Supplementary Material

Molecular triangulation (MT) algorithm

The modified MT method was applied to identify candidate genes associated with phenotype. The permutation-based p-values were combined across phenotypes for each gene and were used to rank genes in the network that included both seed and non-seed genes. The MT method adopts two different scores (*trn* and *cnt*) and two different background models (*ini* and *rwr*).

The first score was calculated as follows: $S^{trn}(I,e;G=\langle N,E\rangle)=\sum_{v\in I} e(v)\frac{1}{1+d_{uv}}$, (1)

Where: $G=\langle N,E\rangle$ =undirected molecular network

N=set of genes and *E* is a set of their interactions

I= set of seed genes

$e(v)$= primary evidence i.e. -log of p-value (<5x10^{-8}) for all $v\in I$, these values provide confidence in the decision to include given gene as a seed node

d_{uv} = length of the shortest path between genes u and v within network G

The second score was measured by: $S^{cnt}(Ti,I,e;G=\langle N,E\rangle)$ (2)

Where,

$T=\{T_1,\ldots,T_M\}, T_i\subset N$ = predefined set of sets of the network genes

Second score is the number of edges in *E* that connect a node from T_i with a node from *I*. The method generates random initials sets I_j in *ini* background model or random networks E_j in the case of the *rwr* model. Later, method computes the score of the given initial set (real score) and its network and the background score for each of the background replicates for every test set T_i. We performed 1000 iterations by shuffling the edges to test for the significance of real

score. The smoothed p-value was calculated as explained by Iossifov *et al* (1). Shortest path length between the genes encoding proteins was calculated using the transcriptional, proteomic and metabolic interaction networks.

The first background model (*ini*) assumes that the set of initial nodes I is sampled uniformly from the network nodes in N. The second background model (*rwr*) assumes instead that the network edges E are attached to nodes using a random rewiring process so that every gene preserves its observed degree (2). Once a score function S and a background model are chosen, we can generate *B replicates* according to the background model (*B* random initial sets I^j in the case of the *ini* background model or *B* random networks E^j in the case of the *rwr* model). We compute the score for the given initial set and the given network $s_i^{real} = S(T_i, I, e; \mathrm{G} = \langle N, E \rangle)$ and the background score s_i^j for each of the background replicates ($s_i^j = S(T_i, I^j, e; \mathrm{G} = \langle N, E \rangle)$ for *ini* and $s_i^j = S(T_i, I, e; \mathrm{G} = \langle N, E^j \rangle)$ for *rwr*) for every test set T_i. We can then assign a *p*-value for each T_i based on how many of the background scores are higher or equal to the real score:

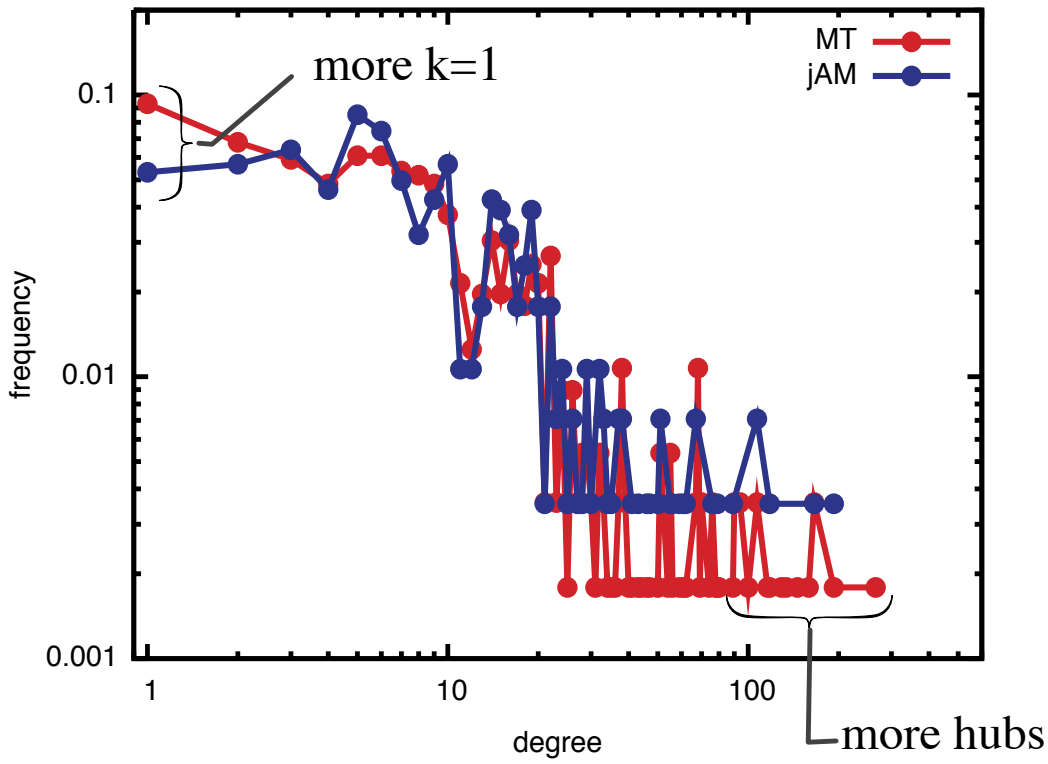
$$\hat{p}_i = \frac{|\{s_i^j : s_i^j \geq s_i^{real}\}|}{B}.$$

