Co-regulatory networks of human serum proteins link genetics to disease

Valur Emilsson[1,2]*†, Marjan Ilkov[1]*, John R. Lamb[3]*†, Nancy Finkel[4], Elias F. Gudmundsson[1], Rebecca Pitts[4], Heather Hoover[4], Valborg Gudmundsdottir[1], Shane R. Horman[3], Thor Aspelund[1,5], Le Shu[6], Vladimir Trifonov[3], Sigurdur Sigurdsson[1], Andrei Manolescu[7], Jun Zhu[8], Örn Olafsson[1], Johanna Jakobsdottir[1], Scott A. Lesley[3], Jeremy To[3], Jia Zhang[3], Tamara B. Harris[9], Lenore J. Launer[9], Bin Zhang[8], Gudny Eiriksdottir[1], Xia Yang[6], Anthony P. Orth[3], Lori L. Jennings[4]‡, Vilmundur Gudnason[1,10]†‡

[1]Icelandic Heart Association, Holtasmari 1, IS-201 Kopavogur, Iceland. [2]Faculty of Pharmacology, University of Iceland, 101 Reykjavik, Iceland. [3]Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA. [4]Novartis Institutes for Biomedical Research, 22 Windsor Street, Cambridge, MA 02139, USA. [5]Centre of Public Health Sciences, University of Iceland, 101 Reykjavik, Iceland. [6]Department of Integrative Biology and Physiology, University of California, Los Angeles CA, USA. [7]School of Science and Engineering, Mentavegur 1, IS-101, Reykjavik University, 101 Reykjavik, Iceland. [8]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [9]Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, Bethesda, MD 20892-9205, USA. [10]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

*These authors contributed equally to this work.

†Corresponding author. Email: valur@hjarta.is (V.E.); v.gudnason@hjarta.is (V.G.); jlamb@gnf.org (J.R.L.)

‡These authors contributed equally to this work.

Proteins circulating in the blood are critical for age-related disease processes, however the serum proteome has remained largely unexplored. To this end, 4,137 proteins covering most predicted extracellular proteins, were measured in the serum of 5,457 Icelanders over 65 years of age. Pair-wise correlation between proteins as they varied across individuals, revealed 27 different network modules of serum proteins, many of which were associated with cardiovascular and metabolic disease states as well as overall survival. The protein modules were controlled by *cis* and *trans* acting genetic variants; which in many cases were also associated with complex disease. This revealed co-regulated groups of circulating proteins that incorporated regulatory control between tissues and demonstrated close relationships to past, current and future disease states.

Human serum contains a dynamic flux of proteins synthesized by tissues and cells of the body (*1*). The secretome is complex and likely involves 15% or more of all proteins (*2*). Secreted proteins and circulating blood cells mediate global homeostasis via inter-cell communication, immune responses, vascular and endothelial cell function, tissue remodeling, fluid exchange and nutrient assimilation (*3*). Defined functional roles for many individual proteins in circulation remains to be ascribed owing to our limited ability to monitor their production, accumulation, and distribution in both model systems and humans.

Hetero-chronic parabiosis experiments that surgically joined the circulation of young and old mice, showed a system-wide effect on the regenerative capacity of organs (*4*, *5*). Thus serum proteins and other circulating factors may directly regulate complex processes such as aging and the development of common chronic diseases. In contrast to monogenic diseases, complex diseases are not caused by proteins acting alone but instead by highly interacting protein networks that may result from genetic and environmental perturbations and ultimately drive physiological states toward disease (*6*–*9*). Since blood mediates coordination between non-adjacent tissues, it is of the highest interest to understand if and how this regulation occurs via serum proteins and their networks.

A custom designed aptamer-based multiplex proteomic platform

To date, high throughput detection and quantification of serum proteins in a large human population has been hampered by the limitations of proteomic profiling technologies. The Slow-Off rate Modified Aptamer (SOMAmer) based technology has emerged as a proteomic profiling platform with high sample throughput and sensitivity of detection (*10*, *11*). We designed an expanded custom version of this platform to include proteins known or predicted to be found in the extracellular milieu, including the predicted extracellular domains of single- and certain multi-pass transmembrane proteins. This resulted in an updated array of 5,034 SOMAmers, 4,783 of which recognize 4,137 individual human proteins (table S1), i.e., some proteins are targeted by more than one aptamer, while the rest recognize non-human targets. We applied the platform to a large population-based sampling of 5,457 subjects of the AGES Reykjavik study (*12*), a prospective

study of deeply phenotyped and genotyped subjects over 65 years of age. Table S2 reports baseline characteristics of the study population, while fig. S1 shows the workflow of the present study. For direct and inferential measures of aptamer specificity see tables S3 to S6 and figs. S2 and S3.

Identification and characterization of serum protein networks

Biological networks are characterized by a non-random distribution of links between objects that are scale-free in nature (*13*). We reconstructed the protein co-regulation network using weighted gene-to-gene co-expression analysis (*14*). Co-regulation networks describe functional relationships that can reflect both physical and non-physical interactions between objects including proteins. The parameters used for constructing the serum protein network, were chosen on the basis of the topological scale-free criterion (*14*), and tuned to enhance signal-to-noise in the protein adjacency matrix (Fig. 1A and fig. S4A, B). This analysis established that serum proteins cluster into 27 highly structured co-regulatory modules ranging in size between 20 and 921 proteins (table S7), while 15% fell outside of these modules. We examined the preservation of the network architecture by randomly splitting the AGES cohort into training and test sets, and applying a suite of statistics as previously described (*15*). We observed strong preservation of the overall structure of the network (Fig. 1B). Permutation testing of the data indicated that these modules were unlikely to have occurred by chance (fig. S4C). Pairwise correlation of proteins could be controlled at any of the many steps from transcription to secretion and clearance, and here we use the term co-regulation to encompass all such regulation between proteins.

We applied various annotation tools, and found the modules were enriched for distinct functional and tissue specific signatures (table S8). This suggests that, to some degree, peripheral proteins cluster according to their function and/or tissue of origin. Note that individual modules frequently contained proteins from many distinct tissues, indicating that individual tissues were not the sole contributor to any module (tables S9 and S10). In fact, comparison of the serum protein modules to 2,672 co-expression mRNA modules constructed from solid tissues (*16*), revealed some but largely an insignificant agreement between the two (fig. S5), suggesting that serum protein co-regulation is distinct from most tissues. Thus the human serum proteome appeared as functionally distinct modules of proteins produced by many tissues of the body.

We characterized each module´s eigenprotein (E$^{(q)}$, where q denotes a module) through a singular value decomposition and transformation of the variable protein levels for any given module. Each E$^{(q)}$ is a unique representation that most closely reflects the collective behavior of that module, explaining on average 40% of the protein covariance. Some modules E$^{(q)}$s were more correlated to others and formed several super-clusters of modules with shared functional categories (table S11 and fig. S6). We assessed if the modules were related to disease status of the AGES donors. Association of module E$^{(q)}$s to various outcomes including coronary heart disease (CHD), heart failure (HF), type-2-diabetes (T2D), visceral and subcutaneous adipose tissue (VAT and SAT, respectively) and metabolic syndrome (MetS), were found (table S12). While some modules showed no association to any disease measure, others were associated to some or all of them (table S12, Fig. 2A to K and fig. S7A to I). Notably, super-clusters, while internally consistent, often showed opposing relationships between super-clusters to CHD related outcomes (table S12, Fig. 2E, H, K and fig. S7B, E, H), which may reflect differential roles of the protein modules in disease. The two modules PM1 and PM23 associated with T2D were also associated with MetS, VAT, SAT and CHD related outcomes in a directionally consistent manner (table S12 and Fig. 2B, K). Finally, modules associated with incident disease and to all-cause or post CHD mortality (table S12, Fig. 2C, F, I and fig. S7C, F, I), indicate that the protein network predicted future events and disease outcome. Note that time between sample collection and diagnosis had no effect on the association of individual proteins to prevalent disease (fig. S8). Thus, individual disease associations of the protein modules link the shared functions of many proteins to common diseases.

