MATERIALS AND METHODS

Study Population

We used the SNP level association p-value results from all 14 of the discovery GWAS analyzed by the transatlantic Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) consortium as well as two GWAS conducted by the Ottawa Heart Institute.[1-3] A summary of the core phenotypic details for these GWAS is presented in **Table 1**. We divided the 16 GWAS into 2 sets: a Stage 1 Discovery Set and a Stage 2 Replication Set. Further details on each of the cohorts included are provided in the Methods section in the online-only Data Supplement.

Genotyping and Quality Control

Details on genotyping and quality control procedures for all GWAS datasets included in this analysis have been previously reported.[1, 3] All datasets were restricted to subjects of white/European descent either through self-report, principal components analysis, or a combination of the two. Prior to performing association testing, all CARDIoGRAM genome-wide datasets were imputed up to ~2.5 millions SNPs using HapMap2 release 22 (build 36) white (CEU; Center d'Etude du Polymorphisme Humain) sample as the reference panel. The SNP call rate filter used on genotyped SNPs pre-imputation was >0.90 and for a majority >0.95. These SNPs were then used to impute approximately 2.5 million HapMap SNPs. Imputed SNPs were then excluded based on missing frequency in cases or controls > 0.02 (Missing), minor allele frequency in cases (MAF cases) or controls (MAF controls) < 0.01, quality of the imputation (INFO) < 0.5, and deviation from Hardy-Weinberg equilibrium in controls (HWE) $p < 0.0001$. One further filter that was applied to the discovery GWAS used in this analysis was the removal of any imputed SNP with an overall SNP call rate of < 0.75. The call rate was applied to imputed SNPs by only counting an imputed call if the posterior probability of one of the three possible genotypes was >90%, i.e. imputed with a high level of certainty. We also removed any SNP that was present in 2 or less of the Stage 1 studies and any SNP that was present in 4 or less of the Stage 2 studies.

Stage 1 and Stage 2 GWAS Meta Analysis

We conducted the meta-analysis of the two sets of GWAS using an analytic approach that was very similar to that used previously by the CARDIoGRAM consortium.[2] . We then proceeded with a fixed-effects inverse-variance-weighted meta-analysis together with a Q- and I- measure of homogeneity. Any SNPs that were significantly heterogeneous based on Q and I statistics (at $p<0.001$) were analyzed using random effects.[4]

Gene Set Enrichment Analysis

Pathway information for gene set enrichment analysis was obtained from the Reactome gene sets available in the Molecular Signatures Database v3.1 (MSigDB).[5, 6] Although a wide choice of pathway databases exist, we selected the expert-authored and manually curated Reactome database due to its transparent structural hierarchy, high generalizability and internally consistent 'reaction-based' data model encompassing a wide variety of biological processes. A total of 639 Reactome pathways were utilized for pathway enrichment analysis, after removing 35 pathways with <10 or >200 gene members[7-9].

For SNP-to-gene mapping, a SNP, Si, was first mapped to gene, Gj (j=1,….N) if Si was located within the primary transcript of the gene or a window of 100 kilobases on either end of the gene (additional details in **Figure S1** in the online-only Data Supplement). Associations between pathways and CAD were examined through a gene-set enrichment analysis (GSEA) procedure[5, 10] via the iGSEA4GWAS tool (Improved Gene Set Enrichment Analysis for Genome-Wide Association Study at http://gsea4gwas.psych.ac.cn/inputPage.jsp).[11] iGSEA4GWAS examines the enrichment of significantly associated variants within or near *a priori* defined gene sets by determining if a particular gene-set ranks higher than a randomly distributed set, based on a running-sum statistic on the ranked list of genes (ranked by association p-values or an equivalent statistic) **(Figure 1)** (additional details in Methods section of online-only Data Supplement). The 'improvement' in iGSEA4GWAS over traditional GSEA approaches is realized by focusing on gene sets with high proportions of significant genes instead of relying solely on the overall gene set significance that may sometimes originate from only a few genes. Pathways achieving a permutation-based nominal p-value of ≤0.05 (at <25% false discovery rate, FDR) in the Stage 1 discovery studies were taken forward for replication in a meta-analysis of the Stage 2 studies. Pathways were *a priori* defined as replicated if they also achieved a p-value of <0.05 (corresponding to FDR<12.5%) in the Stage 2 set of GWAS. Additional analysis was also conducted to test for the effects of linkage disequilibrium (LD) patterns among SNPs and among pathway genes on GSEA results (Methods section of online-only Data Supplement).

Bioinformatic Analysis to Prioritize Genes in Replicated Pathways

After identifying the replicated pathways, we sought to recover higher level functional interactions between the pathways, as well as between genes within a pathway. To accomplish this, we mapped the

genes from the replicated pathways onto well-curated interaction networks, and assessing the networks for (i) the probability that such networks can arise by chance (ii) the presence of biologically relevant clustering of genes within the network, and (iii) the relative contributions of pathway genes on the topology of pathway networks.

(i) Statistical evaluation of networks: To statistically evaluate the degree to which networks derived from the query genes could arise at random, we first mapped a total of 770 candidate genes (derived from the GWAS replicated pathways) to the highly curated and high confidence InWeb protein-protein interaction network (PPI)[12]. Next, we created random networks via 1000 rounds of within-degree, node-label permutation of the InWeb PPI network, and compared parameters of network connectivity (node-degree and edge number) in sub-networks arising from the candidate genes in the original vs. the random networks. This analysis was conducted in the Disease Association Protein-Protein Link Evaluator (DAPPLE) software environment[13].

