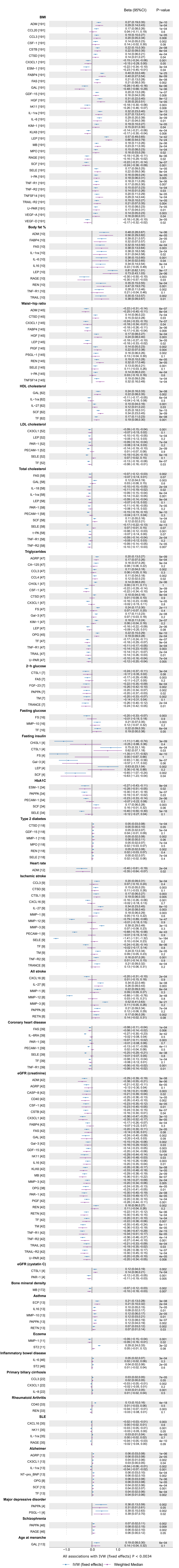


	Model (SNPs)	Beta per SD protein (95%CI)	P-value	P(het) All vs Cis	Evidence of causality
BMI					
RAGE	Pan (11) Cis (1)	-0.04 (-0.05,-0.03) -0.08 (-0.10,-0.05)	1.1e-09 4.9e-10	0.01	Strong
CTSD	Pan (5) Cis (1)	-0.02 (-0.03,-0.01) -0.01 (-0.02, 0.01)	4.3e-03 0.27	0.35	Intermediate
CXCL1	Pan (2) Cis (1)	-0.03 (-0.04,-0.02) -0.01 (-0.03, 0.00)	4.8e-06 0.06	0.08	Intermediate
CXCL16	Pan (9) Cis (1)	-0.02 (-0.04,-0.01) -0.03 (-0.06, 0.00)	1.1e-03 0.04	0.70	Intermediate
Dkk-1	Pan (5) Cis (1)	-0.03 (-0.05,-0.02) -0.01 (-0.03, 0.01)	6.2e-06 0.35	0.14	Intermediate
FABP4	Pan (2) Cis (1)	0.10 (0.06, 0.13) 0.07 (0.02, 0.12)	3.6e-07 7.2e-03	0.39	Intermediate
MMP-1	Pan (5) Cis (1)	0.02 (0.01, 0.03) 0.01 (0.00, 0.02)	5.1e-04 0.16	0.21	Intermediate
NT-pro_BNP	Pan (1) Cis (1)	0.02 (0.01, 0.04) 0.02 (0.01, 0.04)	3.9e-03 3.9e-03	1.00	Intermediate
ST2	Pan (6) Cis (1)	0.01 (0.01, 0.02) 0.01 (0.00, 0.01)	1.1e-04 5.3e-03	0.53	Intermediate
TNF-R2	Pan (5) Cis (1)	-0.03 (-0.04,-0.01) -0.01 (-0.03, 0.01)	2.1e-03 0.39	0.25	Intermediate
Waist-hip ratio					
ADM	Pan (4) Cis (1)	0.00 (-0.02, 0.01) 0.08 (0.05, 0.10)	0.88 6.6e-09	0.00	Strong
CSF-1	Pan (1) Cis (1)	0.06 (0.04, 0.08) 0.06 (0.04, 0.08)	3.8e-08 3.8e-08	1.00	Strong
Gal-3	Pan (6) Cis (1)	0.02 (0.01, 0.03) 0.01 (0.00, 0.03)	3.8e-04 0.02	0.56	Intermediate
HB-EGF	Pan (6) Cis (1)	-0.06 (-0.07,-0.04) -0.06 (-0.10,-0.01)	1.2e-12 7.9e-03	0.94	Intermediate
LOX-1	Pan (7) Cis (1)	0.04 (0.02, 0.06) 0.00 (-0.05, 0.04)	1.1e-06 0.93	0.07	Intermediate
PAR-1	Pan (2) Cis (1)	-0.03 (-0.05,-0.01) 0.00 (-0.03, 0.03)	4.2e-03 0.82	0.12	Intermediate
RAGE	Pan (11) Cis (1)	0.02 (0.01, 0.04) 0.03 (0.00, 0.05)	1.4e-03 0.02	0.59	Intermediate
TF	Pan (4) Cis (1)	-0.02 (-0.03,-0.01) 0.00 (-0.02, 0.02)	5.0e-05 0.96	0.06	Intermediate
TRAIL	Pan (7) Cis (1)	-0.01 (-0.02,-0.01) 0.00 (-0.02, 0.02)	2.0e-03 0.80	0.35	Intermediate