The networks were scale-free in nature where a few protein nodes were highly connected. These are often referred to as network hubs which organize network connectivity and information flow (*17*). Studies of gene networks from solid tissues suggest that hub nodes tend to be essential and evolutionarily conserved (*18*). The scaled intra-module connectivity $\left(K_i = \dfrac{k_i}{k_{max}} \right)$ of proteins was compared to various outcomes, where protein hubs showed stronger association to phenotypic measures than less well connected proteins (Fig. 3A to L and fig. S9A to I). We observed strong reproducibility of the hub status of proteins through the network preservation analysis (fig. S10). In summary, the hub proteins consistently showed stronger association to various disease related outcomes compared to proteins located in the periphery of the network. Thus the structure of the protein network derived solely from measured protein variation, was aligned with disease and phenotype variation across donors.

Genetic variants influence serum protein levels

Genome-Wide Association Studies (GWAS) have identified thousands of common DNA sequence variants affecting human diseases (*19*). By integrating GWAS signals with genetic variants that affect intermediate traits like mRNA and/or protein, the identification of the causal candidates and

pathways can be enhanced (*6–9*). Using Bonferroni correction, we identified 1,046 significant *cis* pSNP-protein associations within a 300kb window across the corresponding protein coding sequences (table S13 and fig. S11A). Cross-referencing *cis* pSNP-proteins to *cis* eSNPs-transcripts data from over 30 tissues/cells (*20*), revealed an overlap of 37.3% (table S14). This suggests that ~60% of the genetic effects on serum protein levels are either mediated by as yet unknown transcriptional effect and/or post-transcriptional mechanisms. The *cis* pSNP-proteins were under-represented among highly connected protein nodes (fig. S11B, C), consistent with the observation that there was reduced selective pressure on non-hub proteins (*21*). We observed examples of *cis* protein effects likely underlying reported GWAS risk loci for T2D, adiposity and/or CHD (fig. S12A and tables S15 and S16).

We also found various lead GWAS SNPs mediate distal *trans* effects on serum proteins (figs. S12B to D and S13A to E, and table S16). Many of these risk loci influenced proteins in both *cis* and *trans* including the risk locus rs579459 for CHD upstream of *ABO* (fig. S12D and table S16), of which many were over-represented in the PM27 module (fig. S12E). Genetic pleiotropy is a well-established phenomenon at the *ABO* locus (fig. S13G and (*22*)). In fact, the ABO blood groups have been linked to many diseases which is in part mediated via their effect on the pro-thrombotic risk factor vWF (*23, 24*), a protein also affected by rs579459 in *trans* (table S16). In light of these results, all *cis*-acting pSNPs were tested directly for *trans* effects on proteins, revealing 16% of all *cis* pSNPs affected in total 911 proteins in *trans* (table S17). Notably, twice as many, or 40.7% of *cis* pSNPs affecting proteins in *trans* matched reported GWAS loci compared to 20.7% among all *cis* pSNPs (tables S15 and S17), suggesting a link between *trans* protein regulation and disease variation. Finally, we confirmed on average 80% of previously reported *cis* and 74% of *trans* effects on plasma proteins in our dataset (table S18 and fig. S14). The finding that GWAS risk loci tend to regulate many proteins in *trans* and that these proteins cluster in the same module, raised the possibility that common DNA sequence variants determine the architecture of the serum protein network.

Linking genetic variants to serum protein networks
Concerns have been expressed regarding the adequacy of GWAS to identify useful disease links, resulting in a call for greater efforts in identifying the gene regulatory networks that integrate the numerous GWAS signals (*25*). Given the hierarchical organization of the serum protein network and our observation that *trans* co-regulated proteins cluster in modules, it is possible that common *cis* and *trans* acting pSNPs underlie the structure of the serum protein network. We adopted a standard single-point GWAS test of linear regression against all module $E^{(q)}$s using a threshold of $P \leq 5 \times 10^{-8}$

for detection of genome-wide significant effects. Even though the $E^{(q)}$ represented many proteins as a complex trait, we often observed strong single SNP associations to the modules $E^{(q)}$s (table S19), which we have dubbed network-associated protein SNPs (npSNPs), of which many were associated with more than one module consistent with their relationship. Given npSNPs were associated with $E^{(q)}$s of distinct protein modules, npSNPs may act via individual proteins in circulation (tables S19 and S20). For instance, the npSNP rs704 (NP_000629.3: p.Thr400Met) in *VTN* affected VTN in *cis* (Fig. 4A to C and fig. S13H), and 698 proteins in *trans* that were significant components of the modules enriched for immune related pathways (Fig. 4D, E and table S20). Furthermore, distinct loci at *APOE* or *BCHE* were associated with the lipoprotein enriched module PM11 (Fig. 5A and fig. S12I and tables S8, S19 and S20), exerting *cis* effects on APOE and BCHE (Fig. 5B), and affecting 64 proteins in *trans* or 89% of all proteins in PM11 (Fig. 5C, D and tables S19 and S20). These results highlight the genetic architecture of the serum protein network and show that the modules and disease variation are intimately connected.

The discoveries of npSNPs and *cis-trans* protein pairs allowed us to assess if the serum protein networks resulted from cross-tissue regulatory control by comparing them to gene expression data from 53 different human tissues and involving only the top tissue specific proteins (table S21 and S22). Many *cis* acting pSNPs affected serum levels of tissue specific proteins and that subsequently affected variable serum levels of other proteins synthesized in distinct tissues (table S21 and S22). These data suggest that the serum protein network arose at least in part via systemic cross-tissue regulation.

Discussion
Deep protein profiling in a large well characterized population of the elderly, revealed the higher-order topology and modularity of the human serum proteome. The study cohort used was population-based thus containing non-diseased donors and donors suffering from various diseases and complications (*12*). As such the relationship between proteins is the consensus of the whole population. Recent studies have revealed a dramatic effect of aging on transcriptional changes that are both tissue and organismal specific (*26–28*). It is possible that the observed network structure of the serum proteins reflects to some extent those changes associated with old age of the population and may indeed reflect a certain set-point of homeostasis or even a decline in adaptive homeostasis in comparison to younger populations. Note that, for all the analyses of the present study we adjusted for the confounder age.

Structural features of the serum protein network resembled that of regulatory networks constructed in solid tissues

(*7, 8*), including association of hub proteins to diseases. However, comparison of *cis-trans* protein pairs and cognate protein modules to gene expression data and co-expression networks derived from solid tissues, suggests that the serum protein network arose in part via systemic cross-tissue regulation. We note that a follow-up replication testing of the many findings from these primary data, including the strong connection of protein hubs to disease traits, is warranted.

We anticipate that additional npSNPs can be identified with this technique, comparative to those found with conventional GWAS analyses of common diseases (*25*). Given that the effect of established GWAS loci is more complex than previously anticipated, this underscores the role of protein networks as the sensors and integrators of complex disease. The strong association of individual proteins and networks to disease states observed here indicates that the serum proteome may be a rich and accessible setting to mine for biomarkers of disease and disease responses to integrate information from tissues in a global regulatory network. As such, coordinated variance of serum proteins may offer unrecognized opportunities for target and biomarker identification in human disease.

