A network of pleiotropic genes for metabolic syndrome and inflammation

Aldi T. Kraja*[†1], Josée Dupuis*[2,3], Daniel I. Chasman*[4], Kari E. North*[5], Alexander P. Reiner*[6], Lisa R. Yanek*[7], Tuomas O. Kilpeläinen*[8], Jennifer A. Smith*[9], Abbas Dehghan*[10], Andrew D. Johnson[11], Mary F. Feitosa[1], Fasil Tekola-Ayele[12], Audrey Y. Chu[4], Ilja M. Nolte[13], Zari Dastani[14], Andrew Morris[15], Sarah A. Pendergrass[16], Yan V. Sun[17], Marylyn D. Ritchie[18], Ahmad Vaez[13], Honghuang Lin[19], Symen Ligthart[10], Letizia Marullo[15,20], Rebecca Rohde[5], Yaming Shao[5], Mark A. Ziegler[21], Hae Kyung Im[22], Renate B. Schnabel[23], Torben Jørgensen[24], Marit E. Jørgensen[25], Torben Hansen[8], Oluf Pedersen[8], Ronald P. Stolk[13], Harold Snieder[13], Albert Hofman[10], Andre G. Uitterlinden[26], Oscar H. Franco[10], M. Arfan Ikram[10], J. Brent Richards[14,27,28], Charles Rotimi[29], James G. Wilson[30], Leslie Lange[31], Santhi K. Ganesh[32], Michael A. Nalls[33], Laura J. Rasmussen-Torvik[34], James S. Pankow[35], Joseph Coresh[36], Weihong Tang[35], W.H. Linda Kao[37], Eric Boerwinkle[38], Alanna C. Morrison[38], Paul M. Ridker[4], Diane M. Becker[7], Jerome I. Rotter[39], Sharon L.R. Kardia[9], Ruth J.F. Loos[40], Martin G. Larson*[3,41], Yi-Hsiang Hsu[42], Michael A. Province[1], Russell Tracy[43], Benjamin F. Voight[44], Dhananjay Vaidya[7], Christopher O'Donnell[11], Emelia J. Benjamin[3,45], Behrooz Z. Alizadeh*[13], Inga Prokopenko*[46], Cross Consortia Pleiotropy (XC-Pleiotropy) Consortium, the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE), the Genetic Investigation of Anthropometric Traits (GIANT) Consortium; the Global Lipids Genetics Consortium (GLGC), the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC), the Global BPgen (GBPG) Consortium, The ADIPOGen Consortium, the Women's Genome Health Study (WGHS), the Howard University Family Study, James B. Meigs*[†47], Ingrid B. Borecki*[†1]

Author Affiliations:

[1]Division of Statistical Genomics, Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, USA;
[2]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA;
[3]National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA;
[4]Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA;
[5]Department of Epidemiology and Carolina Center for Genome Sciences, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, NC, USA;
[6]Department of Epidemiology, University of Washington, Seattle, WA, USA;
[7]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA;
[8]The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark;
[9]Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA;
[10]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands;
[11]National Heart, Lung and Blood Institute (NHLBI) Division of Intramural Research and NHLBI's Framingham Heart Study, Framingham, MA, USA;
[12]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA;

[13]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands;
[14]Department of Epidemiology, Biostatistics and Occupational Health. Jewish General Hospital, Lady Davis Institute, McGill University Montreal, Quebec, Canada;
[15]The Welcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK;
[16]Department of Biochemistry and Molecular Biology, Eberly College of Science and The Huck Institutes of the Life Sciences, The Pennsylvania State University, PA, USA;
[17]Department of Epidemiology, Rollins School of Public Health, and Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, USA;
[18]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA;
[19]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA;
[20]Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy;
[21]Division of Biostatistics, MSIBS Program, Washington University School of Medicine, St. Louis, MO, USA;
[22]Department of Health Studies, University of Chicago, IL, USA;
[23]Department of General and Interventional Cardiology University Heart Center Hamburg-Eppendorf, Hamburg, Germany;
[24]Research Centre for Prevention and Health, Glostrup Hospital, Glostrup, Denmark;
[25]Steno Diabetes Center, Gentofte, Denmark;
[26]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands;
[27]Department of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University, Canada;
[28]Department of Twin Research, King's College, London, UK;
[29]Inherited Disease Research Branch, National Institutes of Health, Bethesda, MD, USA;
[30]University of Mississippi, Medical Center, MS, USA;
[31]Department of Genetics, University of North Carolina, NC, USA;
[32]Department of Internal Medicine, University of Michigan, MI, USA;
[33]National Institute of Aging, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA;
[34]Northwestern University Feinberg School of Medicine, Chicago, IL, USA;
[35]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis MN, USA;
[36]Department of Medicine, Epidemiology, Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA;
[37]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA;
[38]Human Genetics Center, University of Texas-Houston Health Science Center at Houston, Houston TX, USA;
[39]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute (LA BioMed), Harbor-UCLA Medical Center, Torrance, CA, USA;
[40]The Genetics of Obesity and Relate Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY, USA;

[41]Department of Mathematics and Statistics, Boston University, Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA;
[42]Hebrew Senior Life Institute for Aging Research, Harvard Medical School and Molecular and Integrative Physiological Sciences, Harvard School of Public Health, Boston, MA, USA;
[43]University of Vermont College of Medicine, Burlington, VT;
[44]Department of Pharmacology and Department of Genetics, University of Pennsylvania - Perelman School of Medicine, Philadelphia PA, USA;
[45]Section of Cardiology and Preventive Medicine, Boston University School of Medicine, Boston, MA, USA;
[46]Department of Genomics of Common Diseases, School of Public Health, Imperial College London, London, W12 0NN, UK;
[47]General Medicine Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

Emails:
A.T.K.: aldi@wustl.edu, J.D.: dupuis@bu.edu, D.I.C.: dchasman@research.bwh.harvard.edu, K.E.N.: kari_north@unc.edu, A.R.: apreiner@u.washington.edu, L.R.Y.: lryanek@jhmi.edu, T.O.K.: tuomas.kilpelainen@sund.ku.dk, J.A.S.: smjenn@umich.edu, A.D.: a.dehghan@erasmusmc.nl, A.D.J.: johnsonad2@nhlbi.nih.gov, M.F.F.: mfeitosa@wustl.edu, F.T.A.: fasil.ayele2@nih.gov, A.C.: aychu@partners.org, I.M.N.: i.m.nolte@umcg.nl, S.L.: s.ligthart@erasmusmc.nl, Z.D.: zari.dastani@mail.mcgill.ca, A.M.: andrew.morris@well.ox.ac.uk, S.A.P.: sap29@psu.edu, Y.V.S.: yan.v.sun@emory.edu, M.D.R.: marylyn.ritchie@psu.edu, A.V.: a.vaez@umcg.nl, H.L.: hhlin@bu.edu, L.M.: lety@well.ox.ac.uk, R.R.: rohde@email.unc.edu, Y.S.: yaming_shao@unc.edu, M.A.Z.: MZiegler23@WUSTL.EDU, H.K.I.: haky@uchicago.edu, R.B.S.: schnabelr@gmx.de,T.J.: tojo@glo.regionh.dk, M.E.J.: maej@steno.dk, T.H.: torben.hansen@sund.ku.dk, O.P.: oluf@sund.ku.dk, R.P.S.: r.p.stolk@umcg.nl, H.S.: h.snieder@umcg.nl, B.R.: brent.richards@mcgill.ca, J.G.W.: jgwilson2@umc.edu, L.L.: leslie_lange@med.unc.edu, S.K.G.: sganesh@med.umich.edu, M.A.N.: nallsm@mail.nih.gov, A.H.: a.hofman@erasmusmc.nl, A.G.U.: a.g.uitterlinden@erasmusmc.nl, O.H.F.: o.franco@erasmusmc.nl, M.A.I.: m.a.ikram@erasmusmc.nl, L.J.R.: ljrtorvik@northwestern.edu, J.S.P.: panko001@umn.edu, J.C.: coresh@jhu.edu, W.T.: tang0097@umn.edu, W.K.L.K.: wkao@jhsph.edu, E.B.: Eric.Boerwinkle@uth.tmc.edu, A.C.M.: Alanna.C.Morrison@uth.tmc.edu, P.M.R.: pridker@partners.org, D.M.B.: DBecker607@aol.com, J.I.R.: jrotter@labiomed.org, S.L.R.K.: skardia@umich.edu, R.J.F.L.: ruth.loos@mssm.edu, M.G.L.: mlarson@bu.edu, Y.H.: YiHsiangHsu@hsl.harvard.edu, M.A.P.: mprovince@wustl.edu, R.T.: russell.tracy@med.uvm.edu, B.V.: bvoight@upenn.edu, D.V.: dvaidya1@jhmi.edu, C.O.: odonnellc@nhlbi.nih.gov, E.J.B.: emelia@bu.edu, B.Z.A.: b.z.alizadeh@umcg.nl, I.P.: i.prokopenko@imperial.ac.uk, J.B.M.: jmeigs@partners.org, I.B.B.: iborecki@wustl.edu

*Writing group.