HDL cholesterol					
GDF-15	Pan (1) Cis (1)	-0.07 (-0.11,-0.03) -0.07 (-0.11,-0.03)	3.4e-04 3.4e-04	1.00	Strong
CSF-1	Pan (1) Cis (1)	-0.08 (-0.14,-0.03) -0.08 (-0.14,-0.03)	2.5e-03 2.5e-03	1.00	Intermediate
CXCL1	Pan (2) Cis (1)	0.05 (0.02, 0.08) 0.01 (-0.02, 0.05)	3.1e-03 0.50	0.14	Intermediate
IL-27	Pan (8) Cis (1)	-0.02 (-0.04,-0.01) -0.01 (-0.02, 0.01)	1.6e-03 0.52	0.11	Intermediate
LEP	Pan (2) Cis (1)	-0.18 (-0.26,-0.09) -0.04 (-0.17, 0.08)	3.9e-05 0.50	0.09	Intermediate
PAR-1	Pan (2) Cis (1)	0.14 (0.09, 0.19) 0.05 (-0.02, 0.13)	2.4e-07 0.17	0.08	Intermediate
RAGE	Pan (9) Cis (1)	0.10 (0.06, 0.13) 0.07 (0.01, 0.12)	8.8e-08 0.02	0.40	Intermediate
SELE	Pan (5) Cis (1)	-0.03 (-0.04,-0.01) 0.01 (-0.09, 0.10)	1.5e-04 0.68	0.48	Intermediate
LDL cholesterol					
FGF-23	Pan (7) Cis (1)	0.12 (0.07, 0.16) -0.04 (-0.20, 0.12)	4.2e-06 0.82	0.06	Intermediate
FS	Pan (3) Cis (1)	0.09 (0.05, 0.13) 0.05 (-0.03, 0.13)	3.2e-05 0.23	0.35	Intermediate
IL-1ra	Pan (2) Cis (1)	0.08 (0.03, 0.13) 0.08 (0.03, 0.13)	2.8e-03 2.5e-03	0.91	Intermediate
TNF-R2	Pan (4) Cis (1)	-0.13 (-0.19,-0.08) -0.05 (-0.13, 0.03)	4.0e-06 0.24	0.12	Intermediate
Total cholesterol					
IL-1ra	Pan (2) Cis (1)	0.11 (0.06, 0.16) 0.12 (0.07, 0.17)	1.2e-05 2.4e-06	0.71	Strong
RAGE	Pan (9) Cis (1)	0.07 (0.03, 0.10) 0.12 (0.06, 0.18)	2.5e-04 4.7e-05	0.14	Strong
HGF	Pan (1) Cis (1)	-0.11 (-0.18,-0.04) -0.11 (-0.18,-0.04)	2.8e-03 2.8e-03	1.00	Intermediate
KIM-1	Pan (6) Cis (1)	0.04 (0.02, 0.06) 0.03 (0.00, 0.05)	1.5e-04 0.02	0.39	Intermediate
SELE	Pan (5) Cis (1)	-0.05 (-0.06,-0.04) 0.04 (-0.06, 0.14)	3.1e-12 0.42	0.08	Intermediate
Triglycerides					
HGF	Pan (1) Cis (1)	-0.15 (-0.21,-0.08) -0.15 (-0.21,-0.08)	3.2e-05 3.2e-05	1.00	Strong
RAGE	Pan (9) Cis (1)	0.06 (0.03, 0.10) 0.13 (0.08, 0.18)	2.3e-04 2.2e-06	0.04	Strong
KIM-1	Pan (6) Cis (1)	0.03 (0.01, 0.05) 0.02 (0.00, 0.04)	5.2e-04 0.02	0.47	Intermediate
