosphatidylcholine (18:0/22:5); in a longitudinal cohort, plasma phosphatidylcholine levels were lower by age 3 mo among children who eventually developed celiac disease (23)
Chronic kidney disease	eGFR	L	–	Calculated measure of kidney function based on age, sex, ethnicity, and creatinine levels
	Creatinine, potassium	L	+	Buildup in the blood when kidneys work less efficiently
Cognitive performance, EA, intelligence	Total Ω-3	L	+	Ω-3 supplementation may improve cognition in young adults, although evidence is mixed (24)
Coronary artery disease	LDL, small LDL, LDL particle #, PCSK9	L/P	+	PCSK9 (along with leptin, likely a proxy for BMI) is the sole proteomic association with CAD
Depression	Glutamate	M	+	Third strongest depression hit; serum levels are elevated in depression and other psychiatric disorders (25)
	IL-18, IL18R1	P/L	+	Serum levels of IL-18 (proinflammatory cytokine) are elevated in moderate-to-severe depression (26)
	GH1	P	–	61% of growth hormone-deficient adults had symptoms of atypical depression (27)
Male pattern baldness	androstenediol disulfate	M	+	Precursor to dihydrotestosterone (DHT), which stimulates hair loss in male pattern baldness
Prostate cancer	16α-hydroxy DHEA 3-sulfate	M	–	Precursor of estriol; estrogen therapy used to treat prostate cancer due to its antiandrogenic activity
Type 1 diabetes	IL2RB	P	+	IL-2 and its receptor play a central role in type 1 diabetes (28)
Type 2 diabetes	Glucose, HbA1C	L	+	The two strongest T2D associations

In all cases, directions (+/–) match the literature. Analyte types are denoted L (clinical labs), P (proteins), and M (metabolites). EA = educational attainment.

exception of those related to Ω-3s. Leptin, PDGFA, PCSK9/LDLR, CRP, and insulin, respectively, support the notions that obesity (42), hypertension (43), coronary artery disease (44), systemic inflammation (45), and insulin resistance (46) are associated with impaired cognition. [PDGFA is a platelet-derived growth factor elevated in mild hypertension (56).] Plasma levels of PAI-1 have been found to be higher in individuals with Alzheimer's and mild cognitive impairment (MCI)