If all of the background scores are lower than the real score, the above definition will assign a *p*-value of 0 (1).

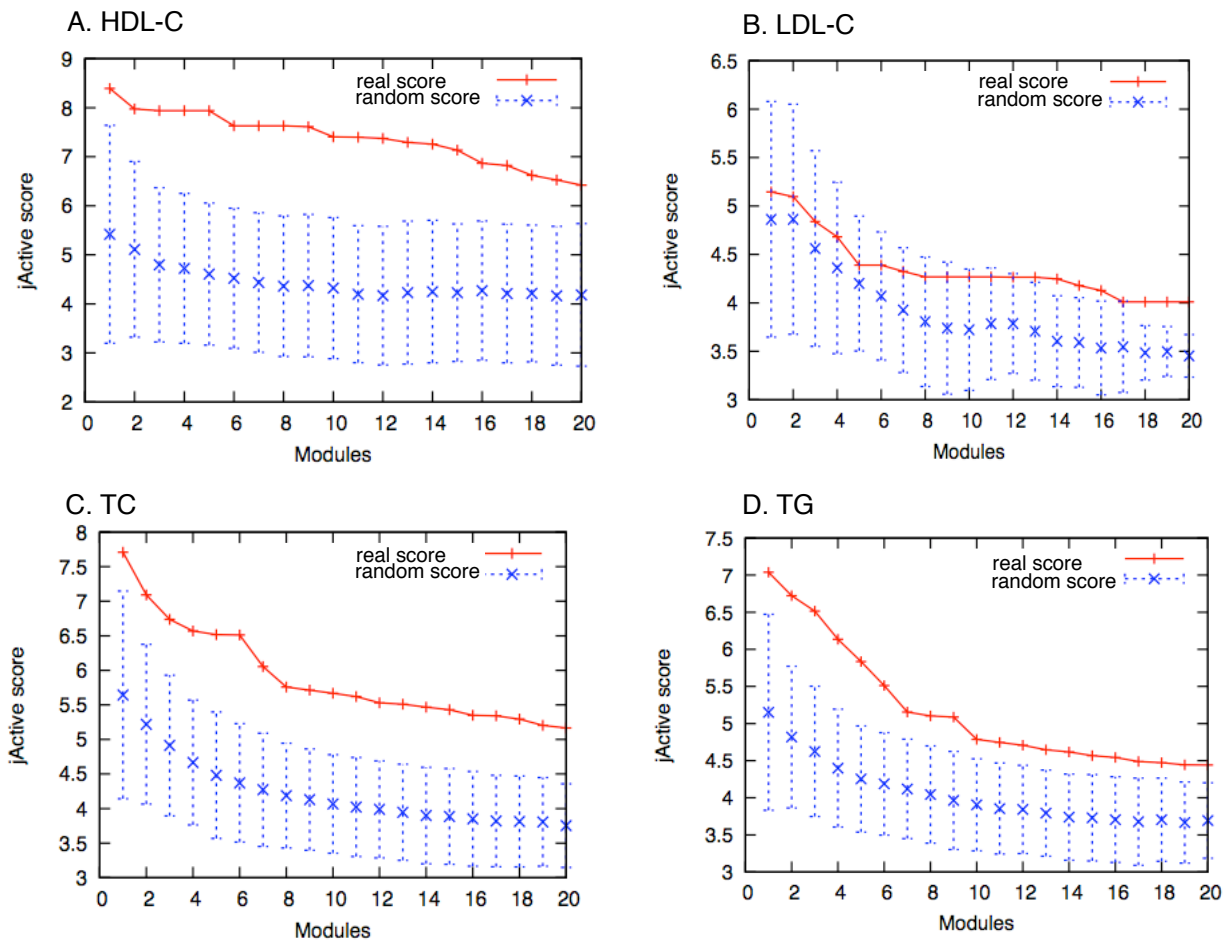
MT uses a null model in which the GWAS signals are uninformative (not linked to the phenotype) and any observed gene clustering within a molecular network is accidental. The hypothesis competing with the null assumes that the GWAS signal is associated with a group of genes within the molecular network.