[†]Correspondence to:

Aldi T. Kraja, DSc, PhD, Division of Statistical Genomics, Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, phone: 314-362-2498, email: aldi@wustl.edu

James B. Meigs, MD, MPH, General Medicine Division, Massachusetts General Hospital, Boston, MA 02114, 617-724-3203, email: jmeigs@partners.org

Ingrid B. Borecki, PhD, Division of Statistical Genomics, Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, phone: 314-362-3690, email: iborecki@wustl.edu

Content **Page**

ABSTRACT

We analyzed 17 traits in 85,523 participants from 14 large epidemiologic studies within the XC-Pleiotropy Consortium. Individuals classified with metabolic syndrome (MetS, NCEP definition), versus those without, showed on average statistically significant different levels for most inflammatory markers studied. Average correlations estimated with two methods, and factor analyses on large simulated data assisted in identifying 8 choice trait combinations for follow-up meta-analyses. Nine correlated meta-analyses using full published genetic results (predominantly of large consortia) for 8 metabolic traits and 6 inflammatory markers, yielded 130 unique SNPs / genes with pleiotropic effects (a gene affecting at least a metabolic trait and an inflammatory marker). Twenty-five genes (two-third new) are proposed as MetS candidates. Pleiotropic effects identified add to the understanding of MetS and the correlated architecture of its risk factors.

INTRODUCTION

Metabolic syndrome (MetS) is a cluster of increased adiposity/ obesity, atherogenic dyslipidemia with high triglyceride and reduced levels of high density lipoprotein cholesterol, hyperglycemia and systemic insulin resistance, and high blood pressure [1]. MetS has become a health and financial burden [2-7]. MetS captures a confluence of clinical disorders, assisting front-line practitioner in identifying risk factors requiring simultaneously clinical attention [1,8].

There are differing opinions of the genetic etiology of MetS and later outcomes, including whether the MetS risk factors are independent in origin. At the phenotypic level, it has been reasoned that the cardiovascular disease (CVD) risk associated with MetS appears to be no greater than the sum of its single traits' risk [9]. Dallmeier *et al.* [10] suggested that the relationship between MetS and inflammation is largely accounted for by MetS components, once the regression model is adjusted with MetS components as continuous traits, suggesting that MetS as a construct generally is no more than the sum of its parts with respect to inflammation. Henneman *et al* (2008) [11] recommended the genetic dissection of MetS be approached by researching individual components, because of their high heritability.

We advocate that parallel with several genes influencing single MetS risk factors, there are biologically relevant genetic variants that influence MetS and inflammatory biomarkers, forming a larger intertwined network. MetS is associated with at least five fold increased risk in developing diabetes mellitus (T2D) and two fold increased heart disease risk [5]. Individuals with MetS, often have increased levels of C-reactive protein, white blood cell count, coagulation factors VII-IX, von Willebrand factor and plasminogen activator inhibitor 1 as well as decreased levels of adiponectin [12-16]. It has been suggested that modified cytokine expression correlated with an increased adipose tissue may be a mechanism for the inflammation influencing lipid and glucose metabolism, as well as blood pressure [12,17,18]. Currently, it remains unclear whether genetic variants identified for individual metabolic traits [19-23] and inflammatory biomarkers [24-28], have pleiotropic effects, thereby influencing the correlated architecture of these traits. As part of the "Pleiotropy among Metabolic traits and Inflammatory-prothrombotic biomarkers" working group (PMI-WG), a sub-group of the *Cross Consortia Pleiotropy (XC-Pleiotropy) Consortium*, we aimed to: 1) assess relations among MetS and inflammatory markers; 2) isolate promising trait combinations, by evaluating correlations among metabolic traits and inflammatory markers, for evaluating the role of pleiotropy in MetS etiology; 3) utilize trait-choice-combinations of the 2-nd objective to perform meta-analyses of several large meta-GWAS-trait consortia and studies published full results (already archived in the XC-Pleiotropy repository), for identifying MetS candidates with potential pleiotropic effects on metabolic traits and inflammatory markers.

RESULTS

1. Association of inflammatory markers with MetS

We studied the association at the phenotypic level between 9 inflammatory biomarkers (C-reactive protein (**CRP**), fibrinogen (**FIB**), plasminogen activator inhibitor 1 (**PAI-1**), interleukin 6 (**IL-6**), interleukin 10 (**IL-10**), intercelular adhesion molecule 1 (**ICAM-1**), white blood cell counts (**WBCC**), tumor necrosis factor alpha (**TNFA**) and adiponectin (**ADIP**)) and 8 MetS risk factors (body mass index (**BMI**), waist circumference (**WAIST**), high density lipoprotein cholesterol (**HDLC**), triglycerides (**TG**), fasting glucose (**GLUC**), fasting insulin (**INS**), systolic

(**SBP**) and diastolic blood pressure (**DBP**) (Supplemental Table 1). The mean age in 14 cohorts (Table 1.a) with a total of 85,523 participants, varied from 25 (SD=+/-3) years in CARDIA to 74 (SD=+/-8) years in the Rotterdam Study. These studies capture a large variability of the human population, from 2.4% MetS prevalence in CARDIA-EA to 58.9% in GENOA-EA. The prevalence of MetS and its components, as well as the mean levels of inflammatory biomarkers in individuals with and without MetS, are summarized in Figure 1 for the Family Heart Study (FamHS) and the Framingham Heart Study (FHS) and summarized for all studies in the Supplemental Figures 1 (a-g). Overall, when comparing mean levels of inflammatory biomarkers after stratifying for individuals with MetS versus those without, the mean levels of biomarkers were significantly different (passing Bonferroni threshold, $p \leq 9.43e\text{-}04$) between the two strata in 85% (45 out of 53) of comparisons. FIB, CRP, PAI-1, ICAM-1, WBCC and TNFA mean levels were elevated, whereas ADIP mean level was lower in individuals with MetS. There were also exceptions such as IL-10 (present only in one study), which did not show significant mean differences between individuals with and without MetS.

2. Correlations among metabolic traits and inflammatory markers

Of interest were the correlations of traits at single studies. Their generalization to a global average correlation matrix can help in inferring combinations of metabolic traits and biomarkers to be used for better understanding pleiotropy in the underlying MetS etiology. First, simulations performed mimicked the correlation substructure of individual studies (Methods.3). The average correlations estimated from 100 replications of the first batch of simulations with 85,523 individuals per replication are presented in Table 3. Second, using Fisher's Z-transformation (Methods.3) produced the average correlation coefficients from 14 cohorts (Supplemental Table 2). The average correlations resulting from the two methods are very similar. Based on p-values, significant correlations between biomarkers and metabolic traits were **(1)** FIB and CRP with the metabolic traits studied; **(2)** ICAM-1 and TNFA with HDLC and TG; and **(3)** ADIP and WBCC with WAIST, HDLC, TG and INS (Supplemental Table 2).

Additionally, the application of factor analysis (Methods.4) on a second batch of simulated data, (with average correlation structure of all studies and no missing observations), yielded useful trait clusters. Clusters produced (Supplemental Figure 2) were: Factor 1 representing a combination of **(4)** BMI, WAIST, INS, CRP, PAI-1 and weaker contributions of HDLC and TG; **(5)** weak contributions of BMI and WAIST were associated in Factor 2 together with stronger contributions of FIB, CRP, IL-6 and WBCC; **(6)** TG and less so HDLC, contributed along with CRP and WBCC in Factor 4; **(7)** HDLC and TG with PAI1 and ADIP in factor 5, and **(8)** GLUC and INS contributed in Factor 6 along with the contribution of PAI-1. Supplemental Table 3 shows results of the coefficients of congruence (CC - formula provided in Methods.4). Similarity of factor 1 from all replications was high (CC=0.99) and less in factor 4. Factor 3 had only contributions from blood pressure and no noteworthy contributions of inflammatory markers and thus was not considered in the correlated meta-analyses. As a result, eight trait clusters were selected for meta-analyses.

3. Correlated meta-analyses

We performed nine correlated meta-analyses (the eight trait-combinations predicted in Results.2, and one including all variables), utilizing full results from mainly large meta-GWAS consortia (Table 1.b) for 8 metabolic traits (BMI [22], WAIST [29], HDLC and TG [23], GLUC and INS [19], SBP

and DBP [21]), and 6 inflammatory markers (CRP [24], PAI-1 [25], ICAM-1 [27], WBCC [26], ADIP [30] and IL-6 [31]). A filter for our meta-analyses results of $-\log_{10}p \geq 8$ and at least one metabolic trait and at least one inflammatory marker results with $-\log_{10}p \geq 3$ was applied. After selecting one best SNP per gene, (either a polymorphism within a gene or intergenic SNP assigned to the closest proxy gene), the results were reduced to 130 unique SNPs and genes (Supplemental Table 4). Of them, 25 genes were selected as candidates for MetS (with associations to at least two metabolic traits from our analyses or GWAS literature and at least one association with inflammatory markers). They may represent 15 loci with pleiotropic effects to MetS and facilitating inflammation. A short description of known functions of 25 genes is provided in Box 1. A summary of our work is shown in Table 2. Additional evidence for 25 genes is condensed in Box 2, in Supplemental Table 5 and in Figure 2 (including regulatory evidence of ENCODE, by utilizing the HaploReg [32] and regulomeDB [33] software).