Table 2. Selected noncanonical PRS–analyte associations

Trait	Analyte	Type	Sign	Notes
Allergic disease	Vanillylmandelate	M	+	Catecholamine metabolite; although epinephrine is famously used to treat anaphylaxis, basal norepinephrine levels are also several-fold increased in severe atopic eczema (32)
ALS	IL-13	P	+	IL-13+ helper and killer T cells are more abundant in ALS and positively correlate with progression (33)
	Total Ω-3, EPA, DHA	L	+	Ω-3 supplementation hastened and Ω-6 delayed
	Total Ω-6	L	–	neurodegeneration in an ALS mouse model (34)
Atrial fibrillation	Pyridoxal	M	–	Metabolite of pyroxidine (vitamin B6), previously found to be inversely associated with nonvalvular atrial fibrillation (35)
Celiac disease	CCL13	P	+	Involved in recruitment and activation of specific immune cell types to inflamed tissues; implicated in several other autoimmune and inflammatory diseases (36)
	CXCL10	P	+	Appears to be involved in immune cell recruitment into the small intestine in celiac disease (37)
Chronotype	4-hydroxyphenyl acetylglutamine	M	+	Precursor to dopamine and norepinephrine (via tyrosine), which have wakefulness-promoting effects
Cognitive performance	AMP (adenosine 5′-monophosphate)	M	–	Product of PDE-mediated degradation of cAMP; PDE inhibitors improve cognition in animal models (38)
Crohn's disease	TIMP4	P	–	Serum TIMP4 levels two thirds lower in Crohn's disease and ulcerative colitis patients than healthy controls (39)
Educational attainment	SERPINE1	P	–	Plasma levels elevated in Alzheimer's and MCI (40); loss-of-function variant associated with lifespan (41)
	LEP, PDGFA, PCSK9, LDLR, CRP, insulin	P/L	–	Obesity (42), hypertension (43), CAD (44), systemic inflammation (45), and insulin resistance (46) harm cognition; LEP and insulin (along with IL-6, another marker of systemic inflammation) also intelligence hits
	DHA esters	M	+	10 of 36 EA-associated metabolites are esters of the Ω-3 fatty acid DHA: 5 glycerophosphocholines (GPCs), 3 glycerophosphoethanolamines (GPEs), docosahexaenoylcholine, docosahexaenoylcarnitine
Height	Dihydrothymine	M	+	Precursor to the DNA base thymine
	Nerve growth factor	P	+	Though a growth factor, does not appear to have been linked to height in the literature
	NOTCH3	P	+	Loss-of-function variants in *NOTCH3* cause Lehman syndrome, which includes short stature and connective tissue abnormalities (47)
	Placental growth factor	P	+	Low maternal serum levels of which predict small for gestational age pregnancy (48)
Intelligence	IGFBP1	P	+	Strongest intelligence hit; deficits associated with insulin resistance, diabetes, cardiovascular disease; mouse overexpression improves insulin sensitivity, lowers blood pressure, reduces atherosclerosis (49)
Parental extreme longevity	LIF	P	–	Pleiotropic cytokine involved in inhibition of differentiation, oncogenesis, and HPA axis stimulation
	ceramide (d16:1/24:1, d18:1/22:1)	M	–	Deletion of the ceramide synthase LAG1 in yeast increases lifespan by 50% (50)
Primary sclerosing cholangitis	4-cholesten-3-one	M	–	Bile acid precursor (primary sclerosing cholangitis affects the bile ducts)
Rheumatoid arthritis	PI3	P	+	a.k.a. elafin; protease inhibitor that mediates the innate immune response; elevated in serum of RA patients (51)
Systolic blood pressure	CASP3, CASP8	P	+	Key apoptotic enzymes; hypertension associated with cardiac hypertrophy and apoptosis (52). Caspase-3 is also a hit for diastolic blood pressure
Systemic lupus erythematosus	MMP9	P	+	Strongest hit for lupus. Serum MMP-9 levels were elevated in lupus patients with neuropsychiatric manifestations, compared to other lupus patients and healthy controls (53)
Ulcerative colitis	trimethyllysine	M	–	Carnitine precursor; propionyl-l-carnitine is an effective treatment for ulcerative colitis (54)
	1-methylhistidine	M	–	Half as abundant in the urine of IBD patients compared to healthy controls (55)

Analyte types are abbreviated as in Table 1.

than healthy controls (40); perhaps relatedly, a loss-of-function variant in *SERPINE1* is associated with increased human lifespan (41). Among other roles, AMP is a degradation product of cAMP by phosphodiesterases (PDEs); PDE inhibitors improve cognition in animal models and have been proposed as procognitive therapeutics in humans (38).

Second, several analytes were associated with genetic risk for amyotrophic lateral sclerosis (ALS), a progressive and still largely mysterious neurodegenerative disease. Going against the common perception that Ω-3 fatty acids are "healthy" and Ω-6s are "unhealthy," we counterintuitively found that Ω-3 levels were elevated, and Ω-6 levels diminished, in individuals at high genetic risk for ALS. Supporting this association, Ω-6s delayed disease progression in a mouse model of ALS, while Ω-3 supplementation hastened progression (34), suggesting dietary modulation of Ω-3 and Ω-6 fatty acids may be a useful therapeutic strategy for ALS. [Conversely, an observational epidemiological study found an inverse correlation between Ω-3 intake and ALS risk, and a nonsignificant correlation with Ω-6 intake (57).] ALS genetic risk is also positively correlated with IL-13, its sole proteomic association: IL-13+ helper and killer T cells have been shown to be more abundant in ALS patients than controls and to positively correlate with progression (33). At least three anti–IL-13 monoclonal antibodies (anrukinzumab, lebrikizumab, and tralokinumab) are currently under development for asthma and atopic dermatitis (58); our results support further investigation of these antibodies as potential ALS therapies.