References:

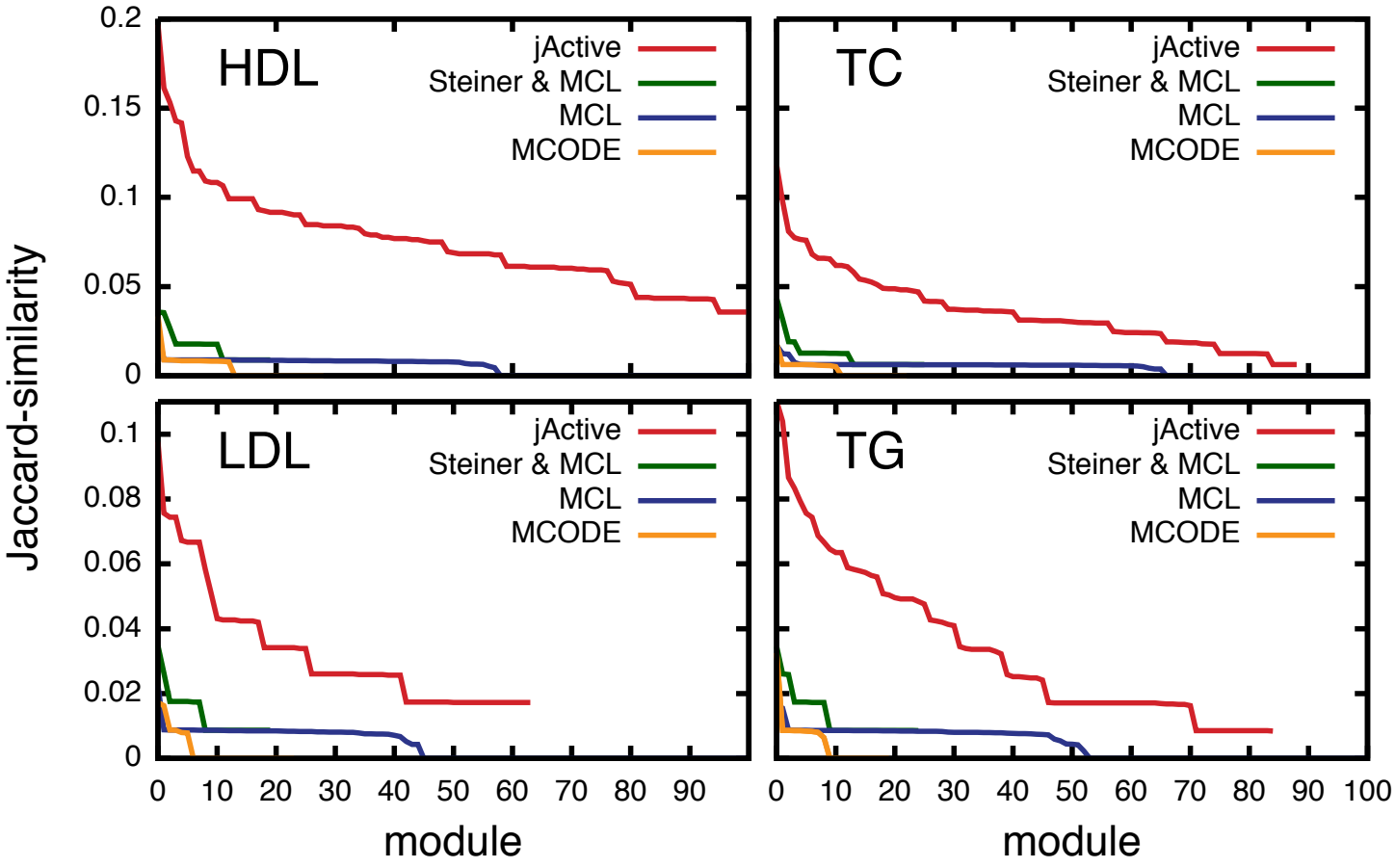
1. Iossifov, I., Rodriguez-Esteban, R., Mayzus, I., Millen, K. J., and Rzhetsky, A. (2009) Looking at cerebellar malformations through text-mined interactomes of mice and humans. *PLoS Comput Biol* 5, e1000559.
2. Feldman, I., Rzhetsky, A., and Vitkup, D. (2008) Network properties of genes harboring inherited disease mutations. *Proceedings of the National Academy of Sciences of the United States of America* 105, 4323-4328.