As shown in Figure 2, we identified three groups of genes for their pleiotropic effects. With a black triangle annotated are the ones with **pleiotropic effects for lipids and inflammation**. In this group were genes, ***MACF1*** [34 35] and ~240K bps distant ***KIAA0754*** on chromosome 1, both significantly associated with HDLC and less with WAIST, TG, GLUC and CRP. On chromosome 2, a rich strand (~1.2M bps in length) of 23 contiguous genes, from *TCF23* to *BRE* was associated with TG and CRP. Although several genes are important in this small chromosomal region, we call this region the ***GCKR*** region, because rs1260326 represents a missense change (from C**T**G (LEU) to C**C**G (PRO)) highly significant in association with TG [23] and CRP [24]. Another independent group of genes on chromosome 2 were: ***GRB14*** and ***COBLL1***. They are positioned about 4.7K bps apart, both were associated with HDLC, TG, less so with PAI-1 and more so with ADIP. An uncharacterized gene ***LOC646736***, which is distant from the above two, is located ~528K bps from the *IRS1* gene. Near *LOC646736* (~23K bps), our meta-results showed associations with HDLC, TG and less so with ADIP. Intron variants of ***BAZ1B***, ***BCL7B***, ***TBL2*** and ***MLXIPL*** (7q11.23) were associated significantly with TG, and demonstrated good associations with HDLC and CRP. ***LPL*** (8p22) was associated significantly with HDLC and TG and less with CRP. ***TOMM40*** (19q13) showed similar patterns. More associations were observed for rs10808546 about 45K near ***TRIB1*** (8q24.13), which was significantly associated with TG and HDLC [23], less so with ADIP, and less so with PAI-1. ***ZNF664*** (12q24.31) associated with TG, HDLC and ADIP.

In Figure 2, with a blue square were annotated genes with **pleiotropic effects for adiposity / obesity and inflammation**. In this group participated ***TFAP2B*** (6p12), which significantly associated with BMI and WAIST and less with CRP; ***HECTD4*** (12q24.13) and ***PTPN11*** (12q24) were associated significantly with ICAM-1 and less with DBP, SBP, HDLC, BMI and WAIST, while ***FTO*** (16q12.2) was significantly associated with BMI, WAIST, less so with CRP, and less with INS.

The third group of genes, depicted in Figure 2 with green diamond shapes, showed **pleiotropic effects for adiposity / obesity, lipids and inflammation**. Among them were ***SLC39A8*** (4q22-q24), with the selected missense variant rs13107325 from our meta-analysis, which associated significantly with HDLC [23], less so with BMI, and less so with ADIP, SBP, DBP and WAIST. The same SNP was associated with blood pressure, hypertension (HTN) [36], and BMI [22]. An interesting group of genes with significant associations with TG and lower associations with

BMI, WAIST, SBP, PAI-1 and WBCC were ***NELFE***, ***SKIV2L*** and ***STK19*** (6p21). They position in the class III region of the major histocompatibility complex of chromosome 6, close to *C2* gene, the last one was also included in our 130- candidate gene' list. ***PDXDC1*** (16p13.11) was associated significantly with ADIP and less so with WAIST and TG. Finally, ***MC4R*** (18q22) was associated significantly with BMI, WAIST, less so with CRP and less with HDLC and TG.

4. Bioinformatics analyses

The overlap of genes with effects on metabolic traits and inflammatory biomarkers suggests prior literature evidence of pleiotropy. First, the keyword searches (Methods.6) using Gene Entrez of NCBI produced a list of 770 genes that had a relationship with at least one of the eight metabolic traits and at least one of the nine biomarkers. Of them, 48 putative pleiotropic genes showed at least a total of 8 relations with metabolic traits and biomarkers, sourced from three species: human, mouse and / or rat (Supplemental Table 6). Highest ranked for possible pleiotropic effects were the *ADIPQ*, *PPARG* and *LEP* genes. Of this list, only four of them (*APOE*, *FTO*, *MMP9* and *VEGFA*) match with our 130 pleiotropic gene list (Supplemental Table 4).

A second source of pleiotropic candidate genes was selected from the literature of previous GWAS (Supplemental Table 7). Eleven genes in this list showed one association with selected biomarkers, but up to four associations with metabolic traits. Among them, *GCKR* was associated with four metabolic traits and CRP, while *TRIB1* and *TOMM40* were associated with HDLC, TG and one biomarker ADIP and CRP, respectively. With the exception of *CSMD1*, ten genes (*GCKR*, *IRS1*, *LYPLAL1*, *TRIB1*, *APOE*, *TOMM40*, *PPP1R3B*, *PEPD*, *BCL7B*, *TMEM18*) are present in the list of 130 pleiotropic candidate genes of metabolic traits and inflammatory markers.

A third source of pleiotropic candidate genes was the gene search for "metabolic syndrome" via dbGaP Association Results Browser, which includes findings of www.genome.gov from large GWAS in humans. This search yielded 30 MetS candidate genes (Supplemental Table 8). Of them, *GCKR*, *C2orf16*, *ZNF512*, *TFAP2B*, *MLXIPL*, *LPL*, *TRIB1*, *MTNR1B*, *FTO*, *TOMM40*, 33% of MetS list represented 7.7% of our 130' genes pleiotropic list, and *GCKR*, *TFAP2B*, *LPL*, *TRIB1*, *FTO*, *TOMM40* 20% of MetS list represented 24% of genes from our 25 MetS candidates (Table 2).

The bioinformatics of our 25 MetS candidate genes shows that only a few contribute to the GeneGO Canonical pathway maps. *PTPN11* and *GRB14* are up-regulated, part of the "Development Angiopoietin Tie2 signaling" (enrichment p= 2.4E-04), conveying anti-inflammatory action. *PTPN11* is part of six other maps, while *LPL* is part of three maps. GeneGO enrichment analysis ranked as the top diseases "Metabolic Syndrome" (p= 9.0E-07); "Obesity" (p= 8.5E-07); and "Insulin Resistance" (p= 5.6E-07). From our list, some of the genes also have been studied for pharmacologic applications. *LPL* is a therapeutic drug target for Ibrolipim (activation) and Gemfibrozil (activation), while *MC4R* is a target for Bremelanotide (activation) and *PTPN11* is a target for Stibogluconate (inhibition).

GeneGO database conveyed an understanding for the 130 candidate pleiotropic genes also. The pathway map of "*ZNF202* role in gene expression in atherosclerosis", was enriched for genes affecting lipid metabolism (p=7.0E-08), while less significant p-values were for other pathways. For process networks, the most common were those about inflammation. Because HLA genes are quite dominant in these pathways, removal of 7 genes, whose names started with HLA, produced a list of 123 pleiotropic candidate genes. The pathway maps remained similar as above, however process networks changed to "Complement system" (Inflammation, p= 5.7E-04), and "Blood vessel morphogenesis" (Development, p= 1.2E-03). For the disease classification, GeneGO reports the top ranking diseases as "Metabolic Syndrome" (p= 1.2E12, *TRIP8*, *BMAL1*, *GCKR*, *C2orf16*, *LPL*, *MMP-9*, *HNF4-alpha*, *NTPBP*, *APOE*, *TRIPs*, *TFAP2A*, *ZNF512*, *VEGF-A*, *AP-2B*, *MC4R*, *Notch*, *RGPR*, *Galpha(s)-specific peptide GPCRs*, *FTO*, *HNF4*, *CCDC121*), Obesity (p= 6.1E-11), "Coronary disease" (p= 1.6E-08), "Macular degeneration" (p= 3.7E-08) and T2D (p= 7.5E-08). In the GO processes ranked at the top were "Glucose homeostasis" (p= 3.0E-09, "Positive regulation of vascular permeability" (p= 8.8E-09) and "Regulation of insulin secretion" (p= 4.0E-07).

Using the Literature Lab software from ACUMENTA for an automated literature interrogation [37], the same list of 25 genes showed association, compared against 1000 random sets of genes, for overnutrition (p=0.0039), obesity (p=0.0041), nutrition disorders (p=0.0053), heart valve diseases (p=0.0112), and fatty liver (p=0.0124). The contributing genes in these disease-MeSH term clusters, ranked by the number of the corresponding publications, were for overnutrition: *MC4R* (46.3%), *FTO* (42.38%), *LPL* (10.39%) and *MLXIPL* (0.62%); similar genes were in ranking order for obesity and nutrition disorders; for heart valve diseases *BAZ1B* (47.02%), *PTPN11* (37.50%), *TBL2* (7.73%), and *BCL7B* (6.63%); and for fatty liver *MLXIPL* (89.50%), *LPL* (8.02%) and *GCKR* (1.82%).

DISCUSSION

This is the first time that a large sample of 85,523 participants with 8 metabolic traits and 9 inflammatory markers are analyzed together for understanding relationships of inflammatory markers and MetS. We are statistically confident that inflammatory markers FIB, CRP, PAI-1, ICAM-1, WBCC and TNFA have elevated mean levels, while ADIP has a lower mean level in individuals with MetS compared to those without. Such findings may hint for presence of pleiotropy. For studying pleiotropy we explored the average correlations of 17 traits generated from 14 cohorts. Correlation estimations and factor analyses yielded eight choice trait-combinations out of 130,305 possible combinations between metabolic traits and inflammatory markers, which we hoped would reflect some latent genetic correlation. During this analytical process we demonstrated also that large data (100 replications of simulations with 85,523 participants each) produces an average correlation matrix similar to the one estimated with Fisher's method.