Third, exactly one protein, leukemia inhibitory factor (LIF), and one pair of metabolites sharing a mass spectrometry peak, ceramides d16:1/24:1 and d18:1/22:1, were associated with the PRS for parental extreme longevity (maternal age at death ≥98 y and/or paternal age at death ≥ 95 y), and both are negatively correlated. LIF is a pleiotropic cytokine broadly involved in the inhibition of cellular differentiation; LIF overexpression has variously been shown to promote the development and metastasis of solid tumors (opposite to its namesake effect on leukemia) (59), inhibit neurogenesis (60) and induce thymic atrophy (61). Serum LIF levels are elevated during both acute and chronic inflammation, and LIF injection stimulates the hypothalamic–pituitary–adrenal (HPA) axis, a possible mechanism by which LIF could modulate aging (62). Our results support clinical investigation of the anti-LIF monoclonal antibodies D25.1.4 and D62.3.2, which have over a 25-y track record of use in basic biology research (63), for aging-related diseases.

Ceramides, a class of sphingolipids, play a role in cellular senescence (64), and deletion of the ceramide synthase LAG1 in yeast increases longevity by 50% (50); therapeutic interventions targeting sphingolipid metabolism have been suggested for neurodegenerative disorders (65). Notably, all ceramides in our metabolomics panel had negative correlations with parental extreme longevity polygenic risk [P = 0.0003–0.1, ACAT (66) combined P = 0.001; Dataset S6], strongly suggesting ceramides as a class, not only ceramides d16:1/24:1 and/or d18:1/22:1, are associated with genetic predisposition to extreme longevity. While broad-spectrum ceramide synthase inhibitors such as fumonisins are associated with substantial toxicity in humans, inhibition of specific ceramide synthases (67) or downstream enzymes could represent viable therapeutic strategies.

Future Prospects

Our use of blood measurements in this study was motivated by greater accessibility compared to other disease-relevant tissues in living individuals (68). The use of blood as a proxy tissue is not as much of a limitation as it might first appear, since proteins from other tissues frequently make their way into the blood. For instance, even with a highly stringent filter of being specifically expressed in exactly one nonblood tissue (according to the Human Protein Atlas and the Genotype-Tissue Expression Project; *Methods*), we saw four examples among our PRS-associated proteins: AMBP (liver), FABP6 (small intestine), GH1 (pituitary), and NT-proBNP (heart muscle). As we have previously argued (69), blood is a surprisingly valuable window into health and disease.

Many expected PRS–analyte associations are missing, often due to lack of power. TGFβ, a canonical tumor suppressor which also contextually promotes cancer progression (70), had only a marginally significant negative correlation with the breast cancer PRS (P = 0.04, FDR = 46%) and a nonsignificant correlation with the PRS for prostate cancer, the other cancer tested. Although PCSK9 was associated with coronary artery disease (CAD), it was not associated with stroke (P = 0.3) despite the known protective effect of loss-of-function *PCSK9* variants on stroke (71). Allergic disease genetic risk had no significant proteomic associations, but among the 20 proteins with marginally significant (P < 0.05) associations were the immune-related proteins HSPB1 (P < 0.003), CEACAM8/CD66b (P < 0.004), BOC (P < 0.007), ACP5 (P < 0.009), CCL24 (P < 0.01), FCGR2B (P < 0.01), DLK1 (P < 0.01), MERTK (P < 0.02), EGFR (P < 0.02), TNSSF13B (P < 0.03), AXL (P < 0.03), XCL1 (P < 0.03), IL-4 (P < 0.04), MMP9 (P < 0.04), IL1R2 (P < 0.04), OSM (P < 0.04), and IL-6 (P < 0.05). Also missing were the vast majority of proteins and metabolites not among the few hundred profiled. Expanding the suite of analytes and individuals measured and broadening the cohort's ethnic and socioeconomic diversity to be more representative of the general population should dramatically increase the biological insights attainable by the analyte–PRS correlation approach outlined here.

Of course, not every disease–analyte correlation ought to appear as a PRS–analyte correlation. Indeed, this can distinguish true risk factors from mere consequences of disease. For instance, we would not expect troponins, a class of proteins (not included in our panel) released into the blood in response to heart injury, to be dysregulated among people at high genetic risk of coronary artery disease who have never had a heart attack. Similarly, although HbA1C is substantially higher in type 1 diabetes cases than controls (Glass's Δ = 2.8), it does not significantly correlate with polygenic risk for type 1 diabetes (P = 0.8), reflecting the autoimmune origin of the disease. Similarly, we might expect plasma levels of medications for a disease, or their downstream perturbations on the metabolome and proteome, to correlate with disease status but not polygenic risk. The subset of disease–analyte correlations that also correlate with polygenic risk are likely to be better therapeutic targets for prevention than analytes that are only dysregulated after the disease manifests.