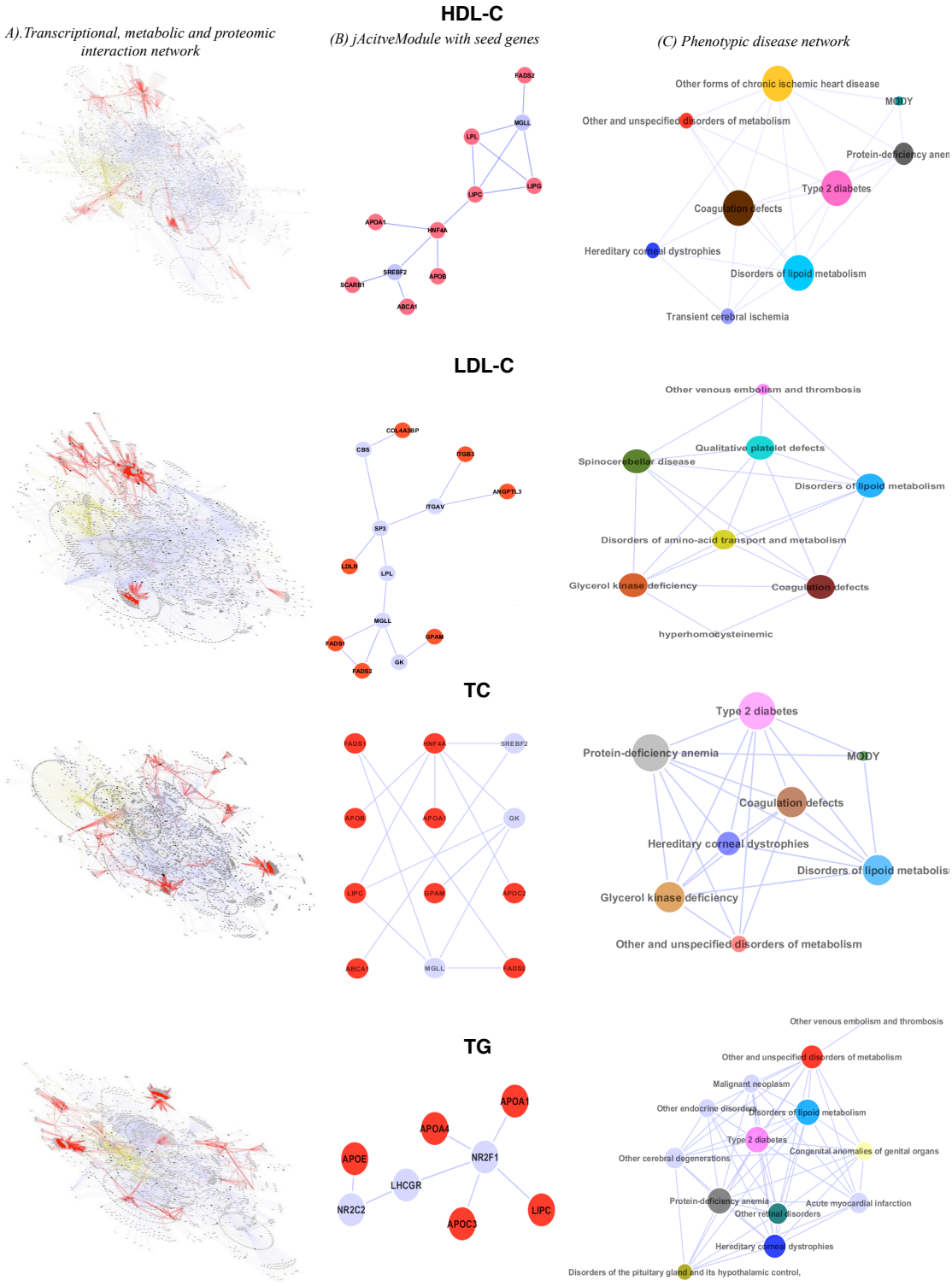
Supplementary fig. S1: Degree distributions of the genes unique to MT and jAM for the four traits.