This is also the first time that 8 metabolic traits and 6 inflammatory markers from mainly largest meta-GWAS are studied in a correlated meta-analyses for inferring pleiotropic variants with effects on MetS. The analysis yielded 130 top ranked genes with putative pleiotropic effects between metabolic traits and inflammatory markers. Twenty-five genes, each represented by a common variant with pleiotropic effects, were considered as contributors to MetS. In principle, genetic makeup and environment enable differences in developing MetS, whereas total burden is

related to how many and direction, of disease predisposing alleles one carries. However, pleiotropy is more than counting alleles. Thus, we considered MetS candidate genes to be the ones that were in significant association with 2 or more components (from our study and GWAS literature), and have an expected pleiotropic effect in the low grade development of inflammation.

Then what does pleiotropy represent at the gene level? Here we focus on a not well known cluster. At first glance, genes *BAZ1B, BCL7B, TBL2* and *MLXIPL*, show pleiotropy by affecting similarly TG, HDLC and CRP. A few SNPs of *BAZ1B* were associated with TG [38], protein C [39], and serum urate concentration [40]. *BCL7B*'s SNPs associated with CRP [24] and with gamma-Glutamyltransferase [41]. *TBL2* associated with TG [23,42,43] and with HDLC [23]. *MLXIPL* was associated significantly with very low density lipoprotein (VLDL) [44], with MetS [45], with TG [46], and with gamma-Glutamyltransferase [47] (Box 2 and Supplemental Table 5). Deletions of these contiguous genes at 7q11.23 have been identified as causing a Williams-Beuren syndrome, a multisystem developmental disorder, where 75% of cases show severe GLUC intolerance [48]. *BAZ1B* and *MLXIPL* may serve as transcription factors. The intron rs17145750 of *MLXIPL*, based on regulomeDB shows some minimal regulatory signature, and from HaploReg affects a PPAR motif. The rest of selected SNPs have also some minimal regulatory properties. A large part of SNPs in these genes are under two *linkage disequilibrium* blocks (HapMap figure not shown), that overlap with each other. It has been shown that *MLXIPL* protein forms a heterodimeric complex and activates, in a glucose-dependent manner, carbohydrate response element (ChoRE) motifs in the promoters of triglyceride synthesis genes. Recently, Herman *et al.* [49] showed in mice that *GLUT4*, officially known as *SLC2A4* (known to be used by insulin for stimulating glucose uptake), regulates the expression of *MLXIPL*. A number of papers with work in mice have proposed *MLXIPL* as another regulator of systemic glucose metabolism, such that it plays a critical role in converting excess carbohydrates to TG by way of *de novo* lipogenesis (DNL) [48,50,51]. Donnelly *et al.* [52] studied 9 non-alcoholic fatty liver disease participants (with excess liver TG) and showed that about 26% of TG in the liver was result of DNL, 59% from serum nonesterified fatty acids, 15% from diet, and a similar pattern of isotope labelling in VLDL, concluding that DNL contributes to the accumulation of hepatic fat. Jeong *et al.* [53], studying expression of *MLXIPL* using ChIP-seq, identified 14 genes as direct targets that affect the paths from GLUC to TG. They also proposed that *MLXIPL* is an activator and repressor based on gene expression patterns of target genes. The role of *MLXIPL* in controlling glucose and lipid homeostasis is complex, because in C57BL/6 mice, global deficiency of *MLXIPL* leads to insulin resistance [50], while in obese mouse with *ob*/*ob* background (leptin deficiency) [50] leads to improved hepatic steatosis and improved insulin resistance. Moreover, Benhamed *at al.* [48] proposed that *MLXIPL* in the mouse liver raises beneficial lipid species. Thus, the pleiotropic effects of *MLXIPL* are more than what they appears in our association tests.

Our findings are supported with additional GWAS results for several genes of three major pleiotropic groups presented in Figure 2. A comprehensive GWAS and functional evidence is reported in Box 1, Box 2 and Supplemental Table 5 as evidence about our findings grouped by a) pleiotropic genes for lipids and inflammation, b) for adiposity / obesity and inflammation, and c) for lipids, adiposity / obesity and inflammation [11,23,30,41,45-47,54-104]. The simple reason that we amass such information, is that our meta-results are the contemporary world's largest studies available. Hence, it is very difficult to find other similar samples as replications for most of the

traits studied, with exception of IL-6, which had the largest SNPs sample (~5M), but smallest individuals' sample in our meta-analyses (n=707). Previous studies have shown that MetS pillars, (excess abdominal fat obesity, atherogenic dyslipidemia, hyperglycemia and insulin resistance, high BP, and a proinflammatory and prothrombotic state [1]) can be modified by genes that affect MetS components individually. That might explain important contributions in developing MetS [10,45]. However, we show that pleiotropic genes also play a role in MetS. Two-third of our 25 MetS pleiotropic candidates are new, compared to 30 MetS candidates' list (Supplemental Table 8). These findings, (graphically summarized in Figure 2), reinforce the importance of inflammatory markers in the etiology of MetS and that pleiotropy can additionally contribute in clustering MetS pillars.

Kristiansson *et al.* [45] recently confirmed that 22 previously identified susceptibility loci for individual MetS risk factors replicated in their GWA and factor analysis. Most of them associated with lipid phenotypes and none with two or more uncorrelated MetS components, but they did not find evidence of pleiotropy of these genes with obesity. Although their findings are right in principle, our study based on very large GWAS meta-analyses indicates, that there are also exceptions. For example, *MC4R* (rs6567160) showed highly significant associations with WAIST, BMI, HDLC and less so with TG and CRP; *NELFE* (rs419788), *SKIV2L* (rs437179), *STK19* (rs389883) were associated strongly with TG, less so with WAIST, SBP, PAI-1, WBCC, and less with BMI; *SLC39A8* (rs13107325) was associated significantly with HDLC and less so with BMI, WAIST, SBP, DBP and ADIP; *MACF1* (rs1537817) was associated significantly with HDLC and less so with CRP, TG, WAIST, and GLUC.

It is interesting to note that in the 130 pleiotropic candidates (Supplemental Table 4), a number of genes are associated noticeably with ADIP and HDLC. Ye and Scherer [105] summarized pleiotropic effects of ADIP by reviewing either recombinant adiponectin protein, or endogenously its overproduction. In adipose tissue, ADIP lowers inflammation and increases glucose uptake, fat storage and adipogenesis; in muscle leads to an increase of fatty acid oxidation; in heart decreases injury and apoptosis; in endothelium decreases oxidative stress and increases angiogenesis and function; in liver increases insulin sensitivity and lowers gluconeogenesis and lipogenesis; in microphages increases insulin sensitivity and lowers inflammation. Is it possible specific alleles of genes associated with ADIP may protect from disease, because of pleiotropic effects with anti-inflammatory properties? Among our findings are rs2785990, ~300K bps from *LYPLAL1* (a lysophospholipase like); *GRB14*, *COBLL1*, and *LOC646736*; rs9853056 of *STAB1* (a scavenger receptor); rs4282054 of *NT5DC2*; *FAM13A*, *SLC39A8* and *ARL15*; rs998584 near *VEGFA*, a glycosylated mitogen that mediates increased vascular permeability, induces angiogenesis, vasculogenesis and endothelial cell growth, promotes cell migration, and inhibits apoptosis; and rs509548 near *HCAR2*, which is expressed in monocytes and macrophages [106]. Other candidates for a specific allele with anti-inflammatory properties could be, *ZNF664* with an unknown function; rs2925979 an intron of *CMIP* involved in T-cells signaling and reported to down regulate *NF-kB* activity [107]; and rs731839 an intron of *PEPD*, gene with a few known roles [108], including elevated activity at the wound site following traumatic injury [109], and appears to activate *EGFR* which would result in cell proliferation [102].

In the list of 130 pleiotropic genes, a few special patterns emerged. The rs2943634 on chromosome 2 was associated with HDLC, TG and ADIP and resides 23.3K bps from an uncharacterized *LOC646736* gene and ~528K bps from *IRS1*. We reported additional evidence for its association with coronary disease as well as a number of neighboring SNPs with T2D, TG, adiposity, and ADIP. These SNPs closer to *LOC646736* gene are not eQTLs of *IRS1*. It is compelling to understand their function toward MetS and relations with *IRS1*. Co-localization might relate with evolutionary functional importance, which is observed in our data for gene clusters. For example, a missense SNP (rs1260326, T=0.3963) of *GCKR* associated with similar traits as rs1919127 (C=0.2647) a missense of *C2orf16*, also as rs23844656 (G=0.2642), an intron of *ZNF512* and rs13002853 (G=0.2593) a variant of *CCDC121*; another cluster was for *DNAH10*, *CCDC92*, and *ZNF664* on chromosome 12, and for *HNF4A*, *PLTP*, *PCIF1*, *ZNF335* and *MMP9* on chromosome 20. Such clusters reminded of a similar pattern previously published on chromosome 11 for *APOA5*, *ZNF259*, and *BUD13*, where a zinc finger probably controls the transcription of nearby genes [62]. It is possible that neighboring gene-variants produce similar results in the associations, because of conserved haplotypes. In the 130 pleiotropic genes, 11 transcription factors (*HEYL*, *SEC16B*, *GTF3C2*, *ZNF512*, *GTF2H4*, *TFAP2B*, *BAZ1B*, *MLXIPL*, *ZNF664*, *MED24*, *HNF4A* and *ZNF335*) represent about 8.5% of the list. Vaquerizas *et al.* [110] reported 1,391 high confidence loci that encode transcription factors, about 6% of the total of human protein coding genes. Thus the 130-genes' list shows patterns that might be common for function conservation. Another feature learned by comparison of 130 pleiotropic candidate genes with the 30 MetS candidate genes (Supplemental Table 8) was that, although *APOA5* and its cluster, as well as *CETP*, *LIPC*, *GALNT2* involved in lipid metabolism are considered important contributors to MetS, they appear not involved directly in the inflammation process.