(ii) Mapping of replicated pathway genes to functionally interacting networks: The 770 genes from the replicated pathways were next mapped to a functional interaction network obtained from Reactome (ReactomeFI, 2012 version), and visualized in Cytoscape (v 2.8.2) (Methods section of online-only Data Supplement). The mapped interactome network was subjected to spectral partition clustering to identify internal modular sub-structures.[14] We subsequently tested the resulting sub-networks for overrepresentation of biological processes via Gene Ontology-Biological Process terms.

(iii) Analysis of network topologies: Lastly, we converted the replicated Reactome pathways into functional network modules and analyzed their network centrality properties, via the Centiscape tool[15]. Assuming that the critical functions of a network are largely governed by central nodes which connect several different neighborhoods of the network, we assessed the relative importance of networks and their constituent nodes (genes) by the centrality measures of 'degree' and 'betweenness' (Methods section , online-only Data Supplement). [16] Although other network centrality descriptors exist, 'degree' and 'betweenness' have been proposed as key correlates of gene and protein function in biological networks. Of them, 'betweenness' has been proposed to the more relevant metric when studying network dysfunction in disease (genes with high 'degree' are usually essential for life and therefore may not be investigatable in the context of disease)[17-19] More specifically, recent data further show that genes with intermediate connectivity (betweenness) have the highest probability of harboring germ-line

disease mutations (compared to essential genes) and correlate with pleiotropy,[20] and crosstalk between functional modules.[21]

Reference List

(1) Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de FU, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoep JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011;43:25-33.

(2) Preuss M, Konig IR, Thompson JR, Erdmann J, Absher D, Assimes TL, Blankenberg S, Boerwinkle E, Chen L, Cupples LA, Hall AS, Halperin E, Hengstenberg C, Holm H, Laaksonen R, Li M, Marz W, McPherson R, Musunuru K, Nelson CP, Burnett MS, Epstein SE, O'Donnell CJ, Quertermous T, Rader DJ, Roberts R, Schillert A, Stefansson K, Stewart AF, Thorleifsson G, Thorsteinsdottir U, Voight BF, Wells GA, Ziegler A, Kathiresan S, Reilly MP, Samani NJ, Schunkert H. Design of the Coronary Artery Disease Genome-Wide Replication and Meta-Analysis (CARDIoGRAM) Study -- A Genome-Wide Association Meta-Analysis Involving More than 22,000 Cases and 60,000 Controls. *Circ Cardiovasc Genet* 2010;3:475-483.

(3) Davies RW, Wells GA, Stewart AF, Erdmann J, Shah SH, Ferguson JF, Hall AS, Anand SS, Burnett MS, Epstein SE, Dandona S, Chen L, Nahrstaedt J, Loley C, Konig IR, Kraus WE, Granger CB, Engert JC, Hengstenberg C, Wichmann HE, Schreiber S, Tang WH, Ellis SG, Rader DJ, Hazen SL, Reilly MP, Samani NJ, Schunkert H, Roberts R, McPherson R. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet* 2012;5:217-25.

(4) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.

(5) Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.

(6) Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011;39:D691-D697.

(7) Pemu PE, Anderson L, Gee BE, Ofili EO, Ghosh S. Early alterations of the immune transcriptome in cultured progenitor cells from obese African-American women. *Obesity (Silver Spring)* 2012;20:1481-90.

(8) Ghosh S, Dent R, Harper ME, Gorman SA, Stuart JS, McPherson R. Gene expression profiling in whole blood identifies distinct biological pathways associated with obesity. *BMC Med Genomics* 2010;3:56.

(9) Ghosh S, Vivar JC, Sarzynski MA, Sung YJ, Timmons JA, Bouchard C, Rankinen T. Integrative pathway analysis of a genome-wide association study of (V)O(2max) response to exercise training. *J Appl Physiol (1985)* 2013;115:1343-59.

(10) Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 2007;81:1278-83.

(11) Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* 2010;38:W90-W95.

(12) Lage K, Karlberg EO, Storling ZM, Olason PI, Pedersen AG, Rigina O, Hinsby AM, Tumer Z, Pociot F, Tommerup N, Moreau Y, Brunak S. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* 2007;25:309-16.

(13) Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, Cotsapas C, Daly MJ. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet* 2011;7:e1001273.

(14) Newman ME. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 2006;103:8577-82.

(15) Scardoni G, Petterlini M, Laudanna C. Analyzing biological network parameters with CentiScaPe. *Bioinformatics* 2009;25:2857-9.

(16) Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. *Genes Dev* 2007;21:1010-24.

(17) Feldman I, Rzhetsky A, Vitkup D. Network properties of genes harboring inherited disease mutations. *Proc Natl Acad Sci U S A* 2008;105:4323-8.

(18) Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol* 2005;22:803-6.

(19) Joy MP, Brock A, Ingber DE, Huang S. High-betweenness proteins in the yeast protein interaction network. *J Biomed Biotechnol* 2005;2005:96-103.

(20) Zou L, Sriswasdi S, Ross B, Missiuro PV, Liu J, Ge H. Systematic analysis of pleiotropy in C. elegans early embryogenesis. *PLoS Comput Biol* 2008;4:e1000003.

(21) Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 2007;3:e59.