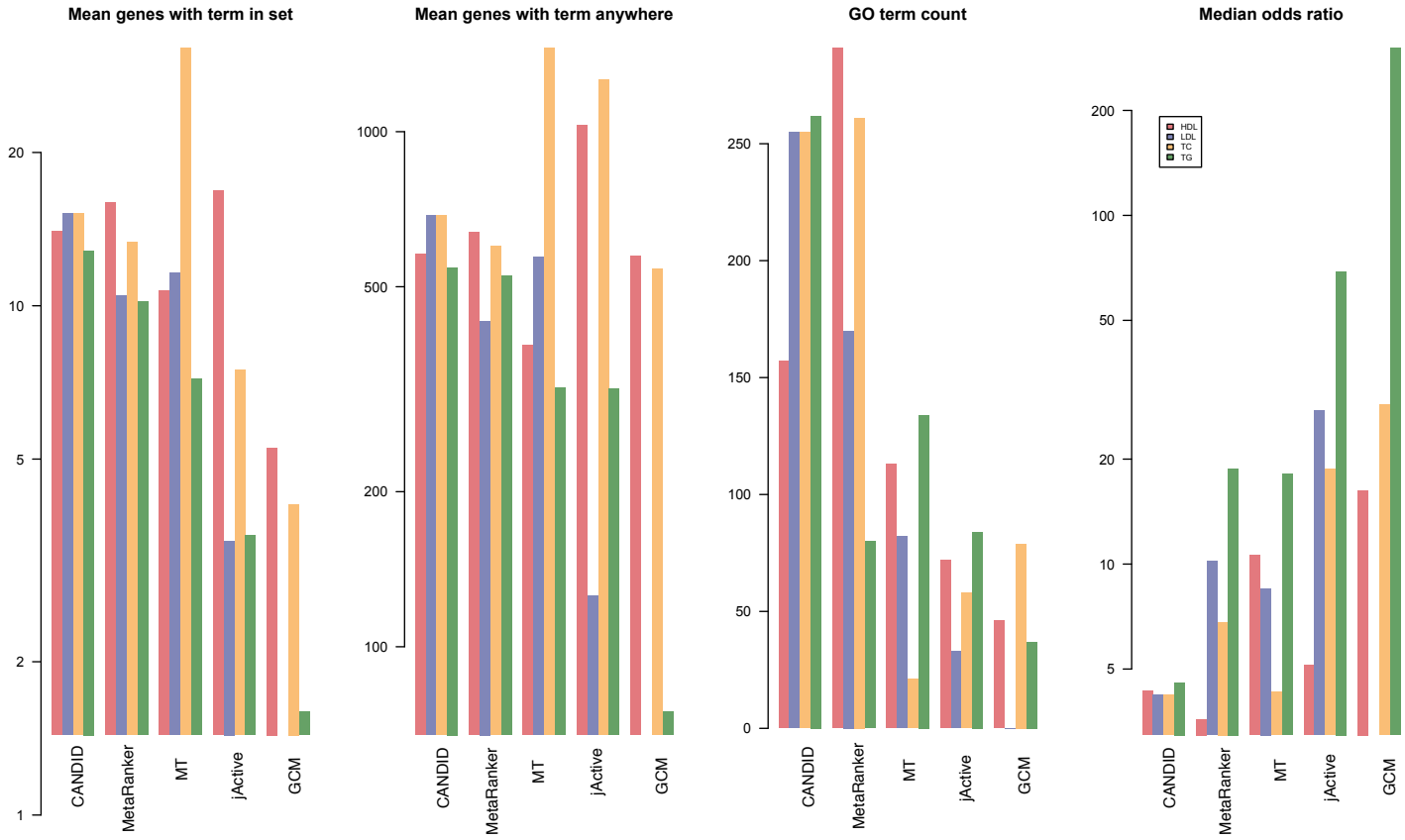
Supplementary Fig. S2: The average *jActive* scores and standard deviations obtained for HDL-C (**A**), LDL-C (**B**), TC (**C**) and TG (**D**) from top twenty modules and for 100 random genes. We identified modules for four traits by applying the criteria of nodes >3 and <50 having *jActive* score >3.