That said, PRS–analyte correlations may still not represent bona fide causal influences of the analyte on the disease. As mentioned earlier, noncausal "proxy" analytes may correlate with causal analytes and thus also correlate with polygenic risk. One way this can happen is horizontal pleiotropy, whereby variants in the PRS causally influence the levels of multiple analytes, only some of which causally influence the trait. Bias in the PRSs themselves, for instance due to population structure or imperfect linkage between the true causal variants and the variants included in the PRS, could also theoretically lead to false-positive (or indeed false-negative) correlations.

In sum, our analysis of the multiomic correlates of genetic risk suggests that individuals at high genetic risk for a trait display dysregulation in many of the same analytes dysregulated in frank disease, as we would predict, and that this signature of dysregulation is frequently detectable in the blood. Our results underscore the concept that polygenic risk scores, far from being a mere statistical tool for disease risk stratification, also reflect underlying disease biology. Our ability to discover hundreds of significant disease–analyte associations from a cohort of only a few thousand people, most without severe disease, exemplifies

the utility of studying genetic risk in multiomics cohorts: since genetic risk scores can be computed for any genome, every person's data are potentially relevant to the study of the prodromal states of many diseases. Applying this approach to the next generation of population-scale multiomics studies offers the promise of discovering therapeutically relevant mechanistic insights across the spectrum of human health and disease.

Methods

Cohort and Dataset. The cohort used in this study consists of participants in the Arivale Scientific Wellness program (Arivale, Inc.) from 2015 to 2019 with self-reported white ancestry who gave informed consent for the use of their deidentified data for scientific research. Within this cohort, genetic data were ascertained through either whole-genome sequencing (n = 2,138) or Illumina MEG single nucleotide polymorphism (SNP) microarray genotyping (n = 1,500) from WuXi Nextcode; proteomic, metabolomic, and clinical laboratory measurements were ascertained as described in Wilmanski et al. (72). Unknown metabolites (denoted by IDs starting with "X -") were excluded from the analysis.

Polygenic Risk Score Generation. GWAS summary statistics for the 54 traits used in this study were downloaded from the GWAS Catalog (13); the traits, studies, and summary statistics URLs are provided as Dataset S1. Summary statistics were first filtered to the variants present in the imputed genomes from the cohort. Summary statistics were then harmonized with 1000 Genomes Phase 3 (73) with respect to reference/alternate allele and strand, using the allele harmonization framework from munge_sumstats.py in the *ldsc* software package (74); ambiguous variants (A/T, C/G, G/C, and T/A) were excluded. Summary statistics were then subset to $P < 1 \times 10^{-3}$, and P value-informed linkage disequilibrium (LD) pruning to $r^2 > 0.2$ in 1000 Genomes Phase 3 was then performed by using the "--indep-pairwise 500kb 0.2" flags to the *plink* v2.00 software package (75) and, to ensure the SNP with the higher P value is always the one pruned, specifying a dummy allele frequency file using the "--read-freq" flag with allele frequencies equal to 0.5 + p/2. The variants that remain, after this filtering and LD pruning, constitute the trait's polygenic risk score, with the variants' effect sizes (beta coefficients for quantitative traits or log odds ratios for case–control studies) constituting the weights of the risk score. Finally, polygenic risk scores were scored on each individual in the study cohort by summing up the published effect size for each variant in the PRS, multiplied by the number of effect alleles the individual carried for that variant; missing genotypes were mean imputed using the effect allele frequency.