2-h glucose					
IL16	Pan (1) Cis (1)	0.08 (0.04, 0.12) 0.08 (0.04, 0.12)	2.4e-05 2.4e-05	1.00	Strong
Insulin secretion					
Dkk-1	Pan (5) Cis (1)	-0.25 (-0.42,-0.08) -0.22 (-0.50, 0.07)	4.0e-03 0.14	0.84	Intermediate
Fasting insulin					
CCL20	Pan (2) Cis (1)	-0.16 (-0.26,-0.06) -0.16 (-0.26,-0.06)	1.6e-03 1.7e-03	0.98	Intermediate
LEP	Pan (2) Cis (1)	0.14 (0.06, 0.22) 0.04 (-0.07, 0.15)	7.4e-04 0.49	0.14	Intermediate
HbA1C					
TF	Pan (3) Cis (1)	0.04 (0.01, 0.06) 0.07 (0.03, 0.10)	1.6e-03 9.1e-04	0.24	Strong
EGF	Pan (1) Cis (1)	-0.08 (-0.14,-0.03) -0.08 (-0.14,-0.03)	3.8e-03 3.8e-03	1.00	Intermediate
FS	Pan (3) Cis (1)	-0.08 (-0.11,-0.04) -0.04 (-0.09, 0.01)	6.1e-06 0.12	0.27	Intermediate
GDF-15	Pan (1) Cis (1)	0.04 (0.01, 0.07) 0.04 (0.01, 0.07)	3.9e-03 3.9e-03	1.00	Intermediate
LEP	Pan (2) Cis (1)	0.17 (0.10, 0.24) 0.07 (-0.03, 0.17)	7.5e-06 0.18	0.11	Intermediate
RAGE	Pan (9) Cis (1)	-0.05 (-0.08,-0.02) -0.06 (-0.10,-0.01)	9.5e-04 0.01	0.75	Intermediate
SCF	Pan (11) Cis (1)	-0.03 (-0.05,-0.01) 0.02 (-0.09, 0.14)	2.0e-03 0.70	0.38	Intermediate
Type 2 diabetes					
PAPPA	Pan (11) Cis (1)	0.01 (-0.02, 0.05) -0.27 (-0.42,-0.11)	0.42 8.6e-04	0.00	Strong
RAGE	Pan (11) Cis (1)	-0.08 (-0.13,-0.04) -0.17 (-0.27,-0.08)	5.8e-04 2.5e-04	0.09	Strong
CCL4	Pan (4) Cis (1)	0.04 (0.02, 0.07) 0.01 (-0.03, 0.06)	1.5e-03 0.55	0.26	Intermediate
EGF	Pan (1) Cis (1)	0.15 (0.05, 0.25) 0.15 (0.05, 0.25)	2.8e-03 2.8e-03	1.00	Intermediate
LOX-1	Pan (7) Cis (1)	0.14 (0.08, 0.20) 0.10 (-0.05, 0.26)	2.2e-06 0.20	0.64	Intermediate
MMP-12	Pan (2) Cis (1)	-0.04 (-0.06,-0.01) -0.04 (-0.06,-0.01)	3.6e-03 3.2e-05	0.96	Intermediate
MPO	Pan (10) Cis (1)	0.10 (0.05, 0.15) 0.02 (-0.06, 0.10)	2.7e-05 0.60	0.09	Intermediate
SELE	Pan (8) Cis (1)	-0.06 (-0.07,-0.04) 0.06 (-0.06, 0.18)	2.1e-09 0.35	0.07	Intermediate
U-PAR	Pan (6) Cis (1)	0.09 (0.03, 0.15) 0.05 (-0.07, 0.17)	2.1e-03 0.42	0.54	Intermediate
