Title

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

Permalink

Journal

ISSN

Authors

Emilsson, Valur

Ilkov, Marjan

Lamb, John R

et al.

Publication Date

DOI

Peer reviewed

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], and **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.

Abstract

Proteins circulating in the blood are critical for age-related disease processes; however, the serum proteome has remained largely unexplored. To this end, 4137 proteins covering most predicted

---

extracellular proteins were measured in the serum of 5457 Icelanders over 65 years of age. Pairwise 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 intercellular 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.

Heterochronic parabiosis experiments that surgically joined the circulation of young and old mice showed a system-wide effect on the re generative 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 caused not 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). Because blood mediates coordination between nonadjacent 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 have 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 (SOMAscan) 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 multipass transmembrane proteins. This resulted in an updated array of 5034 SOMAmers, 4783 of which recognize 4137 individual human proteins (table S1)—i.e., some proteins are targeted by more than one aptamer, whereas the rest recognize nonhuman targets. We applied the platform to a large population-based sampling of 5457 participants in the AGES Reykjavik study (12), a prospective study of deeply phenotyped and genotyped individuals older than 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 coexpression analysis (14). Co-regulation networks describe functional relationships that can reflect both physical and nonphysical 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 ratio in the protein adjacency matrix (Fig. 1A and fig. S4, A and 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), whereas 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 that 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. 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). Indeed, comparison of the serum protein modules to 2672 coexpression 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 that of 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 superclusters 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). Whereas some modules showed no association to any disease measure, others were associated to some or all of them (table S12; Fig. 2, A to K; and fig. S7, A to I). Notably, superclusters, although internally consistent, often showed opposing relationships between superclusters to CHD-related outcomes (table S12; Fig. 2, E, H, and K; and fig. S7, B, E, and 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. 2, B and K). Finally, modules associated with incident disease and to all-cause or post-CHD mortality (table S12; Fig. 2, C, F, and I; and fig. S7, C, F, and I) indicate that the protein network predicted future events and disease outcome. Time between diagnosis and sample collection 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 intramodule connectivity $\left(K_i = \dfrac{k_i}{k_{max}}\right)$ of proteins to various outcomes, where protein hubs showed stronger association to phenotypic measures than less well connected proteins (Fig. 3, A to L, and fig. S9, A 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). Protein single-nucleotide polymorphisms (pSNPs) are DNA sequence variants associated with allelic imbalance in protein levels. Using Bonferroni correction, we identified 1046 significant cis pSNP– protein associations within a 300-kb window across the corresponding protein-coding sequences (table S13 and fig. S11A). Cross-referencing cis pSNP-protein pairs to cis eSNPs (expression SNPs)– transcript data from more than 30 tissues and cells (20) revealed an overlap of 37.3% (table S14). This suggests that ~60% of the genetic effects on serum protein levels are mediated either by an as yet unknown transcriptional effect and/or by posttranscriptional mechanisms. The cis pSNP-proteins were underrepresented among highly connected protein nodes (fig. S11, B and 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 that various lead GWAS SNPs mediate distal trans effects on serum proteins (figs. S12, B 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 overrepresented in the PM27 module (fig. S12E). Genetic pleiotropy is a well-established phenomenon at the *ABO* locus (fig. S13G) (22). Indeed, the ABO blood groups have been linked to many diseases, which is in part mediated through their effect on the prothrombotic risk factor vWF (von Wille-brand factor)

(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 that 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 termed network-associated protein SNPs (npSNPs), of which many were associated with more than one module consistent with their relationship. Given that 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. 4, A to C, and fig. S13H), and 698 proteins in trans that were significant components of the modules enriched for immune-related pathways (Fig. 4, D and E, and table S20). Furthermore, distinct loci at *APOE* or *BCHE* were associated with the lipoprotein-enriched module PM11 (Fig. 5A, 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. 5, C and 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 whether 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 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 and thus contained nondiseased 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 pronounced 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 the 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 those 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 coexpression networks derived from solid tissues suggests that the serum protein network arose in part via systemic cross-tissue regulation. 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, comparable 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 in this study 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES AND NOTES