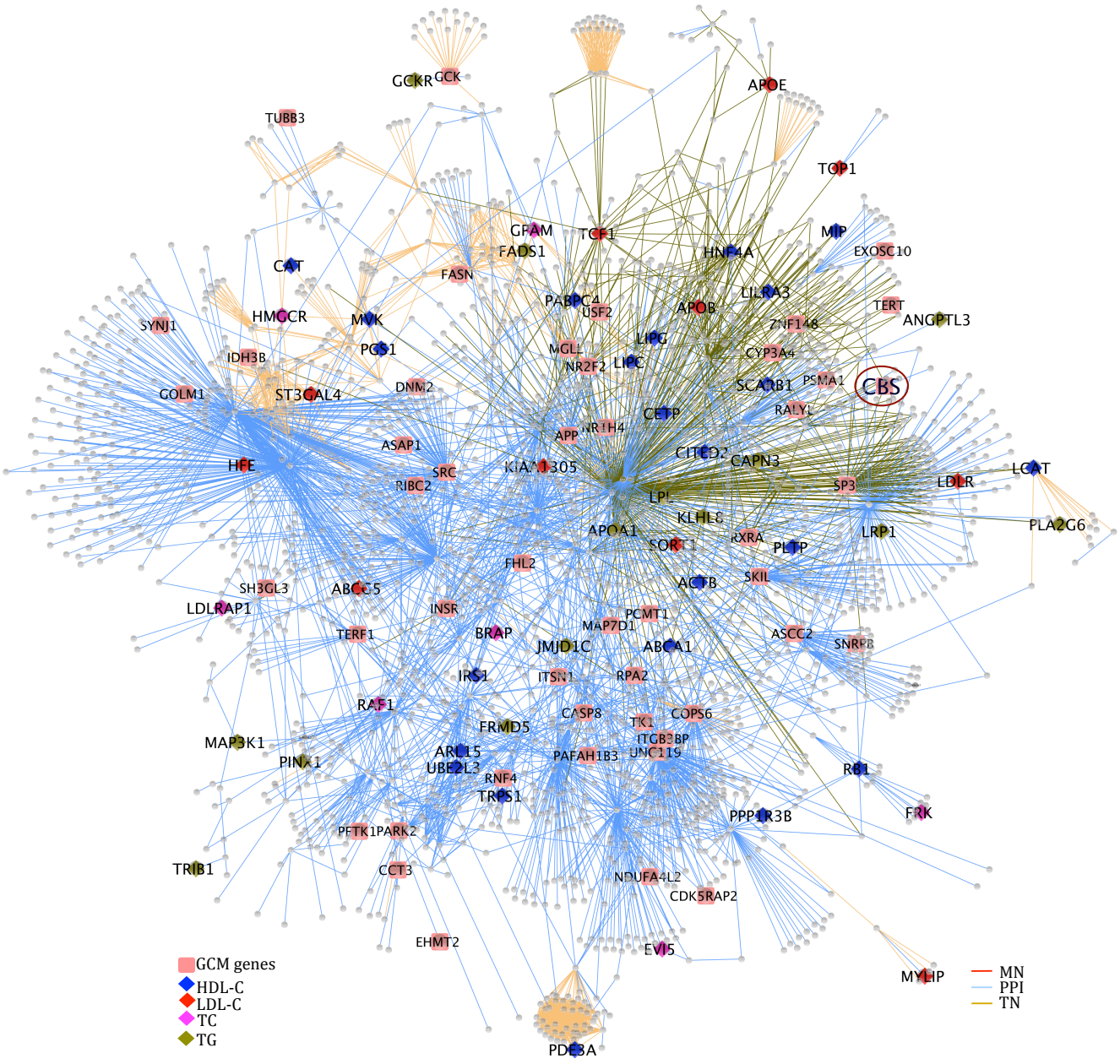
Supplementary Fig.S3: Comparison of *jActiveModule* with Steiner-MCL approach for seed genes coverage in interactome.



Supplementary Fig.S4: Depiction of comorbidity associations between diseases in a module. For each trait interactome (**A**), the unique ICD9 codes for genes in a module (**B**) were converted to construct a phenotypic disease network (**C**). Seed nodes are coloured as red in B. The sizes of the disease nodes in **C** are based on their connectivity with other diseases in a module.



Supplementary Fig.S5: GO term enrichment tests of candidate genes priortized by MT, *jActiveModule*, GCM, CANDID and MetaRanker.