Although we ponder the 48 candidates gene list (Supplemental Table 6) as a "soft" one, because of its vast research of human, mouse and / or rat considered, the bioinformatics searches exhibited that other genes with pleiotropic effects among metabolic traits themselves and also with inflammation remain to be discovered. The bioinformatics research provided additional information not only in support to our findings, but also to a finer understanding of gene effects as is the case of *BAZ1B*, *PTPN11*, *TBL2*, *BCL7B* for heart valve disease, and *MLXIPL*, *LPL* and *GCKR* in relation to fatty liver disease as revealed by Literature Lab.

However, our study has weaknesses also. First, our analyses operate on meta-pvalues, without accounting contributing beta-directions at each SNP location of each study, because some studies did not share beta-coefficients results. Second, we are utilizing for individual contributing metabolic traits and inflammatory markers, thresholds that are lower than genome-wide thresholds ($-\log_{10}p \geq 3$), with the hope to capture "low hanging fruits", and that filtering relying on meta $-\log_{10}p \geq 8$ will weed out false positives. Third, with association results available it was not possible to evaluate mediation, as presented for the rs9939609 of *FTO*, and BMI on other traits by Freathy *at al*.[111] and Fall *et al*.[112].

In conclusion, pleiotropic effects of identified genes add in understanding MetS and its risk factors correlation structure. Among genes with pleiotropic effects in our study some of them give hope that for specific alleles they may have contributions to protect from MetS, including inflammation.

ONLINE METHODS

A Brief Summary of Implemented Methods

The international collaboration of XC-Pleiotropy was founded in early 2011 for studying pleiotropy by using published GWAS results. The PMI-WG built a collaborative network within the XC-Pleiotropy (Supplement 1), for implementing the first two objectives. Seventeen metabolic traits and inflammatory markers are studied (Methods.1), from 14 large-scale cohort studies (dependent on cohort-specific assay availability, Table 1.a and Supplement 2). Together these data represent a total of 85,523 individuals (Supplemental Table 1). Laboratory methods for obtaining these traits are described in Supplement 2. Traits adjustments for medication use and other covariates are provided in Methods.2. Methods of estimating correlations with simulations and Fisher's Z-transformation are provided in Methods.3, and factor analysis in Methods.4. Each study was approved by its local ethics board and each participant provided written, informed consent.

Moreover, we utilize (for implementing the 3d objective) published full results from meta-GWAS-trait consortia and studies (Table 1.b). With them we performed correlated meta-analyses [113,114] (Methods.5) for identifying pleiotropic variants for metabolic traits and inflammatory markers. In this paper, a gene is considered pleiotropic when affects at least a metabolic trait and a biomarker. In this framework, our study includes published results for **BMI** [22], **WAIST** [29], **HDLC** and **TG** [23], fasting **GLUC** and fasting **INS** [19], **SBP** and **DBP** [21]. In addition, our meta-analyses included inflammatory markers, **CRP** [24], **PAI-1** [25], **WBCC** [26], **ADIP** [30], **ICAM-1** [27], and **IL-6** [31]. Because **IL-10** was not significant in the correlation of traits, and **FIB** as well as **TNFA** GWAS results were not available, (although analyzed when studying correlations), the three last traits are not present in the meta-analyses. The reported allele frequencies were based on GIANT BMI. When a SNP was not studied in GIANT BMI, then MAGIC GLUC meta-analysis allele frequencies were used. We used also bioinformatics approaches for appraising pleiotropy (Methods.6).

1. Traits studied

To evaluate the associations between inflammatory biomarkers and MetS risk factors, seventeen traits were studied at the phenotypic level. Metabolic traits included BMI (kg/m^2), WAIST (in cm), HDLC (mg/dL), fasting (at least 8 hours) TG (mg/dL), fasting INS (mU/L), fasting GLUC (mg/dL), SBP and DBP (mm Hg), as average of three seating measures, or the 2-nd and 3-rd ones. We use the term "inflammatory biomarkers" for brevity when referring to the inflammatory – prothrombotic markers fibrinogen (FIB) (mg/dL), CRP (mg/L), PAI-1 (IU/mL), tumor necrosis factor alpha (TNF-alpha) (pg/mL), ICAM-1 (ng/mL), IL-6 (pg/mL), interleukin 10 (IL-10) (pg/mL), WBCC (10e9/L) and ADIP (μg/mL). The studies had a variable number of traits, dependent on the assays performed. The metabolic traits studied belong to domains of adiposity/ obesity (BMI, WAIST), lipids (HDLC, TG), glucose metabolism and insulin (GLUC, INS), and blood pressure (SBP, DBP). In addition, FIB and PAI-1 represented prothrombotic markers, and CRP, IL-6, IL-10, ICAM-1, TNF-alpha, WBCC and ADIP represented markers of immune or inflammatory response. In the study of correlations, we sought to find the average correlation among all traits for 14 cohorts through simulations and average correlation using Fisher's Z-transformation. The MetS definition, data analyses methods, adjustments for

medications use (for blood pressure and lipids medications) and covariates were similar for all contributing cohorts and described in Methods.2.

2. MetS definition, variables' adjustments for medications and other covariates

A participant was classified with MetS when thresholds were passed for three or more out of five traits of the NCEP improved threshold [115]: *WAIST ≥ 102 cm for men / WAIST ≥ 88 cm for women*; *GLUC ≥ 100 mg/dL*; *TG ≥ 150 mg/dL*; *HDLC < 40 mg/dL for men / HDLC < 50 mg/dL for women*; *SBP ≥ 130 mmHg / DBP ≥ 85 mmHg*. The MetS was defined based on the improved National Cholesterol Education Program [115] using original traits adjusted for medication use only (in all cohorts, except for WGHS, which did not measure GLUC), representing (B) set of data (see Supplemental Tables 9-22). T2D was defined as following: *(GLUC ≥ 126 mg/dL AND age ≥ 40 years) OR ((using anti-diabetic medications OR insulin) AND diabetes onset age ≥ 40 years)*.

The average blood pressure was adjusted for individuals using antihypertensive medication(s) as follows, *SBP = measured SBP + 15 mmHg*; and *DBP = measured DBP + 10 mmHg* [116]. For individuals using anti-hyperlipidemic medications, their lipid levels were adjusted respectively as follows, *HDLC = measured HDLC / (1+0.04419)*; and *TG = measured TG / (1-0.17159)*. For lipids, adjusting constants are produced as a summary of Wu *et al* work [117] and also of our additional unpublished summary follow-up, which combined together to a total of 92 clinical trials (for HMG-CoA reductase inhibitors, Fibric Acid Derivatives, Cholesterol Absorption inhibitor, Nicotinic acid derivatives, Bile sequestrants and Fish oil) including 53,005 participants for HDLC and 53,432 participants for TG. All participating studies set to missing GLUC and INS values for individuals that were taking insulin or diabetic medications. Before performing any analysis, the participating studies made sure that each variable had a normal distribution, or transformed them to near normal. For example, a natural log transformation worked well for TG in general for all cohorts. In the FamHS, GLUC had a high kurtosis, thus applying a Box-Cox power transformation it was found, that $1/\text{GLUC}^2$ transformation worked well in acquiring a near-normal distributed GLUC. As a result, for any bivariate correlations in the FamHS that included GLUC, correlations coefficients were multiplied by (-1), because power transformation for GLUC reversed the sign compared to original corresponding correlations. As an empirical check, when compared to FHS, the GLUC correlations in FamHS were very similar, although a transformation of GLUC was implemented in the FamHS. In addition, phenotypes were adjusted for polynomial age trend (age and age^2), sex and important study specific covariates (e.g. field center) which were included in the regression model for generating the final data for analysis: standardized residuals, i.e. with mean 0 and variance of 1.

In the Supplemental Tables 9-22, we present statistics for individual studies for (A) original variables, (B) original variables adjusted only for medication use, and (C) residuals from regression with mean 0 and variance 1 of variables obtained from adjusting (B) data for additional covariates (as mentioned above within Methods.2). In the correlation statistical analyses we use the (C) data.

3. Correlation statistical analysis and simulations

We grouped participants' data in strata with- and without MetS (M_1 versus M_0), for analyzing mean differences of inflammatory biomarkers in these two subgroups for each cohort. We used (B) data and pooled t-test for testing mean differences between the two: $(\bar{x}_1 - \bar{x}_2)$, with sample

sizes n_1 and n_2 via $t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, where $s_p = \sqrt{\frac{(n_1-1)* s_1^2 + (n_2-1)* s_2^2}{n_1+n_2-2}}$ is the pooled standard deviation and $n_1 + n_2 - 2$ degrees of freedom. In general, the MetS subgroup sample size was smaller than non-MetS one, but the variances between M_1 and M_0 subgroups were similar. The mean differences p-values were tested against a conservative Bonferroni p-threshold for an α=0.05 experiment-wise, which corresponded for 53 tests to a p= 9.43e-04. Statistics of MetS, its components as well as of biomarkers by cohort are summarized in Figure 1 and Supplemental Figures 1, (a-g). The biomarker boxplot graph comparisons were built by using simulations via "rnorm" function in R with mean, standard deviation and sample size corresponding to subgroups with- and without MetS from the original data.
The above analysis was followed by correlation analyses (including up to 17 traits), performed on (C) data near normally distributed, adjusted for medication use and covariates. All pairwise correlations were performed using Pearson correlation procedure (using SAS v. 9.3 or R v. 2.15.1, presented in Supplemental Tables 9-22).