1. Schwenk JM et al., J. Proteome Res 16, 4299–4310 (2017). [PubMed: 28938075]

2. Uhlén M et al., Science 347, 1260419(2015). [PubMed: 25613900]

3. Stastna M, Van Eyk JE, Proteomics 12, 722–735 (2012). [PubMed: 22247067]

4. Conboy IM et al., Nature 433, 760–764 (2005). [PubMed: 15716955]

5. Villeda SA et al., Nature 477, 90–94 (2011). [PubMed: 21886162]

6. Schadt EE, Nature 461, 218–223 (2009). [PubMed: 19741703]

7. Zhang B et al., Cell 153, 707–720 (2013). [PubMed: 23622250]

8. Emilsson V et al., Nature 452, 423–428 (2008). [PubMed: 18344981]

9. Chen Y et al., Nature 452, 429–435 (2008). v [PubMed: 18344982]

10. Davies DR et al., Proc. Natl. Acad. Sci. U.S.A 109, 19971–19976 (2012). [PubMed: 23139410]

11. Gold L et al., PLOS ONE 5, e15004(2010). [PubMed: 21165148]

12. Harris TB et al., Am. J. Epidemiol 165, 1076–1087 (2007). [PubMed: 17351290]

13. Barabási AL, Albert R, Science 286, 509–512 (1999). [PubMed: 10521342]

14. Zhang B, Horvath S, Stat. Appl. Genet. Mol. Biol 4, e17(2005).

15. Langfelder P, Luo R, Oldham MC, Horvath S, PLOS Comput. Biol 7, e1001057(2011). [PubMed: 21283776]

16. Shu L et al., PLOS Genet. 13, e1007040(2017). [PubMed: 28957322]

17. Jeong H, Mason SP, Barabási AL, Oltvai ZN, Nature 411, 41–42 (2001). [PubMed: 11333967]

18. Barabási AL, Gulbahce N, Loscalzo J, Nat. Rev. Genet 12, 56–68 (2011). [PubMed: 21164525]

19. Muñoz M et al., Nat. Genet 48, 980–983 (2016). [PubMed: 27428752]

20. Staley JR et al., Bioinformatics 32, 3207–3209 (2016). [PubMed: 27318201]

21. Mähler N et al., PLOS Genet. 13, e1006402(2017). [PubMed: 28406900]

22. Pickrell JK et al., Nat. Genet 48, 709–717 (2016). [PubMed: 27182965]

23. Franchini M, Lippi G, BMC Med 13, 7(2015). [PubMed: 25592962]

24. Franchini M, Capra F, Targher G, Montagnana M, Lippi G, Thromb. J 5, 14(2007). [PubMed: 17894864]

25. Boyle EA, Li YI, Pritchard JK, Cell 169, 1177–1186 (2017). [PubMed: 28622505]

26. Alfego D, Rodeck U, Kriete A, PLOS ONE 13, e0190457(2018). [PubMed: 29293662]

27. Yang J et al., Sci. Rep 5, 15145(2015). [PubMed: 26477495]

28. Zahn JM et al., PLOS Genet. 3, e201(2007). [PubMed: 18081424]

29. Langfelder P, Zhang B, Horvath S, Bioinformatics 24, 719–720 (2008). [PubMed: 18024473]
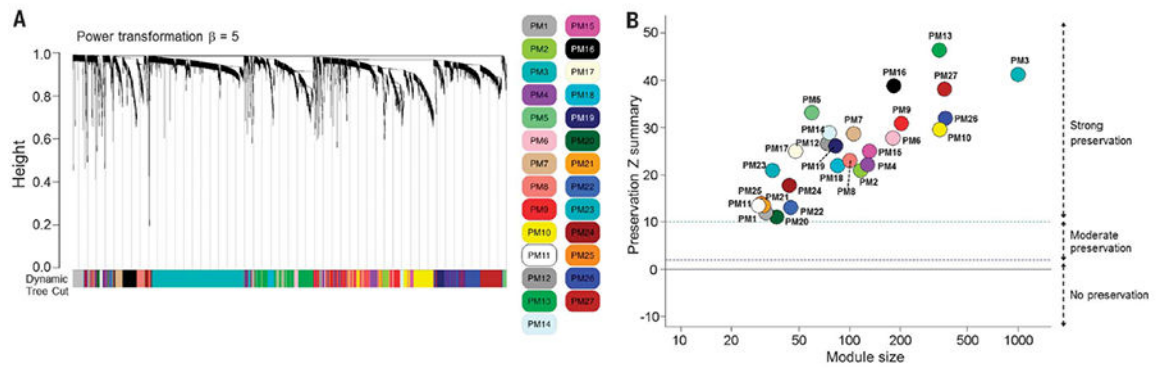
30. Csardi G, Nepusz T, Complex Syst 1695, 1(2006).