REFERENCES AND NOTES

1. J. M. Schwenk, G. S. Omenn, Z. Sun, D. S. Campbell, M. S. Baker, C. M. Overall, R. Aebersold, R. L. Moritz, E. W. Deutsch, The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res.* **16**, 4299–4310 (2017). doi:10.1021/acs.jproteome.7b00467 Medline

2. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). doi:10.1126/science.1260419 Medline

3. M. Stastna, J. E. Van Eyk, Secreted proteins as a fundamental source for biomarker discovery. *Proteomics* **12**, 722–735 (2012). doi:10.1002/pmic.201100346 Medline

4. I. M. Conboy, M. J. Conboy, A. J. Wagers, E. R. Girma, I. L. Weissman, T. A. Rando, Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433**, 760–764 (2005). doi:10.1038/nature03260 Medline

5. S. A. Villeda, J. Luo, K. I. Mosher, B. Zou, M. Britschgi, G. Bieri, T. M. Stan, N. Fainberg, Z. Ding, A. Eggel, K. M. Lucin, E. Czirr, J.-S. Park, S. Couillard-Després, L. Aigner, G. Li, E. R. Peskind, J. A. Kaye, J. F. Quinn, D. R. Galasko, X. S. Xie, T. A. Rando, T. Wyss-Coray, The ageing systemic milieu negatively regulates neurogenesis and cognitive function. *Nature* **477**, 90–94 (2011). doi:10.1038/nature10357 Medline

6. E. E. Schadt, Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009). doi:10.1038/nature08454 Medline

7. B. Zhang, C. Gaiteri, L.-G. Bodea, Z. Wang, J. McElwee, A. A. Podtelezhnikov, C. Zhang, T. Xie, L. Tran, R. Dobrin, E. Fluder, B. Clurman, S. Melquist, M. Narayanan, C. Suver, H. Shah, M. Mahajan, T. Gillis, J. Mysore, M. E. MacDonald, J. R. Lamb, D. A. Bennett, C. Molony, D. J. Stone, V. Gudnason, A. J. Myers, E. E. Schadt, H. Neumann, J. Zhu, V. Emilsson, Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013). doi:10.1016/j.cell.2013.03.030 Medline

8. V. Emilsson, G. Thorleifsson, B. Zhang, A. S. Leonardson, F. Zink, J. Zhu, S. Carlson, A. Helgason, G. B. Walters, S. Gunnarsdottir, M. Mouy, V. Steinthorsdottir, G. H. Eiriksdottir, G. Bjornsdottir, I. Reynisdottir, D. Gudbjartsson, A. Helgadottir, A. Jonasdottir, A. Jonasdottir, U. Styrkarsdottir, S. Gretarsdottir, K. P. Magnusson, H. Stefansson, R. Fossdal, K. Kristjansson, H. G. Gislason, T. Stefansson, B. G. Leifsson, U. Thorsteinsdottir, J. R. Lamb, J. R. Gulcher, M. L. Reitman, A. Kong, E. E. Schadt, K. Stefansson, Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008). doi:10.1038/nature06758 Medline

9. Y. Chen, J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellini, S. Wang, M.-F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis, E. E. Schadt, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008). doi:10.1038/nature06757 Medline

10. D. R. Davies, A. D. Gelinas, C. Zhang, J. C. Rohloff, J. D. Carter, D. O'Connell, S. M. Waugh, S. K. Wolk, W. S. Mayfield, A. B. Burgin, T. E. Edwards, L. J. Stewart, L. Gold, N. Janjic, T. C. Jarvis, Unique motifs and hydrophobic interactions shape the binding of modified DNA ligands to protein targets. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19971–19976 (2012). doi:10.1073/pnas.1213933109 Medline

11. L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T. Fitzwater, D. Flather, A. Forbes, T. Foreman, C. Fowler, B. Gawande, M. Goss, M. Gunn, S. Gupta, D. Halladay, J. Heil, J. Heilig, B. Hicke, G. Husar, N. Janjic, T. Jarvis, S. Jennings, E. Katilius, T. R. Keeney, N. Kim, T. H. Koch, S. Kraemer, L. Kroiss, N. Le, D. Levine, W. Lindsey, B. Lollo, W. Mayfield, M. Mehan, R. Mehler, S. K. Nelson, M. Nelson, D. Nieuwlandt, M. Nikrad, U. Ochsner, R. M. Ostroff, M. Otis, T. Parker, S. Pietrasiewicz, D. I. Resnicow, J. Rohloff, G. Sanders, S. Sattin, D. Schneider, B. Singer, M. Stanton, A. Sterkel, A. Stewart, S. Stratford, J. D. Vaught, M. Vrkljan, J. J. Walker, M. Watrobka, S. Waugh, A. Weiss, S. K. Wilcox, A. Wolfson, S. K. Wolk, C. Zhang, D. Zichi, Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLOS ONE* **5**, e15004 (2010). doi:10.1371/journal.pone.0015004 Medline

12. T. B. Harris, L. J. Launer, G. Eiriksdottir, O. Kjartansson, P. V. Jonsson, G. Sigurdsson, G. Thorgeirsson, T. Aspelund, M. E. Garcia, M. F. Cotch, H. J. Hoffman, V. Gudnason, Age, Gene/Environment Susceptibility-Reykjavik Study: Multidisciplinary applied phenomics. *Am. J. Epidemiol.* **165**, 1076–1087 (2007). doi:10.1093/aje/kwk115 Medline

13. A. L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999). doi:10.1126/science.286.5439.509 Medline

14. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, e17 (2005). doi:10.2202/1544-6115.1128 Medline

15. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and reproducible? *PLOS Comput. Biol.* **7**, e1001057 (2011). doi:10.1371/journal.pcbi.1001057 Medline

16. L. Shu, K. H. K. Chan, G. Zhang, T. Huan, Z. Kurt, Y. Zhao, V. Codoni, D.-A. Trégouët, J. Yang, J. G. Wilson, X. Luo, D. Levy, A. J. Lusis, S. Liu, X. Yang; Cardiogenics Consortium, Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLOS Genet.* **13**, e1007040 (2017). doi:10.1371/journal.pgen.1007040 Medline

17. H. Jeong, S. P. Mason, A. L. Barabási, Z. N. Oltvai, Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001). doi:10.1038/35075138 Medline

18. A. L. Barabási, N. Gulbahce, J. Loscalzo, Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011). doi:10.1038/nrg2918 Medline

19. M. Muñoz, R. Pong-Wong, O. Canela-Xandri, K. Rawlik, C. S. Haley, A. Tenesa, Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat. Genet.* **48**, 980–983 (2016). Medline

20. J. R. Staley, J. Blackshaw, M. A. Kamat, S. Ellis, P. Surendran, B. B. Sun, D. S. Paul, D. Freitag, S. Burgess, J. Danesh, R. Young, A. S. Butterworth, PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016). doi:10.1093/bioinformatics/btw373 Medline

21. N. Mähler, J. Wang, B. K. Terebieniec, P. K. Ingvarsson, N. R. Street, T. R. Hvidsten, Gene co-expression network connectivity is an important determinant of selective constraint. *PLOS Genet.* **13**, e1006402 (2017). doi:10.1371/journal.pgen.1006402 Medline

22. J. K. Pickrell, T. Berisa, J. Z. Liu, L. Ségurel, J. Y. Tung, D. A. Hinds, Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016). doi:10.1038/ng.3570 Medline

23. M. Franchini, G. Lippi, The intriguing relationship between the ABO blood group, cardiovascular disease, and cancer. *BMC Med.* **13**, 7 (2015). [doi:10.1186/s12916-014-0250-y](doi:10.1186/s12916-014-0250-y) Medline

24. M. Franchini, F. Capra, G. Targher, M. Montagnana, G. Lippi, Relationship between ABO blood group and von Willebrand factor levels: From biology to clinical implications. *Thromb. J.* **5**, 14 (2007). [doi:10.1186/1477-9560-5-14](doi:10.1186/1477-9560-5-14) Medline

25. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017). doi:10.1016/j.cell.2017.05.038 Medline