Correlations with Polygenic Risk. Since genetic risk is non-time-varying, longitudinal analyte data were collapsed by taking the median across time points. Temporal covariates (weekday, month, season) were specified only for individuals where they agreed across all time points and otherwise set to missing. Analytes with more than 10% missingness across individuals and individuals with self-reported nonwhite ancestry (East Asian, Hispanic Latino or Spanish origin, South Asian, Ashkenazi Jewish, Black or African-American, Asian, Middle Eastern or North African, Native Hawaiian or other Pacific Islander, Sephardic Jewish, American Indian or Alaska Native, Afro-Caribbean, or Other) were prefiltered from the analysis. Missing analyte values were then imputed from nonmissing individuals using the average of the five nearest neighbors (*sklearn.impute.KNNImputer* in Python), and analytes where more than half of individuals had the same value after imputation were filtered out. Analytes, PRSs, and covariates were then shifted and scaled to have mean 0 and variance 1 across points in their interquartile range (*sklearn.preprocessing.RobustScaler* in Python). Association testing was then performed for each analyte–PRS pair via multilinear robust linear regression with M-estimation (*statsmodels.robust.robust_linear_model.RLM* in Python), with the analyte's abundance as the dependent variable and the PRS as the independent variable, including as covariates age at the first time point, sex, US state, weekday, month, season, whether the individual had whole-genome sequencing or genotyping array data, and the top eight genotype principal components across individuals. For each analyte type, only individuals with both DNA sequence data and abundances for that analyte type were included in the regression.

Glass's Δ Calculation. Glass's Δs were calculated by 1) regressing each analyte on the covariates mentioned above, across individuals, 2) subtracting the predictions from the measured analyte values, and 3) computing, on this residualized data, the difference in mean analyte abundances between cases and controls, divided by the sample SD in controls.

Tissue Specificity Analysis. Of the 136 proteins that correlate with at least 1 PRS, 15 were listed as "tissue-enriched" (at least fourfold higher mRNA level in a particular tissue than in any other tissue; www.proteinatlas.org/humanproteome/tissue/tissue+specific) according to the Human Protein Atlas (76): AGER, AMBP, AZU1, CCL19, FABP6, GH1, IL13, IL22RA1, IL4, LTA, MMP10, NT-proBNP (*NPPB*), PGF, SERPINA12, SLAMF1. Manual inspection of expression profiles in the Genotype-Tissue Expression Project (68) (e.g., https://gtexportal.org/home/gene/AGER) revealed that only seven appeared truly tissue-specific; of these, the adipokine SERPINA12 (skin) and cytokines IL13 (testis) and LTA (appendix) have known roles in adipose or immune cell types. This left the four tissue-specific proteins mentioned in the main text: AMBP (liver), FABP6 (small intestine), GH1 (pituitary), and NT-proBNP (heart muscle).

Data Availability. All study data are available upon request. URLs for the summary statistics in the GWAS catalog used to calculate the 54 polygenic risk scores are listed in Dataset S1.

ACKNOWLEDGMENTS. We gratefully acknowledge Allison Kudla for the graphic design of Fig. 1. We also thank Noa Rappaport, Tomasz Wilmanski, Sergey Kornilov, James Yurkovich, Shizhen Qin, and other members of the L.H.–N.D.P. laboratory and the Institute for Systems Biology for helpful discussions and Nik Schork for a critical reading of the paper. This research was supported by the National Institute on Aging through the Longevity Consortium (Grant 2U19AG023122-11A1 [to N.D.P.]).