Supplementary Fig.S6: Seed and GCM genes for the four lipoprotein traits. GCM genes are rectangular and the seed genes are indicated as diamonds. The cystathionine beta-synthase gene (*CBS*) with a SNP that we found significantly associated in the MDC-CC cohort is circled in the interacome. All the metabolic interactions (MN) are indicated as orange lines, protein-protein interactions (PPI) as blue lines and transcription networks (TN) as yellow lines.

Supplementary table S1: Number of genes filtered in each step applied in our approach for selecting additional candidate genes for lipoprotein traits

Trait	MT genes	MT-p-values	MT-jActive genes (jAM)	GCM genes	co-GCM genes	Co-GCM genes with p<0.05 in GLGC-GWAS
HDL-C	192	5.4E-5	107	45	40	19
LDL-C	142	4.7E-5	42	20	19	9
TC	207	6.5E-5	84	27	25	13
TG	119	4.4E-5	45	25	20	12

Supplementary table S2: Comparison of commonly predicted genes between CANDID, MetaRanker and the three steps in our approach (MT, *jActiveModule*, comorbidity analysis).

	HDL- MT	p-value (Fischer' s test)	LDL-MT	p-value (Fischer' s test)	TC-MT	p-value (Fischer' s test)	TG-MT	p-value (Fischer' s test)
CANDID	3.3%	0.022	5.4%	<0.0001	7.5 %	<0.0001	2.7%	0.29
MetaRanker	2.7 %	0.099	3.4%	0.016	7.2%	<0.0001	4.5%	0.005

	HDL- jActive	p-value (Fischer' s test)	LDL-jActive	p-value (Fischer' s test)	TC-Jactive	p-value (Fischer' s test)	TG-jActive	p-value (Fischer' s test)
CANDID	3.1%	0.22	5.4%	0.1	12.3 %	<0.0001	5.3%	0.1
MetaRanker	4.2 %	0.03	8.1%	0.001	12.3%	<0.0001	10.5%	<0.0001

	HDL- GCM	p-value (Fischer' s test)	LDL- GCM	p-value (Fischer' s test)	TC-GCM	p-value (Fischer' s test)	TG- GCM	p-value (Fischer' s test)
CANDID	7%	0.0048	5.3%	0.55	15.6 %	<0.0001	10%	0.0087
MetaRanker	4.7 %	0.158	10.5%	0.006	15.6%	<0.0001	15%	<0.0001

	HDL- CANDID	p-value (Fischer' s test)	LDL-CANDID	p-value (Fischer' s test)	TC-CANDID	p-value (Fischer' s test)	TG-CANDID	p-value (Fischer' s test)
MetRanker	4.5%	<0.0001	9%	<0.0001	6%	<0.0001	3%	0.004

Supplementary table S3: Literature mining results for GCM genes. Gene order is based on descending p-values in GLGC GWAS data for the four traits. The table consist the description of candidate genes those found to be related to lipid metabolism.

Gene	Trait	Description	Reference
APP	HDL		
RXRA	HDL	RXRA variants rs11185660 (P = 0.0021) has be reported to be associated with susceptibility to low HDL-C and CHD	Peloso et al. J Lipid Res.51:3524-32 (2010).
FASN	HDL	Found to be linked to visceral fat accumulation, impaired insulin sensitivity and increased lipetin and RBF4 suggesting its role in lipogenic pathway	Diabetologia. 50, 1472-80 (2007)
INSR	HDL	n-3 fatty acids diets diminish arterial LDL-cholesterol deposition in mice with insulin resistance (Insr heterozygous knockout mice), and this is associated with changes in arterial LpL levels and distribution	Chang et al. Arterioscler Thromb Vasc Biol. 30:2510-7 (2010).
CYP3A4	HDL	The CYP3A4 enzymes generate major oxysterols that enter the circulation. The oxysterols activate-via nuclear receptors-ATP-binding cassette (ABC) A1 and other genes, leading to the elimination of excess cholesterol and protecting arteries from atherosclerosis.	Luoma PV. Eur J Clin Pharmacol. 64:841-50 (2008).
VASP	HDL	VASP phosphorylation flow cytometric assessment has been reported as a tool to evaluate the responsiveness to clopidogrel in coronary heart disease (CHD) patients	Hezard et al. Platelets. 16;474-481 (2005)
SMURF2	HDL	Smurf2 as regulators of TGF-beta signaling: new targets for managing myofibroblast function and cardiac fibrosis.	Cunnington et al. Can. J. Physiol. Pharmacol. 87:764-772(2009)
PSMA1	HDL		
PCMT1	HDL		
DVL3	HDL	Dvl3 is required for cardiac outflow tract	Etheridge et al. PLoS Genet. 4:e1000259(2008)
FHL2	HDL	Deletion of the FHL2 gene attenuates the formation of atherosclerotic lesions after a cholesterol-enriched diet.	Chu et al. Life Sci. 86:365-71 (2010).