We then used two parallel approaches, simulation and Fisher's Z-transformation, to confirm our results, when averaging correlation coefficients of each bivariate trait combination among all cohorts. First, simulation processes were implemented to produce the average correlation matrix and the final correlated simulated data across all studies (N = 85,523 individuals) based on (C) data. Simulation 1 was performed following these steps: using N (largest number of participants per study) and variance-covariance matrices (from above single studies) we simulated multivariate normal distributions (MVN(0,1)) of dimension (p-variables, N-participants) of each study, using an R multivariate normal generating ("mvrnorm") function of the MASS library [118]. Since in simulations we used the largest number of participants per study, next, we introduced (in random patterns) missing values in traits when they were not available in all participants of a specific cohort. Thus, 100 replications of simulated data imitated correlations and sample size of the original cohorts. When pooled they formed all studies' set. These data represented all traits, but with corresponding per trait missing values. Correlations of simulated data were evaluated via Pearson pairwise correlation, which produced a full variance-covariance matrix, representing a simulated approximation of the average correlation matrix of single studies. The covariance matrix (correlations among metabolic traits, metabolic traits and biomarkers, and among biomarkers) of simulation 1 are presented in Table 2. Simulation 2 (again 100 replications) were implemented by using the first simulation's average variance-covariance matrix, to produce multivariate normal (MVN(0,1)) with p = 16 variables and N = 85,523 individuals and no missing values. Simulation 2 were built with the purpose of implementing factor analyses on them.

Second, we performed Fisher's Z-transformation for averaging correlations of real (C) data (Supplemental Table 2). Assuming that correlations of any two independent bivariate samples (r_1 and r_2) of n_1 and n_2 sample sizes for the same trait combinations are random samples from a larger population, a combined correlation estimate ($\bar{r}$) can be computed. Application of the Z transformation of the two sample correlations follows: $Z_1 = tanh^{-1}(r_1)$ and $Z_2 = tanh^{-1}(r_2)$, where $tanh$ is hyperbolic tangent and the Z can be calculated as $Z = 0.5 \ln\left(\frac{1+r}{1-r}\right) = artanh(r)$, where $artanh$ is hyperbolic arctangent applied to each correlation coefficient.
The weighted average $\bar{Z}$ of the corresponding Z values is

$$\bar{Z} = \frac{(n_1 - 3)Z_1 + (n_2 - 3)Z_2}{n_1 + n_2 - 6},$$

where the weights are inversely proportional to their variances ($V(\bar{Z}) = 1/(n_1 + n_2 - 6)$). Thus, a combined correlation estimate is $\bar{r} = \tanh(\bar{Z})$. We extended averaging correlation coefficients for each bivariate trait combination to include up to 14 cohorts' correlation estimations, by writing a SAS macro program that implements Fisher's Z-transformation averaging via SAS MIANALYSE procedure. The IL-10 was dropped from these analyses, because it was present only in one study.

4. Factor analysis

Factor analyses with "Varimax" rotation were performed in SAS, v. 9.3. The purpose of using a multivariate statistical analysis was to identify latent clusters of traits that may help in identifying MetS and inflammatory markers underlying etiology. "Varimax" rotation creates orthogonal clusters of correlated variables. The objective is to maximize the independence of the clusters of correlated variables that contribute onto specific factors. An absolute value of a loading 0.4 or larger (which represents a correlation of an original variable to a factor when the data are standardized) is considered in the scale of correlations as a significant contribution. For accounting about stochastic processes in the 100 simulations, 100 factor analyses (p=16, N=85,523) with "Varimax" rotation were considered (Supplemental Figure 2). A coefficient of congruence was calculated as: ($CC = \frac{\sum_{n=1}^{ntraits} l_1 l_2}{\sqrt{(\sum_{n=1}^{ntraits} l_1^2)(\sum_{n=1}^{ntraits} l_2^2)}}$), where $l_1$represents loadings of a factor in a replication and l_2 represents loadings of a similar factor in another replication and $ntraits$ is the number of traits contributing to a particular factor [119]. This similarity coefficient was calculated for all similar factors in the 100 replications (respectively 100*99/2=4950 times, as an average similarity measure of comparable factor configurations in the simulations (Supplemental Table 3).

The average correlations among eight metabolic traits and nine inflammatory biomarkers predict to some extent, especially via factor analyses, which are useful trait combinations that may reflect underlying MetS etiology, out of 130,305 possible trait combinations.

5. Correlated meta-analysis

We performed a correlations analysis of 8 metabolic traits and 9 inflammatory markers, as a premise in identifying useful combinations that may help in discovering genetic pleiotropy. Based on such analysis we had selected 8 trait combinations. This large number of results combined requires an unbiased method for meta-analyzing them. When meta-component scans are not independent, it can inflate type-I error, since at each location in the genome, a false-positive finding for one of the scans has an enhanced probability of being a false positive in any correlated scan. Province and Province and Borecki [113,114] developed a method for correcting bias via a correlated meta-analysis, which only requires the GWAS results and does not need the individual genotype / phenotype data. The basic idea is that for a trait of interest, the vast majority of the genome is under the null hypothesis of no genotype-phenotype association, which is only mildly contaminated with a relatively few SNPs that are under the alternative. Thus, the method performs sampling of GWAS genome via the polychoric correlation estimator [120], (using

SAS PROC FREQ). It is the estimate of the NxN correlation matrix, Σ between N scans, that is used to correct the final meta-estimates for this correlation.

In this article, the meta-analyses were based on p-values combinations, which involved the Fisher's 1925 [121] method of combining p-values at each location of the genome [122]. This technique uses the fact that for N scans, $\sum -2\ln(p_i) \sim \chi^2$ with 2n degrees of freedom, so the tail probability provides the meta-p-value. Unfortunately, in the case of correlated GWAS, this sum is no longer distributed as a simple chi-square. Instead, in the correlated meta-method, Province uses an inverse-normal transform, $Z_i = \varphi^{-1}(p_i)$ forming the N dimensional vector $\underline{Z}$ of all Z_is. He then applies the basic theorem of multidimensional statistics that for matrix $\underline{D}$, if $\underline{Z} \sim N(\underline{0},\Sigma)$ then $\underline{D}\,\underline{Z} \sim N(\underline{0},\underline{D}\Sigma\underline{D}')$. In particular, when D is a 1xN vector of all 1's, $SUM(\underline{Z}) = \underline{D}\,\underline{Z} \sim N(0,SUM(\Sigma))$, whose tail probability gives the meta-Z p-value. In this case, for estimating Σ, the SNP p-values are dichotomized across the genome as ($P \leq 0.5$; $P > 0.5$). The software was developed in SAS by Province M. A. [113] and an interface was built with SAS/InterNet to perform parallel computing of each meta-analysis within the Division of Statistical Genomics, Washington University computing cluster.

6. Bioinformatics of selected genes

Another approach we used for appraising pleiotropy was searching Gene Entrez of NCBI (http://www.ncbi.nlm.nih.gov/gene/) for genes related to each of the traits studied: "body mass index", "waist circumference", "high density lipoprotein cholesterol", "triglycerides", "insulin", "glucose", "systolic blood pressure", "diastolic blood pressure", "fibrinogen", "C-reactive protein", "plasminogen activator 1", "interleukin 6", "interleukin 10", "intercellular adhesion molecule 1", "tumor necrosis factor alpha", "adiponectin" and "white blood cell counts". Our search was limited only to human, mouse and rat species. Identified genes represent publication evidence of their contribution to a trait based on linkage, association, function, expression etc. All single traits gene lists were merged by gene name and selected for most contributions among metabolic traits and inflammatory biomarkers, selected with a minimum threshold of 8 contributions between the two of them (Supplemental Table 5).

For the same terms, searches were implemented also at http://www.genome.gov/gwastudies/. These data represent large genome wide studies with at least 100,000 SNPs and with a high statistical significance in the overall (initial GWAS + replication) population [123]. Genes identified as possible candidates were checked via Association Results Browser of dbGaP of NCBI http://www.ncbi.nlm.nih.gov/projects/gapplusprev/sgap_plus.htm. The same database was used to identify genes reported to associate with "metabolic syndrome". Results are reported in Supplemental Tables 5 and 6. The importance of gene lists identified was mined by means of GeneGO (http://thomsonreuters.com/products_services/science/systems-biology/) and Literature Lab of ACUMENTA (http://acumenta.com/) software. The GeneGO, enrichment analysis consists of matching gene IDs of possible targets for the "common", "similar" and "unique" sets with gene IDs in functional ontologies in MetaCore, MetaDrug, MetaBase, Specialty modules, and System toxicology. The probability of a random intersection between a set of IDs the size of target list with ontology entities is estimated in p-value of hypergeometric intersection. The lower p-value means higher relevance of the entity to the dataset, which shows in higher rating for the entity. Literature Lab on the other hand, is an interface between experimentally-derived gene lists and scientific literature in a curated vocabulary of 24,000 biological and biochemical

terms. It employs statistical and clustering analysis on over 14 million PubMed abstracts (01/01/90 to the present) to identify pathways (809 pathways), diseases, compounds, cell biology and other areas of biology and biochemistry. The analysis engine compares statistically the submitted gene set to 1,000 random gene sets generated on-the-fly to identify term relationships that are associated with the gene set more than by chance alone.