Heart rate					
CTSL1	Pan (8) Cis (1)	0.03 (0.01, 0.06) 0.05 (0.00, 0.11)	4.2e-03 0.06	0.52	Intermediate
Atrial fibrillation					
CHI3L1	Pan (4) Cis (1)	-0.04 (-0.06,-0.02) -0.04 (-0.07,-0.02)	4.3e-04 4.3e-04	0.92	Strong
IL-6RA	Pan (1) Cis (1)	-0.04 (-0.06,-0.03) -0.04 (-0.06,-0.03)	4.4e-09 4.4e-09	1.00	Strong
SPON1	Pan (1) Cis (1)	0.14 (0.08, 0.20) 0.14 (0.08, 0.20)	1.2e-06 1.2e-06	1.00	Strong
Dkk-1	Pan (5) Cis (1)	0.13 (0.08, 0.18) 0.03 (-0.06, 0.12)	2.2e-06 5.48	0.07	Intermediate
HB-EGF	Pan (6) Cis (1)	0.19 (0.12, 0.25) 0.12 (-0.01, 0.25)	2.5e-09 0.07	0.38	Intermediate
LEP	Pan (2) Cis (1)	0.27 (0.13, 0.40) 0.13 (-0.05, 0.31)	9.3e-05 0.16	0.21	Intermediate
Ischemic stroke					
MMP-12	Pan (2) Cis (1)	-0.09 (-0.13,-0.05) -0.10 (-0.14,-0.06)	2.8e-06 5.5e-07	0.80	Strong
CD40	Pan (3) Cis (1)	-0.17 (-0.28,-0.06) -0.11 (-0.27, 0.04)	3.0e-03 0.14	0.57	Intermediate
ADM	Pan (2) Cis (1)	-0.07 (-0.12,-0.03) -0.07 (-0.12,-0.03)	9.0e-04 1.4e-03	0.96	Intermediate
SELE	Pan (8) Cis (1)	-0.07 (-0.10,-0.04) 0.10 (-0.10, 0.29)	4.0e-06 0.34	0.10	Intermediate
TF	Pan (5) Cis (1)	-0.17 (-0.23,-0.11) -0.07 (-0.18, 0.04)	3.5e-08 0.12	0.10	Intermediate
TM	Pan (4) Cis (1)	0.13 (0.08, 0.19) 0.08 (0.02, 0.14)	3.1e-06 0.01	0.22	Intermediate
All stroke					
CD40	Pan (2) Cis (1)	-0.08 (-0.12,-0.04) -0.08 (-0.12,-0.04)	1.3e-04 1.6e-04	0.98	Strong
MMP-12	Pan (2) Cis (1)	-0.09 (-0.13,-0.05) -0.09 (-0.13,-0.05)	4.3e-06 1.2e-06	0.83	Strong
MMP-3	Pan (3) Cis (1)	0.06 (0.02, 0.10) 0.05 (0.02, 0.09)	3.7e-03 6.2e-03	0.93	Intermediate
SELE	Pan (8) Cis (1)	-0.06 (-0.08,-0.03) 0.06 (-0.12, 0.24)	1.0e-04 0.52	0.22	Intermediate
TF	Pan (4) Cis (1)	-0.14 (-0.20,-0.09) -0.05 (-0.15, 0.05)	6.3e-07 0.36	0.10	Intermediate
TM	Pan (4) Cis (1)	0.11 (0.06, 0.16) 0.07 (0.02, 0.13)	3.1e-05 0.01	0.39	Intermediate
Coronary heart disease					