CASP8	HDL		
ASCC2	HDL		
TERT	HDL	Telomerase reverse transcriptase promotes cardiac muscle cell proliferation	Oh et al. Proc Natl Acad Sci U S A. 98:10308-13 (2001)
RNF4	HDL		
SKIL	HDL		
RALYL	HDL		
PFTK1	HDL		
EHMT2	HDL		
SH3GL2	LDL	Intima-media thickness (IMT) of the carotid arteries, as measured by B-mode ultrasonography, is a quantitative trait that strongly predicts CVD and is increasingly used in clinical decision-making. A SNP-**rs2593404 near SH3GL2** was found to be associated with IMT. But, this was not signficant after the correction.	Lanktree et al. Stroke. 40:3173-9 (2009)
SH3GL3	LDL	A genome-wide linkage analysis to identify QTLs for plasma HDL-C levels in a well-characterized U.S. cohort consisting of multiplex families (GeneQuest). Fine mapping of the 15q25 QTL, we studied D15S983 and two SNPs, **rs1491579 (SH3GL3)** and rs1638634 adjacent to marker D15S655 with the highest LOD score.	Yang et al. J Lipid Res. 51:1442-51 (2010).
UNC119	LDL		
IDH3B	LDL		
RPA2	LDL		
CDK5RAP2	LDL		
ITGB3BP	LDL		
NDUFA4L2	LDL		
SH3GL2	TC		

CBS	TC	Hepatic steatosis in CBS(-/-) mice is caused by or associated with abnormal lipid metabolism	Namekata et al. J Biol Chem. 279:52961-9 (2004).
USP33	TC		
DCP2	TC		
EXOSC10	TC		
PSMA7	TC		
ZNF148	TC	Nuclear proteins capable of binding to the MMP3 promoter is transcription factor ZBP89 (also named ZNF148) which acts as a transcriptional enhancer	Cardiovasc Res. 69,636-45 (2006)
SYNJ1	TC		
PARK2	TC		
SNRPB	TC		

Supplementary table S4: Analysis of a combined associated effect of the four SNPs genotyped in MDC-CC cohort (Summing the number of risk alleles for each individual).

COMBINEDRISK_FOURSNPS_RISK_ALLELE COUNT: INSR.NDUFA4/2. CBS. DNM2		Cholesterol (mmol/l)	Low density lipoprotein (mmol/l)	High density lipoprotein (mmol/l)	Triglycerides (mmol/l)
0	Mean	6.0889	4.0537	1.417	1.3919
	N	305	298	304	306
	Std. Deviation	1.10461	1.00802	0.37605	1.1281
1	Mean	6.1709	4.1605	1.4029	1.3182
	N	1106	1081	1094	1108
	Std. Deviation	1.16989	1.01722	0.39618	0.73698
2	Mean	6.17	4.1702	1.3668	1.3952
	N	1559	1518	1540	1556
	Std. Deviation	1.07246	0.96281	0.36709	0.77564
3	Mean	6.1376	4.1486	1.392	1.3274
	N	1163	1141	1155	1164
	Std. Deviation	1.04363	0.96425	0.37188	0.76794
4	Mean	6.2178	4.2033	1.3839	1.4091
	N	466	448	456	465
	Std. Deviation	1.02687	0.94645	0.36046	0.81499
5	Mean	6.2491	4.1964	1.3264	1.4713
	N	115	111	113	115
	Std. Deviation	1.23936	1.05123	0.34472	0.78866
6	Mean	6.6592	4.65	1.1817	1.815
	N	12	12	12	12
	Std. Deviation	1.16357	1.03089	0.23417	0.60157
Total	Mean	6.1649	4.1601	1.385	1.3645
	N	4726	4609	4674	4726
	Std. Deviation	1.09147	0.98013	0.37483	0.79738
		P (linear regression adjusting for age and sex):	**P (linear regression adjusting for age and sex):**	**P (linear regression adjusting for age and sex):**	**P (linear regression adjusting for age and sex):**
		0.101	**0.057**	**0.041**	**0.026**