26. D. Alfego, U. Rodeck, A. Kriete, Global mapping of transcription factor motifs in human aging. *PLOS ONE* **13**, e0190457 (2018). doi:10.1371/journal.pone.0190457 Medline

27. J. Yang, T. Huang, F. Petralia, Q. Long, B. Zhang, C. Argmann, Y. Zhao, C. V. Mobbs, E. E. Schadt, J. Zhu, Z. Tu; GTEx Consortium, Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci. Rep.* **5**, 15145 (2015). doi:10.1038/srep15145 Medline

28. J. M. Zahn, S. Poosala, A. B. Owen, D. K. Ingram, A. Lustig, A. Carter, A. T. Weeraratna, D. D. Taub, M. Gorospe, K. Mazan-Mamczarz, E. G. Lakatta, K. R. Boheler, X. Xu, M. P. Mattson, G. Falco, M. S. H. Ko, D. Schlessinger, J. Firman, S. K. Kummerfeld, W. H. Wood 3rd, A. B. Zonderman, S. K. Kim, K. G. Becker, AGEMAP: A gene expression database for aging in mice. *PLOS Genet.* **3**, e201 (2007). doi:10.1371/journal.pgen.0030201 Medline

29. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008). doi:10.1093/bioinformatics/btm563 Medline

30. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal. Complex Syst.* **1695**, 1 (2006).

31. American Diabetes Association, Diagnosis and classification of diabetes mellitus. *Diabetes Care* **36** (suppl. 1), S67–S74 (2013). [doi:10.2337/dc13-S067](doi:10.2337/dc13-S067) Medline

32. A. Agarwala, S. Virani, D. Couper, L. Chambless, E. Boerwinkle, B. C. Astor, R. C. Hoogeveen, J. Coresh, A. R. Sharrett, A. R. Folsom, T. Mosley, C. M. Ballantyne, V. Nambi, Biomarkers and degree of atherosclerosis are independently associated with incident atherosclerotic cardiovascular disease in a primary prevention cohort: The ARIC study. *Atherosclerosis* **253**, 156–163 (2016). doi:10.1016/j.atherosclerosis.2016.08.028 Medline

33. Y. Hathout, E. Brody, P. R. Clemens, L. Cripe, R. K. DeLisle, P. Furlong, H. Gordish-Dressman, L. Hache, E. Henricson, E. P. Hoffman, Y. M. Kobayashi, A. Lorts, J. K. Mah, C. McDonald, B. Mehler, S. Nelson, M. Nikrad, B. Singer, F. Steele, D. Sterling, H. L. Sweeney, S. Williams, L. Gold, Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7153–7158 (2015). doi:10.1073/pnas.1507719112 Medline

34. J. Candia, F. Cheung, Y. Kotliarov, G. Fantoni, B. Sellers, T. Griesman, J. Huang, S. Stuccio, A. Zingone, B. M. Ryan, J. S. Tsang, A. Biancotto, Assessment of Variability in the SOMAscan Assay. *Sci. Rep.* **7**, 14248 (2017). [doi:10.1038/s41598-017-14755-5](doi:10.1038/s41598-017-14755-5) Medline

35. K. J. Max Kuhn, *Applied Predictive Modeling* (Springer, 2013).

36. J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi Jr., M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, L. A. Garraway, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012). doi:10.1038/nature11003 Medline

37. B. MacLean, D. M. Tomazela, S. E. Abbatiello, S. Zhang, J. R. Whiteaker, A. G. Paulovich, S. A. Carr, M. J. Maccoss, Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal. Chem.* **82**, 10116–10124 (2010). [doi:10.1021/ac102179j](doi:10.1021/ac102179j) Medline

38. Y. Mohammed, D. Domański, A. M. Jackson, D. S. Smith, A. M. Deelder, M. Palmblad, C. H. Borchers, PeptidePicker: A scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J. Proteomics* **106**, 151–161 (2014). doi:10.1016/j.jprot.2014.04.018 Medline

39. B. MacLean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, M. J. MacCoss, Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010). doi:10.1093/bioinformatics/btq054 Medline

40. J. A. Vizcaíno, R. G. Côté, A. Csordas, J. A. Dianes, A. Fabregat, J. M. Foster, J. Griss, E. Alpi, M. Birim, J. Contell, G. O'Kelly, A. Schoenegger, D. Ovelleiro, Y. Pérez-Riverol, F. Reisinger, D. Ríos, R. Wang, H. Hermjakob, The PRoteomics IDEntifications (PRIDE) database and associated tools: Status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013). [doi:10.1093/nar/gks1262](doi:10.1093/nar/gks1262) Medline

41. M. M. Chan, R. Santhanakrishnan, J. P. C. Chong, Z. Chen, B. C. Tai, O. W. Liew, T. P. Ng, L. H. Ling, D. Sim, K. T. G. Leong, P. S. D. Yeo, H.-Y. Ong, F. Jaufeerally, R. C.-C. Wong, P. Chai, A. F. Low, A. M. Richards, C. S. P. Lam, Growth differentiation factor 15 in heart failure with preserved vs. reduced ejection fraction. *Eur. J. Heart Fail.* **18**, 81–88 (2016). doi:10.1002/ejhf.431 Medline

42. P. G. van Peet, A. J. de Craen, J. Gussekloo, W. de Ruijter, Plasma NT-proBNP as predictor of change in functional status, cardiovascular morbidity and mortality in the oldest old: The Leiden 85-plus study. *Age (Dordr.)* **36**, 9660 (2014). [doi:10.1007/s11357-014-9660-1](doi:10.1007/s11357-014-9660-1) Medline

43. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). [doi:10.1186/1471-2105-9-559](doi:10.1186/1471-2105-9-559) Medline

44. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A. L. Barabási, Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002). doi:10.1126/science.1073374 Medline

45. S. L. Carter, C. M. Brechbühler, M. Griffin, A. T. Bond, Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20**, 2242–2250 (2004). doi:10.1093/bioinformatics/bth234 Medline

46. M. C. Oldham, S. Horvath, D. H. Geschwind, Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17973–17978 (2006). doi:10.1073/pnas.0605938103 Medline

47. R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000). [doi:10.1038/35019019](doi:10.1038/35019019) Medline

48. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002). doi:10.1103/RevModPhys.74.47

49. J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004). doi:10.1038/nature02555 Medline

50. G. Chauhan, C. R. Arnold, A. Y. Chu, M. Fornage, A. Reyahi, J. C. Bis, A. S. Havulinna, M. Sargurupremraj, A. V. Smith, H. H. H. Adams, S. H. Choi, S. L. Pulit, S. Trompet, M. E. Garcia, A. Manichaikul, A. Teumer, S. Gustafsson, T. M. Bartz, C. Bellenguez, J. S. Vidal, X. Jian, O. Kjartansson, K. L. Wiggins, C. L. Satizabal, F. Xue, S. Ripatti, Y. Liu, J. Deelen, M. den Hoed, S. Bevan, J. C. Hopewell, R. Malik, S. R. Heckbert, K. Rice, N. L. Smith, C. Levi, P. Sharma, C. L. M. Sudlow, A. M. Nik, J. W. Cole, R. Schmidt, J. Meschia, V. Thijs, A. Lindgren, O. Melander, R. P. Grewal, R. L. Sacco, T. Rundek, P. M. Rothwell, D. K. Arnett, C. Jern, J. A. Johnson, O. R. Benavente, S. Wassertheil-Smoller, J.-M. Lee, Q. Wong, H. J. Aparicio, S. T. Engelter, M. Kloss, D. Leys, A. Pezzini, J. E. Buring, P. M. Ridker, C. Berr, J.-F. Dartigues, A. Hamsten, P. K. Magnusson, M. Traylor, N. L. Pedersen, L. Lannfelt, L. Lindgren, C. M. Lindgren, A. P. Morris, J. Jimenez-Conde, J. Montaner, F. Radmanesh, A. Slowik, D. Woo, A. Hofman, P. J. Koudstaal, M. L. P. Portegies, A. G. Uitterlinden, A. J. M. de Craen, I. Ford, J. W. Jukema, D. J. Stott, N. B. Allen, M. M. Sale, A. D. Johnson, D. A. Bennett, P. L. De Jager, C. C. White, H. J. Grabe, M. R. P. Markus, U. Schminke, G. B. Boncoraglio, R. Clarke, Y. Kamatani, J. Dallongeville, O. L. Lopez, J. I. Rotter, M. A. Nalls, R. F. Gottesman, M. E. Griswold, D. S. Knopman, B. G. Windham, A. Beiser, H. S. Markus, E. Vartiainen, C. R. French, M. Dichgans, T. Pastinen, M. Lathrop, V. Gudnason, T. Kurth, B. M. Psaty, T. B. Harris, S. S. Rich, A. L. deStefano, C. O. Schmidt, B. B. Worrall, J. Rosand, V. Salomaa, T. H. Mosley, E. Ingelsson, C. M. van Duijn, C. Tzourio, K. M. Rexrode, O. J. Lehmann, L. J. Launer, M. A. Ikram, P. Carlsson, D. I. Chasman, S. J. Childs, W. T. Longstreth, S. Seshadri, S. Debette; Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Stroke Genetics Network