IL-6RA	Pan (1) Cis (1)	-0.05 (-0.07,-0.03) -0.05 (-0.07,-0.03)	2.0e-07 2.0e-07	1.00	Strong
PIGF	Pan (2) Cis (1)	-0.31 (-0.43,-0.18) -0.35 (-0.51,-0.19)	3.6e-06 1.6e-05	0.67	Strong
MMP-12	Pan (2) Cis (1)	-0.05 (-0.09,-0.02) -0.06 (-0.10,-0.02)	4.0e-03 2.2e-03	0.88	Intermediate
OPG	Pan (5) Cis (1)	-0.11 (-0.18,-0.04) 0.01 (-0.10, 0.11)	2.8e-03 0.88	0.07	Intermediate
SCF	Pan (13) Cis (1)	-0.13 (-0.18,-0.08) -0.04 (-0.31, 0.24)	8.2e-07 0.80	0.52	Intermediate
SELE	Pan (8) Cis (1)	-0.06 (-0.09,-0.04) 0.06 (-0.13, 0.25)	1.1e-06 0.54	0.21	Intermediate
TF	Pan (4) Cis (1)	-0.15 (-0.20,-0.09) -0.08 (-0.18, 0.02)	3.1e-07 0.12	0.25	Intermediate
eGFR (creatinine)					
FS	Pan (4) Cis (1)	0.16 (0.12, 0.20) 0.12 (0.05, 0.18)	7.1e-17 3.2e-04	0.23	Strong
ST2	Pan (5) Cis (1)	-0.02 (-0.03,-0.01) -0.02 (-0.03, 0.00)	2.7e-03 0.02	0.71	Intermediate
TNF-R2	Pan (5) Cis (1)	-0.09 (-0.13,-0.04) -0.02 (-0.09, 0.04)	2.2e-04 0.46	0.12	Intermediate
eGFR (cystatin C)					
CXCL16	Pan (10) Cis (1)	0.12 (0.04, 0.20) 0.14 (-0.03, 0.32)	3.6e-03 0.11	0.79	Intermediate
FS	Pan (4) Cis (1)	0.17 (0.09, 0.25) 0.11 (-0.03, 0.29)	1.1e-04 0.12	0.49	Intermediate
TNF-R2	Pan (5) Cis (1)	-0.25 (-0.36,-0.15) -0.10 (-0.24, 0.04)	1.5e-06 0.15	0.09	Intermediate
Bone fracture					
CX3CL1	Pan (5) Cis (1)	0.08 (0.03, 0.13) 0.11 (0.04, 0.17)	1.7e-03 9.6e-04	0.53	Strong
Dkk-1	Pan (5) Cis (1)	0.11 (0.06, 0.17) 0.36 (0.27, 0.45)	2.2e-05 2.1e-14	0.00	Strong
TRANCE	Pan (7) Cis (1)	-0.06 (0.10,-0.02) -0.10 (-0.21, 0.00)	2.5e-03 0.06	0.45	Intermediate
Bone mineral density					
CSF-1	Pan (1) Cis (1)	-0.04 (-0.06,-0.02) -0.04 (-0.06,-0.02)	1.3e-04 1.3e-04	1.00	Strong
Dkk-1	Pan (5) Cis (1)	-0.18 (-0.20,-0.17) -0.57 (-0.59,-0.54)	3.4e-129 9.2e-03	0.00	Strong
EGF	Pan (1) Cis (1)	0.07 (0.04, 0.10) 0.07 (0.04, 0.10)	1.6e-06 1.6e-06	1.00	Strong
IL-18	Pan (6) Cis (1)	-0.03 (-0.04,-0.02) -0.08 (-0.10,-0.06)	9.0e-06 2.2e-16	0.00	Strong