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; and a summary Z score >10 (green dotted line) indicates strong evidence of preservation. See fig. S10 for the $\frac{2}{3}$ versus $\frac{1}{3}$ split of the cohort.
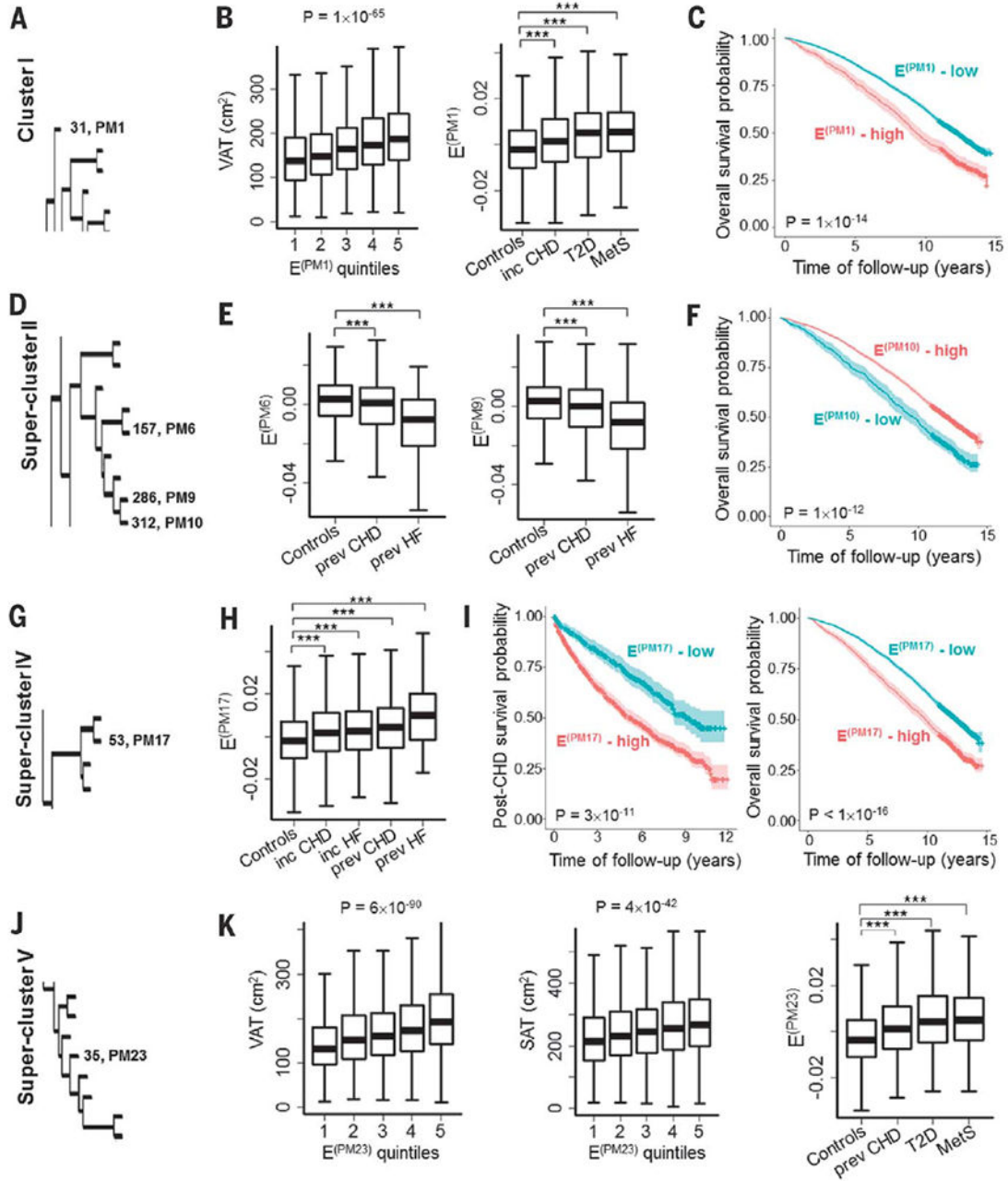
Fig. 2. The relationship between module's E(q) to disease-related measures.
(**A**) The module PM1 is a single cluster of 31 proteins. (**B**) Positive associations of $E^{(PM1)}$ quintiles to variation (cm^2) in visceral adipose tissue (VAT), incident coronary heart disease (inc CHD), type 2 diabetes (T2D), and the metabolic syndrome (MetS), ***$P < 1 \times 10^{-10}$. (**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 supercluster II. (**E**) Inverse association of the $E^{(PM6)}$ and $E^{(PM9)}$ to prevalent CHD (prev CHD) and prevalent heart failure (prev HF), ***$P \quad 1 \times 10^{-9}$. (**F**) Reduced overall survival for low $E^{(PM10)}$ levels (cyan curve) compared to high $E^{(PM10)}$