(SiGN), and the International Stroke Genetics Consortium, Identification of additional risk loci for stroke and small vessel disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* **15**, 695–707 (2016). doi:10.1016/S1474-4422(16)00102-2

51. E. J. Foss, D. Radulovic, S. A. Shaffer, D. R. Goodlett, L. Kruglyak, A. Bedalov, Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLOS Biol.* **9**, e1001144 (2011). doi:10.1371/journal.pbio.1001144 Medline

52. C. Gaiteri, Y. Ding, B. French, G. C. Tseng, E. Sibille, Beyond modules and hubs: The potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav.* **13**, 13–24 (2014). doi:10.1111/gbb.12106 Medline

53. D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. E. Dolan, N. J. Cox, Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLOS Genet.* **6**, e1000888 (2010). doi:10.1371/journal.pgen.1000888 Medline

54. Å. Johansson, S. Enroth, M. Palmblad, A. M. Deelder, J. Bergquist, U. Gyllensten, Identification of genetic variants influencing the human plasma proteome. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4673–4678 (2013). doi:10.1073/pnas.1217238110 Medline

55. S. Kim, S. Swaminathan, M. Inlow, S. L. Risacher, K. Nho, L. Shen, T. M. Foroud, R. C. Petersen, P. S. Aisen, H. Soares, J. B. Toledo, L. M. Shaw, J. Q. Trojanowski, M. W. Weiner, B. C. McDonald, M. R. Farlow, B. Ghetti, A. J. Saykin; Alzheimer's Disease Neuroimaging Initiative (ADNI), Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLOS ONE* **8**, e70269 (2013). doi:10.1371/journal.pone.0070269 Medline

56. S. Enroth, A. Johansson, S. B. Enroth, U. Gyllensten, Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat. Commun.* **5**, 4684 (2014). doi:10.1038/ncomms5684 Medline

57. Y. Liu, A. Buil, B. C. Collins, L. C. Gillet, L. C. Blum, L.-Y. Cheng, O. Vitek, J. Mouritsen, G. Lachance, T. D. Spector, E. T. Dermitzakis, R. Aebersold, Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11**, 786 (2015). doi:10.15252/msb.20145728 Medline

58. B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S. Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic, D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M. Plenge, J. Danesh, H. Runz, A. S. Butterworth, Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018). doi:10.1038/s41586-018-0175-2 Medline

59. K. Suhre, M. Arnold, A. M. Bhagwat, R. J. Cotton, R. Engelke, J. Raffler, H. Sarwath, G. Thareja, A. Wahl, R. K. DeLisle, L. Gold, M. Pezer, G. Lauc, M. A. El-Din Selim, D. O. Mook-Kanamori, E. K. Al-Dous, Y. A. Mohamoud, J. Malek, K. Strauch, H. Grallert, A. Peters, G. Kastenmüller, C. Gieger, J. Graumann, Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017). doi:10.1038/ncomms14357 Medline

60. I. Bhattacharya, Z. Manukyan, P. Chan, A. Heatherington, L. Harnisch, Application of Quantitative Pharmacology Approaches in Bridging Pharmacokinetics and Pharmacodynamics of Domagrozumab From Adult Healthy Subjects to Pediatric Patients With Duchenne Muscular Disease. *J. Clin. Pharmacol.* **58**, 314–326 (2018). Medline

61. K. Kondás, G. Szláma, M. Trexler, L. Patthy, Both WFIKKN1 and WFIKKN2 have high affinity for growth and differentiation factors 8 and 11. *J. Biol. Chem.* **283**, 23677–23684 (2008). doi:10.1074/jbc.M803025200 Medline

62. H. Sun, Y. Zhu, H. Pan, X. Chen, J. L. Balestrini, T. T. Lam, J. E. Kanyo, A. Eichmann, M. Gulati, W. H. Fares, H. Bai, C. A. Feghali-Bostwick, Y. Gan, X. Peng, M. W. Moore, E. S. White, P. Sava, A. L. Gonzalez, Y. Cheng, L. E. Niklason, E. L. Herzog, Netrin-1 Regulates Fibrocyte Accumulation in the Decellularized Fibrotic Sclerodermatous Lung Microenvironment and in Bleomycin-Induced Pulmonary Fibrosis. *Arthritis Rheumatol.* **68**, 1251–1261 (2016). Medline

63. G. T. Consortium; GTEx Consortium, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi:10.1126/science.1262110 Medline

64. J. Wang, D. Duncan, Z. Shi, B. Zhang, WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).

doi:10.1093/nar/gkt439 Medline

65. J. E. Shoemaker, T. J. S. Lopes, S. Ghosh, Y. Matsuoka, Y. Kawaoka, H. Kitano, CTen: A web-based platform for identifying enriched cell types from heterogeneous microarray data. *BMC Genomics* **13**, 460 (2012). doi:10.1186/1471-2164-13-460 Medline

66. D. Warde-Farley, S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. D. Bader, Q. Morris, The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38** (suppl. 2), W214–W220 (2010). doi:10.1093/nar/gkq537 Medline

67. W. Huang, B. T. Sherman, R. A. Lempicki; W. Huang da, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009). doi:10.1038/nprot.2008.211 Medline

68. D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, H. Parkinson, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42** (D1), D1001–D1006 (2014). doi:10.1093/nar/gkt1229 Medline