1. L. Hood, S. H. Friend, Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat. Rev. Clin. Oncol.* **8**, 184–187 (2011).
2. B. B. Sun *et al.*, Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
3. S. M. Schüssler-Fiorenza Rose *et al.*, A longitudinal big data approach for precision health. *Nat. Med.* **25**, 792–804 (2019).
4. W. Zhou *et al.*, Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature* **569**, 663–671 (2019).
5. N. D. Price *et al.*, A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* **35**, 747–756 (2017).
6. P. Mahlknecht, K. Seppi, W. Poewe, The concept of prodromal Parkinson's disease. *J. Parkinsons Dis.* **5**, 681–697 (2015).
7. M. K. Larson, E. F. Walker, M. T. Compton, Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders. *Expert Rev. Neurother.* **10**, 1347–1359 (2010).
8. R. C. Peterson, Mild cognitive impairment: Transition between aging and Alzheimer's disease. *Neurologia* **15**, 93–101 (2000).
9. T. H. McGlashan, Commentary: Progress, issues, and implications of prodromal research: An inside view. *Schizophr. Bull.* **29**, 851–858 (2003).
10. R. A. Sperling *et al.*, Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 280–292 (2011).
11. R. Mayeux, Biomarkers: Potential uses and limitations. *NeuroRx* **1**, 182–188 (2004).
12. S. Burgess, C. N. Foley, V. Zuber, Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu. Rev. Genomics Hum. Genet.* **19**, 303–327 (2018).
13. A. Buniello *et al.*, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
14. A. Torkamani, N. E. Wineinger, E. J. Topol, The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
15. S. B. Heymsfield, T. A. Wadden, Mechanisms, pathophysiology, and management of obesity. *N. Engl. J. Med.* **376**, 1490–1492 (2017).
16. M. J. Kendall, M. Farr, M. J. Meynell, C. F. Hawkins, Synovial fluid in ankylosing spondylitis. *Ann. Rheum. Dis.* **32**, 487–492 (1973).
17. D. Smith *et al.*, A rare IL33 loss-of-function mutation reduces blood eosinophil counts and protects from asthma. *PLoS Genet.* **13**, e1006659 (2017).
18. K. Eder, N. Baffy, A. Falus, A. K. Fulop, The major inflammatory mediator interleukin-6 and obesity. *Inflamm. Res.* **58**, 727–736 (2009).
19. N. J. Timpson *et al.*, C-reactive protein levels and body mass index: Elucidating direction of causation through reciprocal Mendelian randomization. *Int. J. Obes.* **35**, 300–308 (2011).
20. J.-P. Després, A. Marette, "Obesity and insulin resistance" in *Contemporary Endocrinology: Insulin Resistance*, G. Reaven, A. Laws, Eds. (Humana Press Inc., 1999), pp. 51–81.
21. H. Pan, J. Guo, Z. Su, Advances in understanding the interrelations between leptin resistance and obesity. *Physiol. Behav.* **130**, 157–169 (2014).