levels (red curve). (**G**) The PM17 is in supercluster IV. (**H**) Positive association of module's $E^{(PM17)}$ to incident CHD and HF as well as prevalent CHD and HF, ***$P < 1 \times 10^{-17}$. (**I**) Reduced postincident CHD survival 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 supercluster V. (**K**) Positive associations of the module $E^{(PM23)}$ quintiles to VAT and subcutaneous adipose tissue (SAT), and prevalent CHD, T2D, and MetS, ***$P < 1 \times 10^{-13}$. 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 is denoted at the branches of the dendrogram. Controls are individuals free of the disease in question.
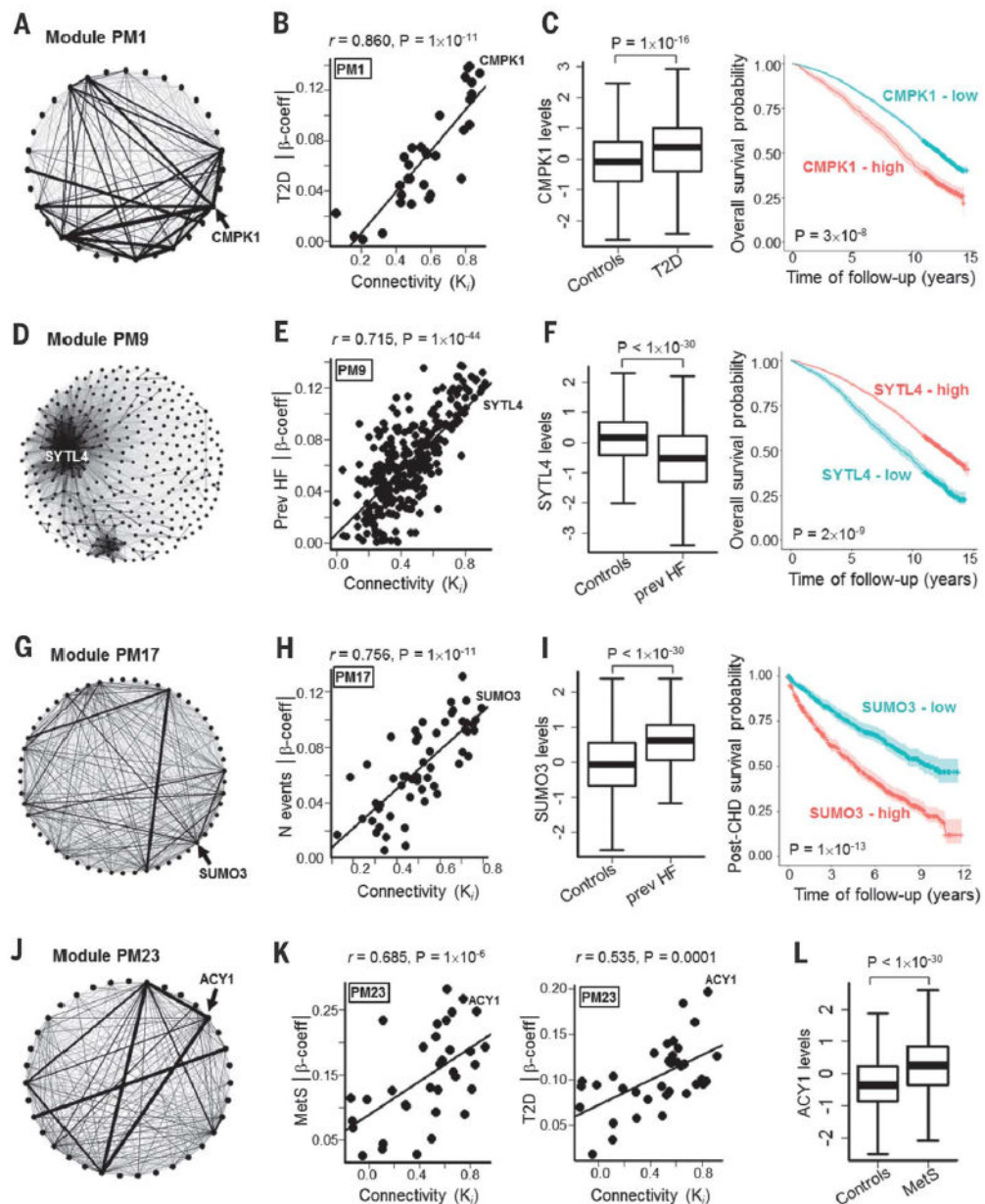
Fig. 3. The relationship between connectivity of proteins and disease-related measures.
(**A**) Circle graph of PM1 highlighting the hub protein CMPK1. (**B**) Positive correlation
between within module connectivity (Ki) (*x* axis) and the absolute value of the effect size of