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aaq1327/DC1
Materials and Methods
Figures S1 to S14
Tables S1 to S22
References (*31–68*)
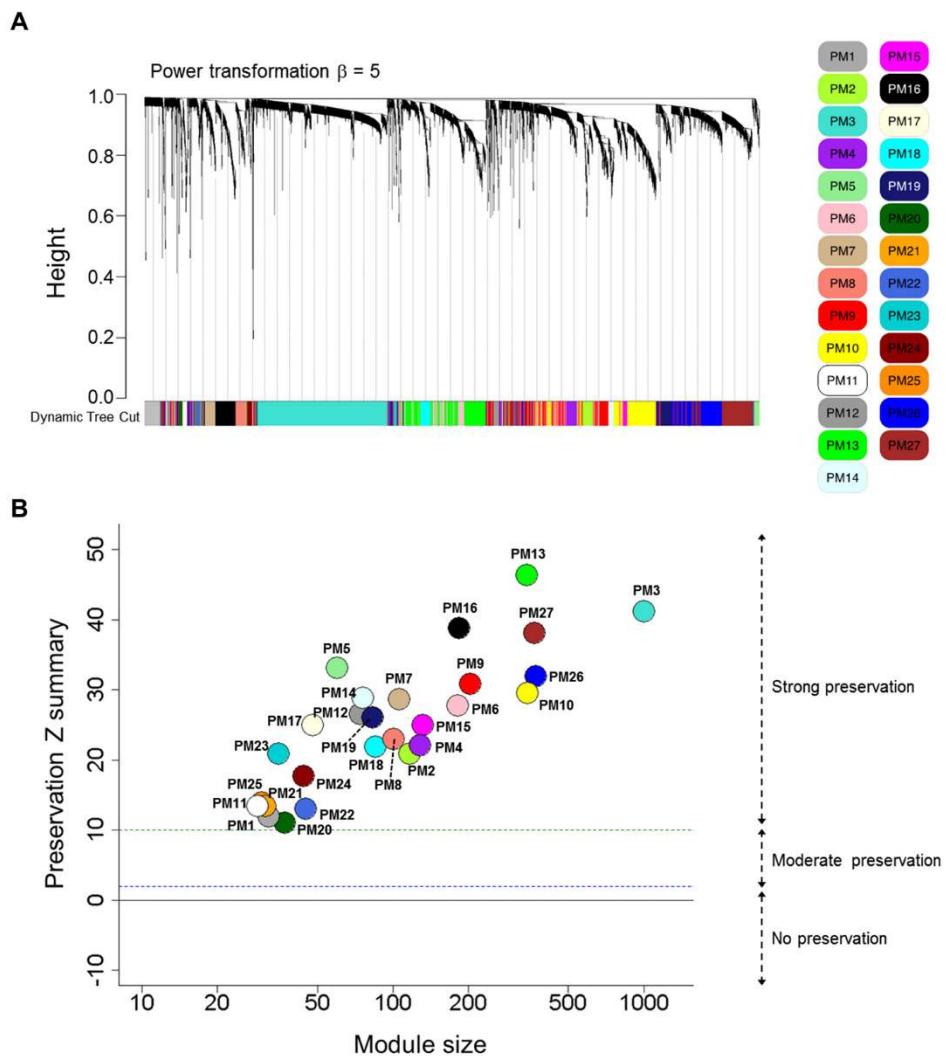
Fig. 1. The serum protein network structure. (**A**) Hierarchical clustering dendrogram using dynamic tree cut (*29*), revealing 27 serum protein modules. Each branch of the dendrogram represents a single protein and the colored bar below denotes its corresponding protein module, as annotated in the legend to the right. The dendrogram height is the distance between proteins (*14*). (**B**) The cohort was randomly split into two equal parts, one for a training set and another for the test set, and a summary Z score statistics (*15*), plotted for each module presented as colored data points. The summary Z score <2 (blue dotted line) indicates no preservation, 2< summary Z score <10 (between the blue and green dotted lines) indicates moderate evidence of preservation, while a summary Z score >10 (green dotted line) indicates strong evidence of preservation. See fig. S10 for the 2/3 vs. 1/3 split of the cohort.
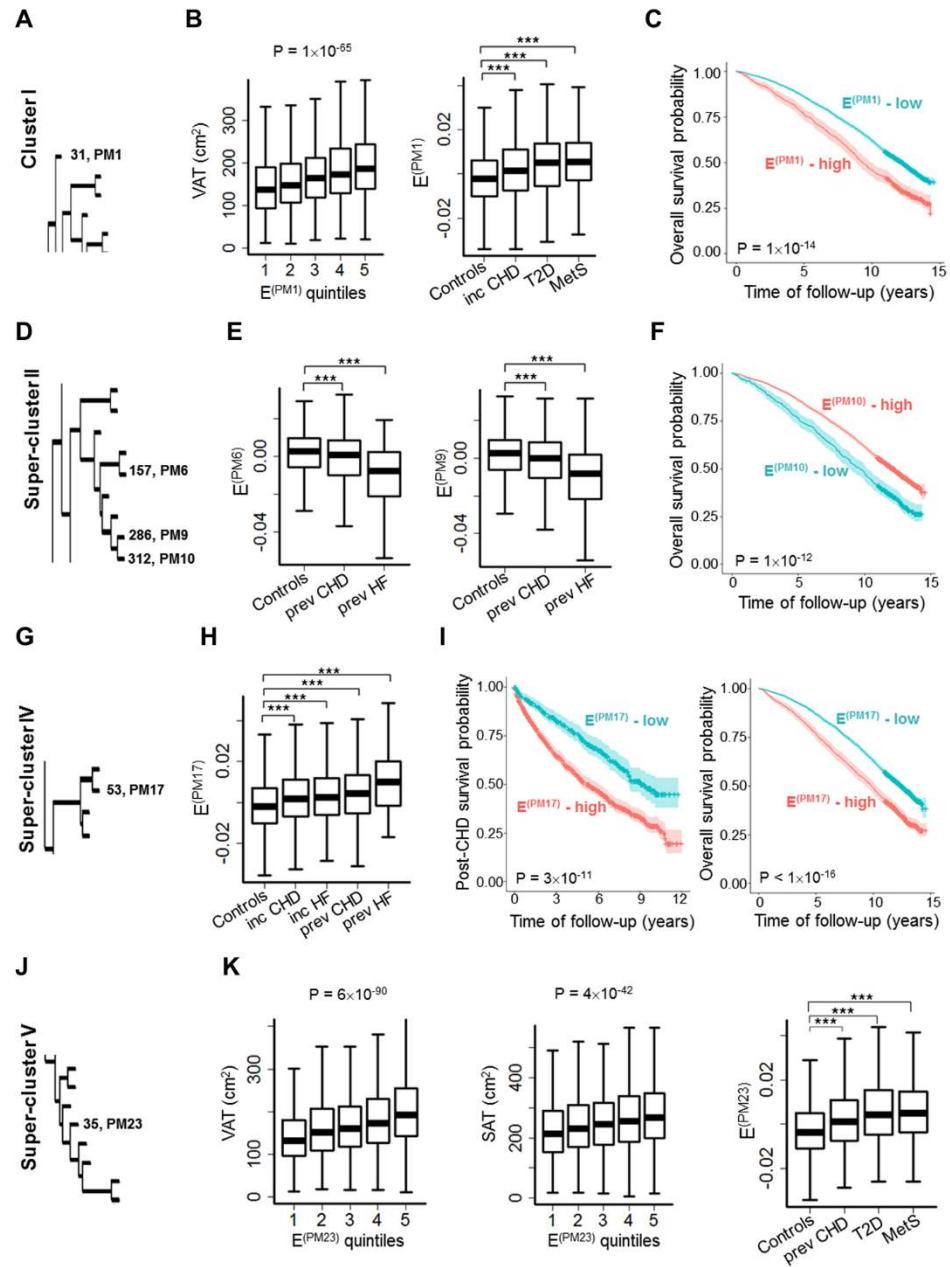
Fig. 2. The relationship between module´s $E^{(q)}$ to disease related measures. (A) The module PM1 is a single cluster of 31 proteins. (B) Positive associations of $E^{(PM1)}$ quintiles to variation (cm^2) in visceral adipose tissue (VAT), incident coronary heart disease (inc CHD), type 2 diabetes (T2D) and the metabolic syndrome (MetS), ***$P<1×10^{-10}$. (C) Overall survival, i.e., with respect to all-cause mortality, was reduced for high $E^{(PM1)}$ levels (red curve) compared to low $E^{(PM1)}$ levels (cyan curve). (D) The modules PM6, PM9 and PM10 are members of super-cluster II. (E) Inverse association of the $E^{(PM6)}$ and $E^{(PM9)}$ to prevalent CHD (prev CHD) and prevalent heart failure (prev HF), ***$P≤1×10^{-9}$. (F) Reduced overall survival for low $E^{(PM10)}$ levels (cyan curve) compared to high $E^{(PM10)}$ levels (red curve). (G) The PM17 is in super-cluster IV. (H) Positive association of module´s $E^{(PM17)}$ to incident CHD and HF as well as prevalent CHD and HF,***$P<1×10^{-17}$. (I) Reduced post-incident CHD survial as well as overall survival for high $E^{(PM17)}$ levels (red