Acknowledgments (Supplement 3)

Box 1. A summary of 25 MetS candidate genes functions.

No*	Gene	Location	Function (References)	Annotating Marker	Allele (Frequency)
			Pleiotropic genes for lipids and inflammation		
1	*MACF1*	1p32-p31	"**Microtubule-actin crosslinking factor 1**"; Produces a protein that forms bridges between different cytoskeletal elements, by stabilizing and guiding microtubule growth along actin filaments. An alternative spliced form associates with the Golgi apparatus (35,36).	rs1537817	T (0.2156)
2	*KIAA0754*	1p34.3	An uncharacterized gene.	rs3768302	G (0.2859)
3	*GCKR*	2p23	"**Glucokinase (hexokinase 4) regulator**"; GCKR's protein is a regulatory protein that inhibits glucokinase in liver and pancreatic islet cells by binding non-covalently to form an inactive complex with the enzyme.	rs1260326	T (0.3963)
4	*GRB14*	2q22-q24	"**Growth factor receptor-bound protein 14**", which likely produces an inhibitory effect on insulin receptor signaling,	rs10184004	T (0.4214)
5	*COBLL1*	2q24.3	"**Cordon bleu**"; a conserved gene involved in neural tube formation (54)	rs10195252	C (0.4205)
6	*LOC646736*	2q36.3	An uncharacterized gene.	rs2943634	A (0.3428)
12	*BAZ1B*	7q11.23	"**Bromodomain adjacent to zinc finger domain, 1B**"; The bromodomain is a structural motif characteristic of proteins involved in chromatin-dependent regulation of transcription. This gene is deleted in Williams-Beuren syndrome.	rs7811265	G (0.191)
13	*BCL7B*	7q11.23	"**B-cell CLL/lymphoma 7B**"; This gene is located at a chromosomal region commonly deleted in Williams syndrome. This gene is highly conserved from C. elegans to human.	rs13233571	T (0.1209)
14	*TBL2*	7q11.23	"**Beta-transducin like 2**"; involved in regulatory functions. This protein is possibly involved in some intracellular signaling pathway. This gene is deleted in Williams-Beuren syndrome.	rs11974409	G (0.1906)
15	*MLXIPL*	7q11.23	"**Helix-loop-helix leucine zipper transcription factor of the Myc/Max/Mad superfamily**"; This protein forms a heterodimeric complex and binds and activates, in a glucose-dependent manner, carbohydrate response element (ChoRE) motifs in the promoters of triglyceride synthesis genes. The gene is deleted in Williams-Beuren syndrome.	rs17145750	T (0.1496)
16	*LPL*	8p22	"**Lipoprotein lipase**"; is expressed in heart, muscle and adipose tissues. Its main functions are the hydrolysis of triglycerides of circulating chylomicrons and very low density lipoproteins, and to serve as a ligand or bridging factor for receptor-mediated lipoprotein uptake. The apolipoprotein APOC2, acts as a coactivator of LPL in the presence of lipids on the luminal surface of vascular endothelium, whereas ANGPTL4 expression in adipose tissue as induced by fasting is proposed as an inhibitor of LPL in adipose tissue to reroute fat from adipose tissue to other tissues (76)	rs3289	C (0.028)
17	*TRIB1*	8q24.13	"**Tribbles pseudokinase 1**";	rs10808546	T (0.4425)
21	*ZNF664*	12q24.31	"**Zinc finger protein 664**";	rs12310367	G (0.3367)
25	*TOMM40*	19q13	"**Translocase of outer mitochondrial membrane 40 homolog (yeast)**"; channel-forming subunit of the translocase of the mitochondrial outer membrane (TOM) complex that is essential for protein import into mitochondria.	rs2075650	G (0.1533)
			Pleiotropic genes for adiposity / obesity and inflammation		
11	*TFAP2B*	6p12	"**Transcription factor AP-2 beta**"; *TFAP2B* is a transcription factor that stimulates cell proliferation.	rs3857599	A (0.1734)
19	*HECTD4*	12q24.13	"**HECT domain containing E3 ubiquitin protein ligase 4**";	rs11066188	A (0.4152)
20	*PTPN11*		"**Protein tyrosine phosphatase, non-receptor type 11**"; *PTPN11* produces a protein tyrosine phosphatase non-receptor 11 involved in cell growth, differentiation, and mitotic cycle.	rs11066320	A (0.421)
23	*FTO*	16q12.2	"**Fat mass and obesity associated**"; Studies in mice and humans indicate a role in nervous and cardiovascular systems and a strong association with body mass index, obesity risk, and type 2 diabetes	rs1558902	A (0.4163)
			Pleiotropic genes for adiposity / obesity, lipids and inflammation		
7	*SLC39A8*	4q22-q24	"**Solute carrier family 39, member 8**"; a solute carrier with structural characteristic of a zinc transporter. It is found in the plasma membrane and mitochondria, and functions in the cellular importation of zinc at the onset of inflammation.	rs13107325	T (0.0748)
8	*NELFE*	6p21.3	"**Negative elongation factor complex member E**"; Represses RNA polymerase II transcript elongation;Localizes to the major histocompatibility complex (MHC) class III region on chromosome 6.	rs419788	T (0.2954)
9	*SKIV2L*	6p21	"**Superkiller viralicidic activity 2-like**"; DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. Some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division.	rs437179	A (0.2956)
10	*STK19*	6p21.3	"**Serine/threonine kinase 19**"; it is possible that phosphorylation of this protein is involved in transcriptional regulation. This gene localizes to the major histocompatibility complex (MHC) class III region on chromosome 6	rs389883	G (0.2954)
18	*ATXN2*	12q24.1	"**Ataxin 2**"; The autosomal dominant cerebellar ataxias are a heterogeneous group of neurodegenerative disorders characterized by progressive degeneration of the cerebellum, brain stem and spinal cord.	rs653178	C (0.4687)
22	*PDXDC1*	16p13.11	"**Pyridoxal-dependent decarboxylase domain containing 1**";	rs4985155	G (0.3319)
24	*MC4R*	18q22	"**Melanocortin 4 receptor**"; A membrane-bound receptor and member of the melanocortin receptor family. Defects in this gene are a cause of autosomal dominant obesity.	rs6567160	C (0.2381)

* The corrsponding number matches with Table 2 order number (In table 2 this corresponds with ordering genes by chromosome and position).

Box 2. Supportive findings for known genes of the 25 MetS pleiotropic candidates.

Group 1: Pleiotropic genes for lipids and inflammatory markers

The ***MACF1*** was associated also with T2D [54]. Recently, Fassett *et al*. [55] using inducible cardiac-specific *MCF1* knockout mice concluded this gene works as a stress induced regulator of cardiomyocyte microtubule distribution and is important for ventricular adaptation to hemodynamic overload.

The **GCKR** rs1260326 was associated with T2D risk, by changing the ability of *GCKR* to sequester glucokinase in the nucleus of hepatocytes [56], and with hepatic fat accumulation along large VLDL and TG levels in obese youth [66]. Rees *et al* [56], suggested that leucine allele elevates hepatic glucose uptake and disposal by increasing active cytosolic *GCK*, which would increase hepatic lipid biosynthesis, but also by decreasing fasting plasma glucose concentrations, which may provide a protective effect on T2D risk. Another *GCKR* SNP was associated with serum albumin [42], decreased levels of amino acids alanine and isoleucine and elevated levels of glutamine [67], with liver enzyme gamma-Glutamyltrasferase [47], and platelet count [61]. *GCKR* associated with serum calcium [63]. *GCKR* already has been proposed as a candidate for MetS for its significant associations with qualitative bivariate TG-BP and WC-TG [62]. The rs2303369, neighboring *GCKR* and an intron of fibronectin type III (*FNDC4*) was associated significantly with menopause [68].

The ***GRB14*** protein has a pleckstrin homology domain, a C-terminal Src homology 2 (SH2) domain, and an intervening ~45 residues known as BPS. *GRB14* and its family members *GRB7* and *GRB10* are recruited by a number of receptor tyrosine kinases [105]. This recruitment is facilitated via phosphotyrosine binding the SH2 domain, while the *INS* and *IGF1* receptors are recruited by the BPS region [60]. Cooney *et al.* [59] noticed an improved glucose tolerance and an enhanced insulin-induced signaling in muscle and liver, but not in adipose tissue in a male mice deficient for *Grb14* ($^{-/-}$). They proposed that *Grb14* was a negative regulator, tissue specific for insulin signaling. In a gene expression study, *Grb14* expression was elevated in adipose tissue of both ob/ob mice and Goto-Kakizaki (non-obese T2D) rats [57]. Our meta-results add to the importance of *GRB14*, which can be viewed as an inhibitor of the insulin receptor and therefore as affecting insulin signaling.