PIGF	Pan (2) Cis (1)	-0.10 (-0.13,-0.08) -0.14 (-0.17,-0.11)	1.2e-14 2.8e-18	0.05	Strong
TRANCE	Pan (7) Cis (1)	0.04 (0.03, 0.05) 0.26 (0.23, 0.29)	8.8e-13 1.3e-62	0.00	Strong
ADM	Pan (4) Cis (1)	0.04 (0.02, 0.06) 0.03 (0.01, 0.06)	9.6e-06 9.6e-03	0.63	Intermediate
CCL3	Pan (2) Cis (1)	0.02 (0.01, 0.03) 0.01 (0.00, 0.02)	2.5e-03 8.0e-03	0.81	Intermediate
CXCL16	Pan (9) Cis (1)	-0.02 (-0.04,-0.01) -0.04 (-0.07,-0.02)	1.2e-03 2.1e-03	0.23	Intermediate
FGF-23	Pan (6) Cis (1)	-0.06 (-0.08,-0.04) -0.11 (-0.06, 0.05)	2.6e-09 7.2e-03	0.06	Intermediate
MMP-3	Pan (3) Cis (1)	-0.01 (-0.02, 0.00) -0.01 (-0.02, 0.00)	3.7e-03 2.9e-03	0.94	Intermediate
OPG	Pan (5) Cis (1)	0.03 (0.02, 0.04) 0.03 (0.01, 0.05)	1.4e-05 4.8e-03	0.97	Intermediate
PAPPA	Pan (11) Cis (1)	-0.02 (-0.03,-0.01) -0.03 (-0.07, 0.02)	4.5e-07 0.26	0.93	Intermediate
RAGE	Pan (11) Cis (1)	-0.04 (-0.05,-0.03) -0.02 (-0.05, 0.00)	2.7e-09 0.07	0.27	Intermediate
SCF	Pan (13) Cis (1)	0.03 (0.02, 0.04) 0.02 (-0.03, 0.08)	2.7e-09 0.45	0.73	Intermediate
SELE	Pan (7) Cis (1)	0.03 (0.03, 0.04) 0.02 (-0.01, 0.01)	9.3e-31 0.19	0.71	Intermediate
TRAIL	Pan (7) Cis (1)	-0.02 (-0.02,-0.01) 0.00 (-0.02, 0.02)	3.8e-04 0.87	0.13	Intermediate
Asthma					
CASP-8	Pan (1) Cis (1)	0.32 (0.14, 0.51) 0.32 (0.14, 0.51)	6.1e-04 6.1e-04	1.00	Strong
IL-6RA	Pan (1) Cis (1)	0.07 (0.04, 0.09) 0.07 (0.04, 0.09)	6.3e-07 6.3e-07	1.00	Strong
ST2	Pan (3) Cis (1)	-0.17 (-0.22,-0.13) -0.18 (-0.22,-0.14)	2.0e-18 4.0e-19	0.85	Strong
CTSL1	Pan (8) Cis (1)	-0.17 (-0.24,-0.08) -0.13 (-0.34, 0.08)	1.2e-04 0.22	0.72	Intermediate
IL-27	Pan (9) Cis (1)	-0.07 (-0.11,-0.03) -0.03 (-0.08, 0.01)	1.3e-03 0.14	0.30	Intermediate
Eczema					
IL-6RA	Pan (1) Cis (1)	0.08 (0.04, 0.11) 0.08 (0.04, 0.11)	1.0e-05 1.0e-05	1.00	Strong
MMP-12	Pan (2) Cis (1)	0.13 (0.06, 0.20) 0.13 (0.06, 0.20)	1.9e-04 2.0e-04	0.99	Strong
Inflammatory bowel disease					
CD40	Pan (2) Cis (1)	-0.14 (-0.21,-0.07) 			