curve) compared to low $E^{(PM17)}$ levels (cyan curve). (J) The module PM23 is a member of super-cluster V. (K) Positive associations of the module $E^{(PM23)}$ quintiles to VAT and subcutaneous adipose tissue (SAT), and prevalent CHD, T2D and MetS, ***$P<1×10^{-13}$. Data were analyzed using forward linear or logistic regression or Cox proportional hazards regression, depending on the outcome being continuous, binary or a time to an event. Kaplan-Meier plots were used to display survival probabilities. The number of proteins per module are denoted at the branches of the dendrogram. Controls are subjects free of the disease in question.
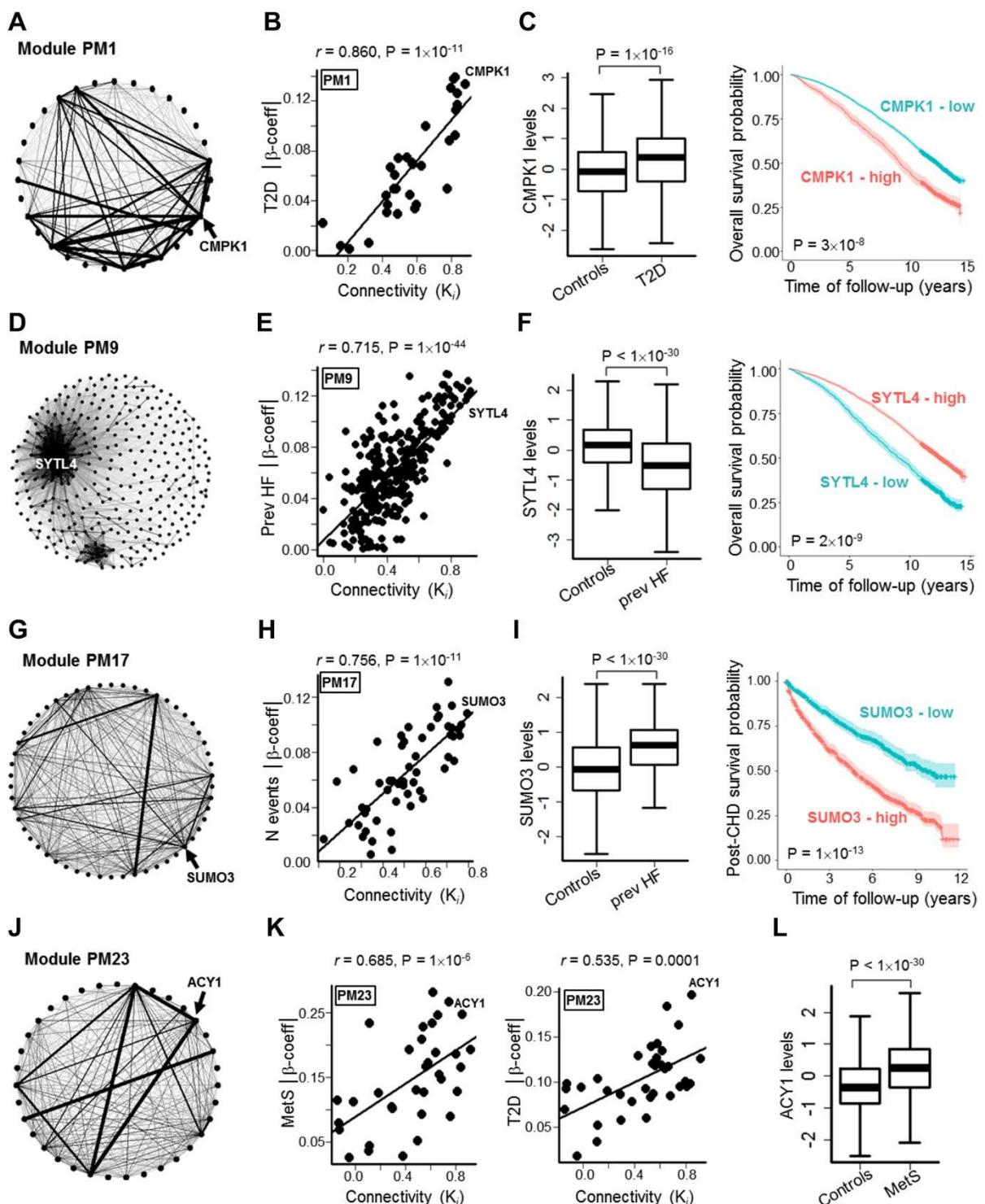
Fig. 3. The relationship between connectivity of proteins and disease related measures. (A) Circle graph of PM1 highlighting the hub protein CMPK1. (B) Positive correlation between within module connectivity (K_i) (x-axis) and the absolute value of the effect size of the association of proteins to type 2 diabetes (T2D) (y-axis). (C) Positive association of CMPK1 to T2D, and reduced overall survival associated with high serum CMPK1 levels (red curve). (D) Spring graph of PM9, highlighting the hub SYTL4. (E) Positive correlation between K_i and the association of proteins to prevalent heart failure (prev HF). (F) Inverse association of SYTL4 to HF, $P<1\times10^{-30}$, and reduced overall survival associated with low serum SYTL4 levels (cyan curve). (G) A circle graph of PM17 highlighting the hub protein SUMO3. (H) Positive correlation between the K_i of proteins and their association to prevalent HF. (I) SUMO3 is positively associated with prevalent HF, and high levels of SUMO3 (red curve) predict reduced survival post incident CHD. (J) A circle graph of the PM23 highlighting the hub ACY1. (K) Positive correlation between the K_i of proteins and their association to both MetS and T2D. (L) Strong positive association of ACY1 to MetS. Network visualization was performed with the igraph package in R (30). Pearson´s r was estimated for correlation between K_i of proteins and their strength of association to disease measures. See Fig. 2 for other relevant statistics.