22. M.-C. Alessi, I. Juhan-Vague, PAI-1 and the metabolic syndrome: Links, causes, and consequences. *Arterioscler. Thromb. Vasc. Biol.* **26**, 2200–2207 (2006).
23. P. Sen *et al*., Persistent alterations in plasma lipid profiles before introduction of gluten in the diet associated with progression to celiac disease. *Clin. Transl. Gastroenterol.* **10**, 1–10 (2019).
24. I. Bauer *et al*., Omega-3 supplementation improves cognition and modifies brain activation in young adults. *Hum. Psychopharmacol.* **29**, 133–144 (2014).
25. D. C. Mathews, I. D. Henter, C. A. Zarate, Targeting the glutamatergic system to treat major depressive disorder: Rationale and progress to date. *Drugs* **72**, 1313–1333 (2012).
26. R. A. Merendino *et al*., Interleukin-18 and CD30 serum levels in patients with moderate-severe depression. *Mediators Inflamm.* **11**, 265–267 (2002).
27. T. Mahajan, A. Crown, S. Checkley, A. Farmer, S. Lightman, Atypical depression in growth hormone deficient adults, and the beneficial effects of growth hormone treatment on depression and quality of life. *Eur. J. Endocrinol.* **151**, 325–332 (2004).
28. M. A. Hulme, C. H. Wasserfall, M. A. Atkinson, T. M. Brusko, Central role for interleukin-2 in type 1 diabetes. *Diabetes* **61**, 14–22 (2012).
29. J. A. A. Stockman, Large-scale, consortium-based genomewide association study of asthma. *Year Bk. Pediatr.* **2012**, 525–527 (2012).
30. H. Takatori, S. Makita, T. Ito, A. Matsuki, H. Nakajima, Regulatory mechanisms of IL-33-ST2-mediated allergic inflammation. *Front. Immunol.* **9**, 2004 (2018).
31. N. Bansal, Prediabetes diagnosis and treatment: A review. *World J. Diabetes* **6**, 296–303 (2015).
32. G. Ionescu, R. Kiehl, Plasma catecholamine levels in severe atopic eczema. *Allergy* **43**, 614–616 (1988).
33. N. Shi *et al*., Increased IL-13-producing T cells in ALS: Positive correlations with disease severity and progression rate. *J. Neuroimmunol.* **182**, 232–235 (2007).
34. E. F. Boumil, R. B. Vohnoutka, Y. Liu, S. Lee, T. B. Shea, Omega-3 hastens and omega-6 delays the progression of neuropathology in a murine model of familial ALS. *Open Neurol. J.* **11**, 84–91 (2017).
35. R. Marcucci *et al*., Hyperhomocysteinemia and vitamin B6 deficiency: New risk markers for nonvalvular atrial fibrillation? *Am. Heart J.* **148**, 456–461 (2004).
36. E. Mendez-Enriquez, E. A. García-Zepeda, The multiple faces of CCL13 in immunity and inflammation. *Inflammopharmacology* **21**, 397–406 (2013).
37. C. Bondar *et al*., Role of CXCR3/CXCL10 axis in immune cell recruitment into the small intestine in celiac disease. *PLoS One* **9**, e89068 (2014).
38. O. A. H. Reneerkens, K. Rutten, H. W. M. Steinbusch, A. Blokland, J. Prickaerts, Selective phosphodiesterase inhibitors: A promising target for cognition enhancement. *Psychopharmacology (Berl.)* **202**, 419–443 (2009).
39. A. N. Kapsoritakis *et al*., Imbalance of tissue inhibitors of metalloproteinases (TIMP) - 1 and - 4 serum levels, in patients with inflammatory bowel disease. *BMC Gastroenterol.* **8**, 55 (2008).
40. J. Oh, H.-J. Lee, J.-H. Song, S. I. Park, H. Kim, Plasminogen activator inhibitor-1 as an early potential diagnostic marker for Alzheimer's disease. *Exp. Gerontol.* **60**, 87–91 (2014).
41. S. S. Khan *et al*., A null mutation in SERPINE1 protects against biological aging in humans. *Sci. Adv.* **3**, eaao1617 (2017).
42. C. Wang, J. S. Y. Chan, L. Ren, J. H. Yan, Obesity reduces cognitive and motor functions across the lifespan. *Neural Plast.* **2016**, 2473081 (2016).
43. C. Iadecola *et al*., Impact of hypertension on cognitive function: A scientific statement from the American Heart Association. *Hypertension* **68**, e67–e94 (2016).
44. L. Zheng *et al*., Coronary artery disease is associated with cognitive decline independent of changes on magnetic resonance imaging in cognitively normal elderly adults. *J. Am. Geriatr. Soc.* **60**, 499–504 (2012).
45. C. B. Wright *et al*., Interleukin-6 is associated with cognitive function: The Northern Manhattan Study. *J. Stroke Cerebrovasc. Dis.* **15**, 34–38 (2006).
46. L. Ma, J. Wang, Y. Li, Insulin resistance and cognitive dysfunction. *Clin. Chim. Acta* **444**, 18–23 (2015).
47. K. W. Gripp *et al*., Truncating mutations in the last exon of NOTCH3 cause lateral meningocele syndrome. *Am. J. Med. Genet. A.* **167A**, 271–281 (2015).