the association of proteins 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 Ki and the association of proteins to prevalent heart failure (prev HF). (**F**) Inverse
association of SYTL4 to HF, $P < 1 \times 10^{-30}$, 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 Ki 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 postincident CHD. (**J**) A circle graph of the PM23 highlighting the hub ACY1. (**K**) Positive correlation between the Ki 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 Ki of proteins and their strength of association to disease measures. See Fig. 2 for other relevant statistics.
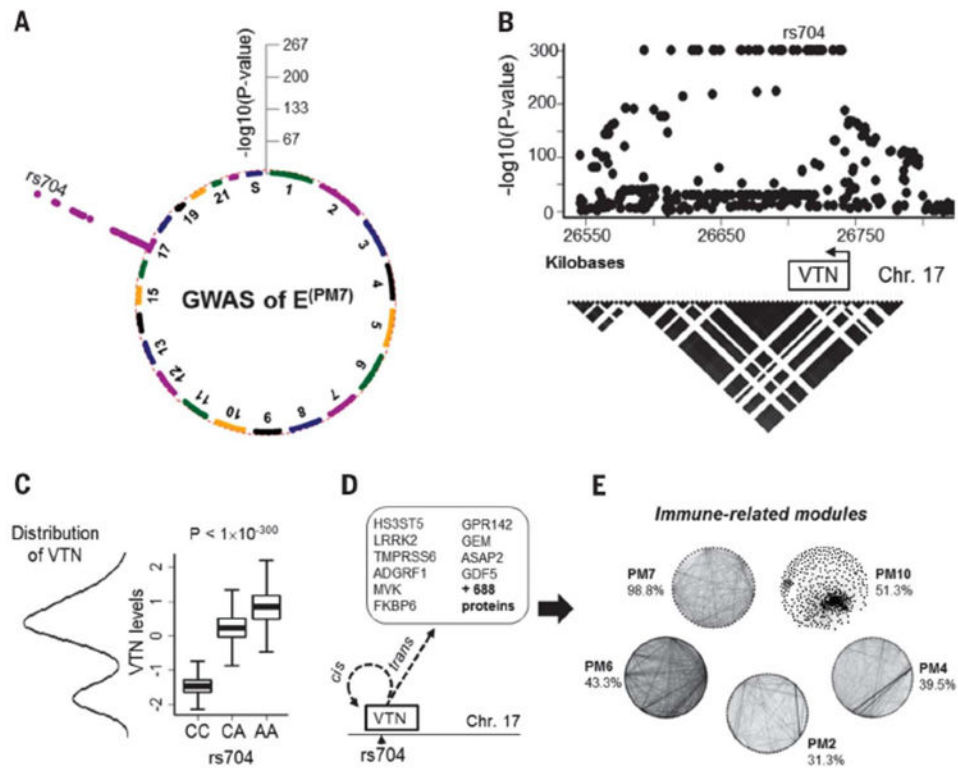
Fig. 4. The network-associated pSNP rs704 regulates modules related to immune functions.
(**A**) A circular Manhattan plot highlights the GWAS results for E$^{(PM7)}$, revealing a single highly significant association at rs704, a missense variant (NP_000629.3: p.Thr400Met) in