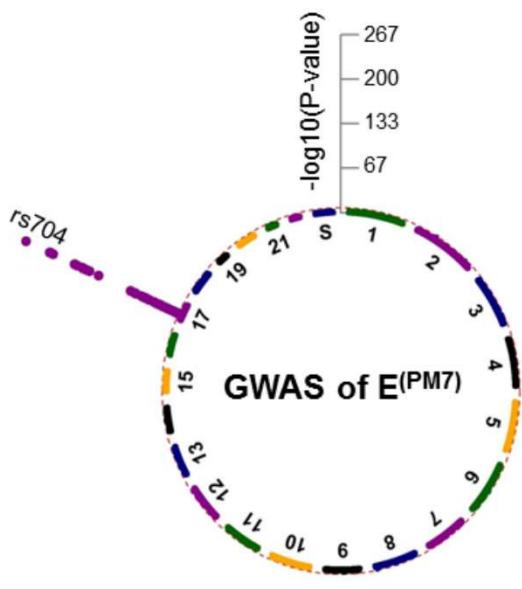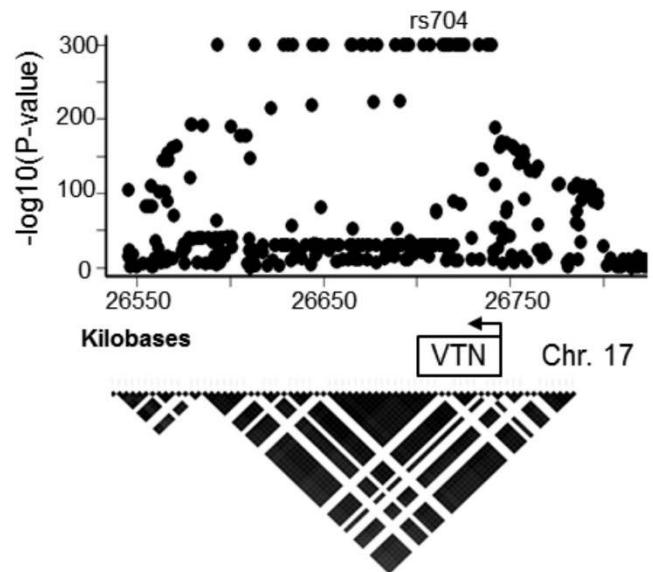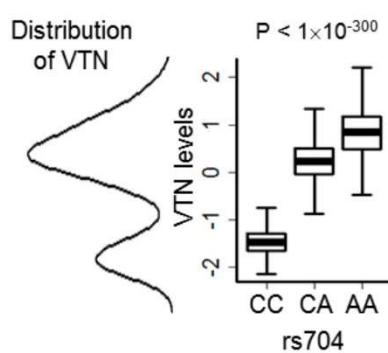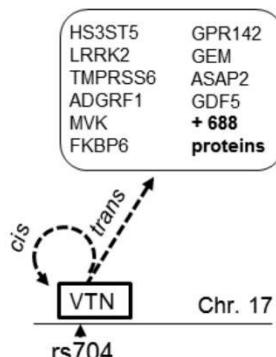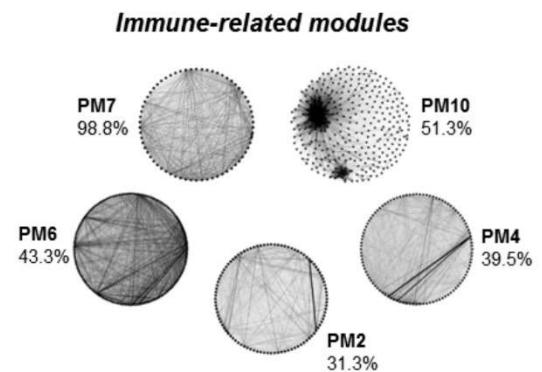
Fig. 4. The network associated pSNP rs704 regulates modules related to immune functions. (**A**) A circular Manhattan plot highlights the GWAS results for $E^{(PM7)}$, revealing a single highly significant association at rs704, a missense variant (NP_000629.3: p.Thr400Met) in VTN. Four other modules within super-cluster II were affected by rs704 (table S20). (**B**) The rs704 variant, and many other linked SNPs in the region, exerts a strong *cis* acting effect on VTN within the 300kb genomic region across the *VTN* gene. The black triangle demonstrates LD (r^2) patterns in the region derived from the AGES cohort data. (**C**) The population distribution of the VTN protein is bimodal explained by the drastic reduction in VTN levels in subjects homozygous for the rs704 C minor allele. (**D**) A schematic presentation of the single *cis* and many *trans* effects mediated by the rs704 in VTN on serum proteins. (**E**) Proteins affected by rs704 in *trans* (and VTN in *cis*) cluster in modules of immune-related functions (table S20). The percentage denotes the fraction of proteins within a given module regulated by the rs704 locus..
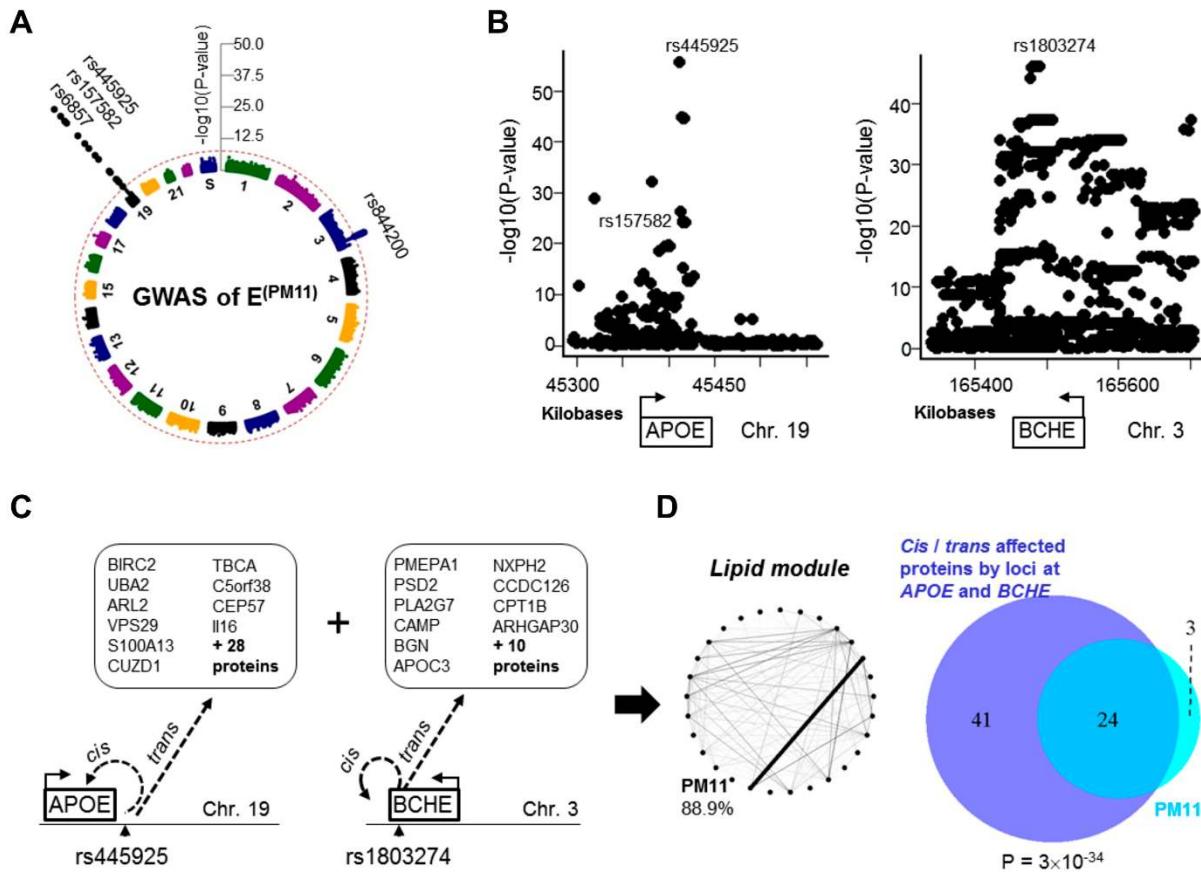
Fig. 5. Network-associated pSNPs at the APOE and BCHE loci regulate proteins of module PM11. (**A**) A circular Manhattan plot for the E$^{(PM11)}$, revealing two distinct genomic loci at chromosomes 3 (*BCHE*) and 19 (*APOE / TOMM40*). The three npSNPs at the *APOE/TOMM40* locus are not correlated (r^2=0). (**B**) The npSNPs at the two genomic regions exert a strong *cis* acting effect on the serum levels of APOE (left) and BCHE (right). (**C**) The npSNPs at the APOE locus affected APOE in *cis* and mediated *trans* effects on 38 proteins, while the npSNP rs1803274, a missense variant (NP_000046.1: pAla567Thr) in *BCHE*, affected BCHE in *cis* and 20 other proteins in *trans* (table S20). (**D**) The distinct npSNPs at the *APOE* and *BCHE* loci regulate 88.9% of all proteins that constitute PM11, Fisher's exact test P=3×10^{-34}, as demonstrated in the Venn diagram to the right. The number 3 refers to proteins in PM11, not regulated by the npSNPs.