Supplementary Notes

Systems biology

The purpose of the network analysis was to identify the most likely chain of events (via protein interactions or other forms of regulation) leading from the trans-pQTL to the plasma protein being affected. For each of the trans-pQTL signals, all genes within a distance of 1 Mb of the SNP were considered to be a possible cis-gene intermediary between the SNP and the plasma protein. For investigating possible chains of events acting via physical interactions, we utilized the protein-protein interaction (PPI) network inBio Map™ (InWeb_InBioMap)¹ (<https://www.intomics.com/inbio/map.html>), and for investigating other potential regulatory chains we generated an association network based on text mining of PubMed abstracts. The inBio Map PPI resource was selected because it is built entirely on experimentally determined interactions, has a large coverage of the human interactome and furthermore, provides a confidence score on the experimental support for each edge in the network allowing for fine-grained control of what to include. In this study, we only used the high confidence subset of the global PPI network (see details below). Possible chains of events were then defined as paths in the PPI network connecting the protein product of the possible cis-gene intermediaries to the plasma protein whose protein level was affected. Under the assumption that the shortest path is also the most likely path, we calculated the statistical significance of each path by a network permutation scheme where the shortest path of each cis-gene intermediary was compared to the shortest path in a set of randomized networks.

For this approach to work, it is important that the random networks are similar to the original PPI network. Hence, they should contain the same number of nodes (proteins) and edges (interactions) as inBio Map. Furthermore, it has been shown that biological networks belong to a class of networks called scale-free networks and that their node degree distribution follow a power-law². Therefore, it is important that the random networks also display this property and in particular, that they follow the same node degree distribution as inBio Map.

Random networks can be generated either by re-wiring the existing network (by pairwise edge swapping) or by building a random network from scratch (*de novo* network generation). However, with the requirement of a conserved node degree distribution, these methods become prohibitively slow (generating 1,000,000 random networks would take months-years even on a multicore supercomputer). Previous approaches to randomization of the complete protein-protein interactome have overcome this limitation by re-wiring the existing network but only swapping edges between certain subsets of nodes in the network (e.g. nodes with the same node degree)³.

We developed an algorithm that creates *de novo* random networks with a nearly (99.97 %) conserved node degree distribution of inBio Map. Since the networks are randomly generated from scratch, they will be completely independent of the original PPI network but they will still maintain the required properties.

Our algorithm exploits the fact that a random network with a conserved node degree distribution can be generated relatively fast, if the network is allowed to contain multi-edges and self-edges (loops). The multi-edges and loops are subsequently removed and the lack of edges is accounted for by generating a new network filling in the missing edges. This procedure is repeated a number of times until the random network is sufficiently similar to the original network.

The algorithm was implemented using the igraph package⁴ for Python (<https://igraph.org/python/>) and is outlined below:

Iterative refinement of *de novo* generated random networks:

1. Calculate the node degree distribution of the original network (N_O)
2. Generate a random network (N_A) with the same node degree distribution as N_O
3. Remove multi-edges and loops from N_A
4. Iterate until the stop criterion is met:
 - a. Calculate the difference between the node degree distribution of the random network (N_A) and the original network (N_O)
 - b. Generate a new random network (N_B) with node degree distribution equal to the difference between N_A and N_O just calculated
 - c. Merge the two random networks N_A and N_B
 - d. Remove multi-edges and loops from the merged network – this is now the new random network (N_A)

After each iteration of the algorithm, the distance between the original network (N_O) and the random network (N_A) was evaluated. The distance was defined as the sum of all pairwise absolute differences in node degree between all nodes (n) in the original network (N_O) and the random network (N_A):

$$distance = \sum_n | degree(n_A) - degree(n_O) |$$

The stop criterion was defined as “the distance being less than or equal to 100” or if “the distance did not change for 5 consecutive iterations”.

The algorithm was applied to the high confidence subset of the PPI network inBio Map (version 2018_04_05), defined by the edges with a confidence score ≥ 0.117 . Unspecific post-translational modifications (PTMs), such as the attachment of ubiquitin to target proteins, are often picked up by the methods used to detect protein-protein interactions. To avoid a bias from PTMs in our analysis a few proteins often associated with PTMs were left out from the analysis (UBB, UBC, UBA52, RPS27A, SUMO1-4). The final PPI network used in the analysis contained 13,033 proteins (nodes) and 147,882 high confidence interactions (edges). The total permutation time, generating 1,000,000 random networks using 20 CPU cores, was 6.7 hours.

The average distance (as defined above) between the random networks and inBio Map was 83.1 with a standard deviation of 10.6. This means that the node degree distribution of the random networks was highly conserved and on average 99.97% identical to the node degree distribution of inBio Map.