VTN. Four other modules within supercluster II were affected by rs704 (table S20). (**B**) The rs704 variant, and many other linked SNPs in the region, exert a strong cis-acting effect on VTN within the 300-kb genomic region across the *VTN* gene. The black triangle demonstrates linkage disequilibrium (r^2) patterns in the region derived from the AGES cohort data. (**C**) The bimodal population distribution of the VTN protein is explained by the drastic reduction in VTN levels in individuals 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.
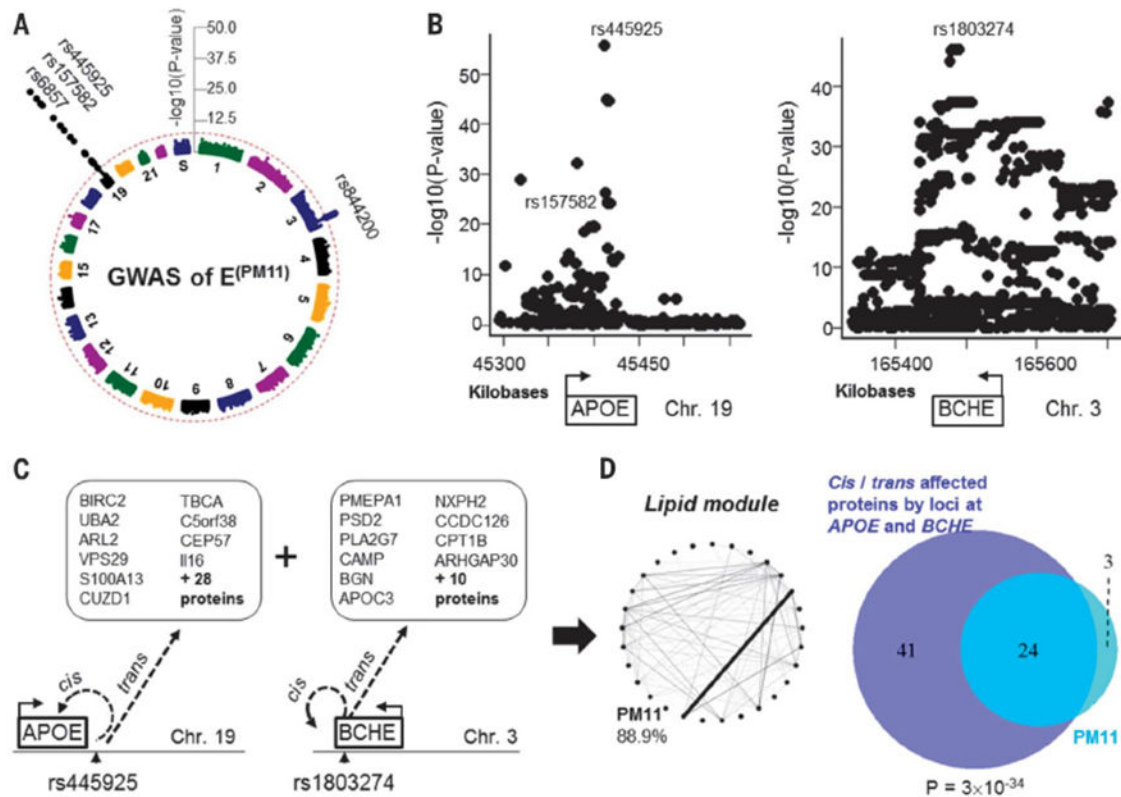
Fig. 5. Network-associated pSNPs at the APOE and BCHE loci regulate proteins of module PM11.

(**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, whereas 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 $P = 3 \times 10^{-34}$, as demonstrated in the Venn diagram to the right. The number 3 refers to proteins in PM11 that are not regulated by the npSNPs.