48. L. C. Y. Poon, E. Zaragoza, R. Akolekar, E. Anagnostopoulos, K. H. Nicolaides, Maternal serum placental growth factor (PlGF) in small for gestational age pregnancy at 11(+0) to 13(+6) weeks of gestation. *Prenat. Diagn.* **28**, 1110–1115 (2008).
49. A. Rajwani *et al*., Increasing circulating IGFBP1 levels improves insulin sensitivity, promotes nitric oxide production, lowers blood pressure, and protects against atherosclerosis. *Diabetes* **61**, 915–924 (2012).
50. N. P. D'mello *et al*., Cloning and characterization of LAG1, a longevity-assurance gene in yeast. *J. Biol. Chem.* **269**, 15451–15459 (1994).
51. A. Olewicz-Gawlik *et al*., Serum alarm antiproteases in systemic sclerosis patients. *Hum. Immunol.* **78**, 559–564 (2017).
52. X. U. Deng *et al*., Reversion of left ventricle remodeling in spontaneously hypertensive rats by valsartan is associated with the inhibition of caspase-3, -8 and -9 activities. *Biomed. Rep.* **3**, 533–536 (2015).
53. H. Ainiala *et al*., Increased serum matrix metalloproteinase 9 levels in systemic lupus erythematosus patients with neuropsychiatric manifestations and brain magnetic resonance imaging abnormalities. *Arthritis Rheum.* **50**, 858–865 (2004).
54. M. G. Scioli *et al*., Propionyl-L-carnitine is efficacious in ulcerative colitis through its action on the immune function and microvasculature. *Clin. Transl. Gastroenterol.* **5**, e55 (2014).
55. N. S. Stephens *et al*., Urinary NMR metabolomic profiles discriminate inflammatory bowel disease from healthy. *J. Crohn's Colitis* **7**, e42–e48 (2013).
56. E. Rossi *et al*., Increased plasma levels of platelet-derived growth factor (PDGF-BB + PDGF-AB) in patients with never-treated mild essential hypertension. *Am. J. Hypertens.* **11**, 1239–1243 (1998).
57. K. C. Fitzgerald *et al*., Dietary ω-3 polyunsaturated fatty acid intake and risk for amyotrophic lateral sclerosis. *JAMA Neurol.* **71**, 1102–1110 (2014).
58. H. Li, K. Wang, H. Huang, W. Cheng, X. Liu, A meta-analysis of anti-interleukin-13 monoclonal antibodies for uncontrolled asthma. *PLoS One* **14**, e0211790 (2019).
59. X. Yue, L. Wu, W. Hu, The regulation of leukemia inhibitory factor. *Cancer Cell Microenviron.* **2**, e877 (2015).
60. S. Bauer, P. H. Patterson, Leukemia inhibitory factor promotes neural stem cell self-renewal in the adult brain. *J. Neurosci.* **26**, 12089–12099 (2006).
61. G. D. Sempowski *et al*., Leukemia inhibitory factor, oncostatin M, IL-6, and stem cell factor mRNA expression in human thymus increases with age and is associated with thymic atrophy. *J. Immunol.* **164**, 2180–2187 (2000).
62. D. Gupta, J. E. Morley, Hypothalamic-pituitary-adrenal (HPA) axis and aging. *Compr. Physiol.* **4**, 1495–1510 (2014).
63. K. J. Kim, M. Alphonso, C. H. Schmelzer, D. Lowe, Detection of human leukemia inhibitory factor by monoclonal antibody based ELISA. *J. Immunol. Methods* **156**, 9–17 (1992).
64. M. E. Venable, J. Y. Lee, M. J. Smyth, A. Bielawska, L. M. Obeid, Role of ceramide in cellular senescence. *J. Biol. Chem.* **270**, 30701–30708 (1995).
65. A. Di Pardo, V. Maglione, Sphingolipid metabolism: A new therapeutic opportunity for brain degenerative disorders. *Front. Neurosci.* **12**, 249 (2018).
66. Y. Liu *et al*., ACAT: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am. J. Hum. Genet.* **104**, 410–421 (2019).
67. S. Schiffmann *et al*., Inhibitors of specific ceramide synthases. *Biochimie* **94**, 558–565 (2012).
68. A. Battle *et al*., Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
69. J. T. Yurkovich, L. Hood, Blood is a window into health and disease. *Clin. Chem.* **65**, 1204–1206 (2019).
70. M. H. Barcellos-Hoff, R. J. Akhurst, Transforming growth factor-β in breast cancer: Too much, too late. *Breast Cancer Res.* **11**, 202 (2009).
71. A. S. Rao *et al*., Large-scale phenome-wide association study of *PCSK9* variants demonstrates protection against ischemic stroke. *Circ. Genom. Precis. Med.* **11**, e002162 (2018).
72. T. Wilmanski *et al*., Blood metabolome signature predicts gut microbiome α-diversity in health and disease. *Nat. Biotechnol.* **37**, 1217–1228 (2018).
73. A. Auton *et al*.; The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
74. B. K. Bulik-Sullivan *et al*.; Schizophrenia Working Group of the Psychiatric Genomics Consortium, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
75. C. C. Chang *et al*., Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
76. M. Uhlén *et al*., Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).