Significance of shortest paths

The purpose of generating random networks was to evaluate the significance of the shortest paths between the cis-gene intermediaries and the plasma proteins for each of the trans-pQTL signals. Thus, for each potential cis-gene intermediary and the plasma protein, the shortest path was also calculated in each of the 1,000,000 random networks and the significance of the observed shortest

path was then calculated as the fraction of random networks with a shortest path, shorter than or equal to the observed shortest path in inBio Map. As multiple paths were tested for each trans-pQTL signal, the Benjamini–Hochberg procedure was used to correct for multiple testing, yielding FDR controlled *q*-values.

Calculation of the shortest paths was actually more time-consuming than the permutation of the network, and the running time of the whole PPI network analysis was 37.3 hours using 28 CPU cores.

The results of the PPI network analysis are available in Supplementary Table 3.

Text mining association network

The inBio Map PPI network consists of experimentally observed physical interactions between proteins. Some interactions between proteins (e.g. protein phosphorylation by a kinase) are very short-lived and may not be picked up by some of the methods (especially co-purification based methods) used to detect physical interactions between proteins but they might still be relevant. Some of these interactions are described in the scientific literature and can be picked up by text mining.

Therefore, as an alternative to the PPI network, we also created a network of statistically significant associations between proteins identified by text mining (TM network). We used the text mining resource inBio Know™ (<https://www.intomics.com/technology/#text-mining>) to identify proteins co-mentioned in PubMed abstracts. inBio Know is based on highly curated protein synonyms for the human proteome and provides a statistical framework to identify proteins that are significantly associated in the scientific literature. Briefly, based on the manually curated protein synonyms, the occurrence of each protein from Uniprot was determined in PubMed abstracts (PubMed download date: 2018_05_03). Only human proteins with the status “reviewed” in Uniprot were considered. The pairwise co-occurrence between proteins was determined and Fisher’s exact test was used to calculate the significance of the association. A conservative, multiple-testing corrected *p*-value cutoff of 10^{-8} was applied to obtain significant associations between proteins.

A TM network was then created by connecting proteins that were significantly associated in PubMed abstracts. The TM network contained 14,635 proteins (nodes) and 193,777 associations (edges).

The algorithm described for the PPI network above was applied to the TM network in order to find significant paths between the potential cis-gene intermediaries and the plasma proteins in the text mining associations. Using the same stop criterion as above the total running time for the analysis was 87.9 hours using 28 CPU cores. Due to the topology of the TM network, the algorithm did not converge as fast as for the PPI network. This was also reflected in the average distance of the random networks, which was 739.0 with a standard deviation of 41.7. The distance could be reduced at the expense of longer running times but since the degree distribution of the random networks was still 99.81% identical to the degree distribution of the original TM network, we did not find that necessary.

Finally, we created a combined PPI and TM network consisting of both the protein-protein interactions and the associations identified by text mining simply by merging the two networks

(PPI+TM network). This network contained 16,150 proteins (nodes) and 314,503 interactions (edges). The running time of the network analysis was 126.9 hours using 28 CPU cores. The average distance of the random networks was 370.3 with a standard deviation of 30.5. The degree distribution of the random networks was 99.94% identical with the degree distribution of the original PPI+TM network.

The results of the TM network analysis and the PPI+TM network analysis are both available in Supplementary Table 3.

References

- 1: Li, T. *et al.* A scored human protein-protein interaction network to catalyze genomic interpretation. *Nature Methods* **14**, 61-4 (2017).
- 2: Albert R., Jeong H. & Barabási A. Error and attack tolerance of complex networks. *Nature* **406**, 378-82 (2000).
- 3: Rossin E. J. *et al.* Proteins Encoded in Genomic Regions Associated with Immune-Mediated Disease Physically Interact and Suggest Underlying Biology. *PLoS Genet* **7**, e1001273 (2011).
- 4: Csárdi G. & Nepusz T. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695 (2006).