The ***COBLL1*** [58] associated with T2D [54]. Adjacent to this gene toward *GRB14* are a number of SNPs that were associate with T2D [106], TG [23] and HDLC [23]. Albrechtsen *et al*. [54] showed that *COBLL1* expresses in pancreatic islets and kidney, and to some degree in skeletal muscle, liver and adipose tissue. They guessed *COBLL1* variants may influence expression of nearby *GRB14* to change insulin sensitivity.

The ***LOC646736*** rs2943634 was associated with coronary disease [65] and T2D [64]. Downstream (~47K bps) from this SNP, an intron of *LOC646736* was associated with T2D [80]. Upstream of our meta-SNP, a few SNPs associates with TG [23], with adiposity [76], and with ADIP [30].

The ***LPL*** is significantly associated with TG and HDLC [23,39,42-45,62,69-73,75,77-79,81]. *LPL* is part of glycerolipid metabolism pathway (map00561, kegg.jp), involved in free fatty acids production, and is also a member of *PPAR* signaling pathway (map03320, kegg.jp).

TRIB1 is reported in associations with TG, HDLC, LDLC [23], with alkaline phosphatase and alanine transaminase [47], with ADIP [30], with Crohn's Disease [83], with bivariate qualitative combinations of HDLC-TG and TG-BP [62]. Recently Akira *et al.* [82] working with *Trib1*$^{(-/-)}$ mice demonstrated that mice lacking Trib1 in hematopoietic cells exhibited severe lipodystrophy due to increased lipolysis, while in a high-fat diet, mice exhibited hypertriglyceridemia, insulin resistance, together with increased proinflammatory cytokine production. They suggested, that *Trib1* is critical for adipose tissue maintenance and suppression of metabolic disorders by controlling the differentiation of tissue-resident anti-inflammatory-like macrophages. The rs10808546 positioned about 45K bps from *TRIB1* is located in a DNAase mark often found in active regulatory elements.

Fox *et al.* [85] report association of ***ZNF664*** with visceral adipose tissue adjusted for BMI and with visceral adipose tissue/subcutaneous adipose tissue ratio for women.

The ***TOMM40*** is positioned at the side of the cluster *APOE*/*APOC4*/*APOC2* and was associated with Alzheimer's disease [86,96], low density lipoprotein cholesterol (LDLC) and HDLC [69], and CRP [69,95]. The rs2075650 of *TOMM40* is part of three signatures of promoter histone marks, part of enhancer histone markers in 6 cell types, it can be involved in a DNase signature, and is part of 8 changed motifs, among them sterol regulatory element binding transcription factor (SREBP).

Box 2 (Continued). Supportive findings for known genes of the 25 MetS pleiotropic candidates.

Group2: Pleiotropic genes for adiposity / obesity and inflammation

An intron of ***TFAP2B***, was associated with the effects of dietary fat intake on weight loss and waist reduction [99]. A few other SNPs of *TFAP2B* associated significantly with BMI [22], adiposity [90] and with a qualitative bivariate WAIST-GLUC combination [62].

The ***PTPN11*** was associated with platelet counts [97], with TG [74], and with carotid arteries [94].

While ***FTO*** contributes to the regulation of the global metabolic rate, energy expenditure, energy homeostasis, regulation of body size and body fat accumulation, its exact function is not known. Other SNPs of *FTO* were associated with BMI [22], body weight [101], adiposity [76], WAIST [87], with T2D [103] and less so with factor1 and factor2 of MetS risk factors [38].

Group 3: Pleiotropic genes for adiposity / obesity, lipids and inflammation

The ***SLC39A8*** protein is found in the plasma membrane and mitochondria, and functions in the cellular transport of of zinc at the onset of inflammation. *SLC39A8* is a negative regulator of *NF-κB* and functions to negatively regulate proinflammatory responses through zinc-mediated down-modulation of IκB kinase (*IKK*) activity [91]. *SLC39A8* and *SLC39A14* are regulated by IL-6 dependent signaling in the liver [92]. In addition, rs230487, which is closer to *NFKB1* than *SLC39A8* was associated with tissue Plasminogen activator [102]. Liu *et al* [91] proposed that *SLC39A8* and *SLC39A14* are important zinc transporters that channel zinc in a tissue-specific manner to fundamentally important intracellular checkpoints, which help to coordinate and balance host defense.

The ***NELFE***, ***SKIV2L*** and ***STK19*** position in the class III region of the major histocompatibility complex of chromosome 6. The three genes are likely involved in transcription regulation and have been found to be associated with Macular Degeneration and Lupus Erythematosus, and rs2072633, an intron of *CFB* – complement factor B, (but only 286 bps from *NELFE* gene) [88] being associated with Multiple Sclerosis.

The association of ***PDXDC1*** with ADIP may indicate that its pleiotropic effect could have protective contributions for inflammation and MetS. Based on the ENCODE information the rs4985155 is located in a transcription factor binding site and to a DNase peak. *The* rs4500751, (chr16:15140211) mapped at *NTAN1* about 10.7K bps from our PDXDC1 meta-SNP, associated with absolute plasma levels and proportions of the phospholipid species with important roles in cell survival and inflammation [84]. Other SNPs associated with blood metabolite concentration [100], and with phospholipids levels in plasma [89].

The ***MC4R*** is a member of melanocortin family. The melanocortins are involved in pigmentation, energy homeostasis, inflammation, immunomodulation, steroidogenesis and temperature control. Stäubert *et al.* [98] found a strong correlation between positional conservation and the functional relevance of missense, nonsense, and frame-shifting mutations of *MC4R* affecting 60 amino acid positions. The mostly heterozygous (dominant) occurring *MC4R* mutations are implicated in 1–6% of early-onset or severe adult obesity cases. Some of the GWAS findings indicated that *MC4R* was associated with BMI [22,29]), obesity [93], body height [104], with body weight [101], WAIST [70], and with HDLC [23].

Tables and Figures

Table 1.a. XC-Pleiotropy Consortium studies for assessing associations among MetS and inflammatory markers and identifying promising trait combinations for evaluating the role of pleiotropy in MetS etiology.

No	Participating studies	Acronym	Cohorts
1	The Atherosclerosis Risk in Communities Study	ARIC	AA and EA
2	The Coronary Artery Risk Development in Young Adults	CARDIA	EA
3	The Johns Hopkins Genetic Study of Atherosclerosis Risk	GeneSTAR	AA and EA
4	The Genetic Epidemiology Network of Arteriopathy	GENOA	AA and EA
5	The Family Heart Study	FamHS	EA
6	The Framingham Heart Study	FHS	EA
7	The INTER99	INTER99	EA
8	The LifeLines Cohort Study		EA
9	The Rotterdam Study	RS	EA
10	The Women's Genome Health Study	WGHS	EA
11	The Women's Health Initiative	WHI	EA
Note: The addition of a suffix –AA in the study name refers to an African American cohort, and –EA refers to a European ancestry cohort.			

Table 1.b. Sources of association tests results analyzed in our 9 correlated meta-analyses.

No	Contributing studies	Acronym	Traits	Studies (N)	Participants (N)	SNPs (N)	Reference
1	The Genetic Investigation of Anthropometric Traits Consortium	GIANT	BMI, WAIST	28	~124,000	~2.5M	22,29
2	The Global Lipids Genetics Consortium	GLGC	HDLC, TG	46	~99,000	~2.5M	23
3	The Meta-Analyses of Glucose and Insulin-related traits	MAGIC	GLUC, INS	21	~46,000, 38,000	~2.5M	19
4	The Global BPgen	GBPG	SBP, DBP	17	~34,000	~2.5M	21
5	The Cohorts of the Heart and Aging Research in Genomic Epidemiology Consortium	CHARGE					
6	and The European Special Population Network	EUROSPAN	CRP	15	~66,185	~2.5M	24
7	and six independent studies						
8	Independent cohorts of European-ancestry		PAI-1	8	~19,599	~2.5M	25
9	The Cohorts of the Heart and Aging Research in Genomic Epidemiology Consortium	CHARGE	WBCC	7	~19509	~2.5M	26
10	ADIPOGen Consortium	ADIPOGen	ADIP	23	~35,355	~2.5M	30
11	The Women's Genome Health Study	WGHS	ICAM-1	1	2,435	~0.3M	27
12	The Howard University Family Study	HUFS	IL-6	1	707	~5.0M	31

Table 2. Meta-analyses results of 9 classes of trait-combinations.

Footnote: Selected are best SNPs per gene with up to three possibilities, within a gene, up to 5 KB from the nearest gene or beyond 5KB to the nearest gene. A SNP to be selected had to fulfill the following conditions: meta -$\log_{10}p \geq 8$ and at least one metabolic trait and one inflammatory marker with -$\log_{10}p \geq 3$. Table 1 represents SNPs that can be considered as contributors to MetS. Yellow color shows SNPs that pass or reach a threshold of -$\log_{10}p \geq 3$, orange color marks SNPs that pass or reach a threshold of -$\log_{10}p \geq 8$, and green indicates SNPs that might show some protective effect against MetS (see Discussion for clusters of ADIP and HDLC). Note: A particular SNP is intergenic and marked with ()_beyond when its location was more than 5K bps; rs- rsname, chrom- chromosome, position in bps, meta_nlog10p – meta -$\log_{10}p$, BMI, WAIST, HDLC, TG, GLUC, INS, SBP, DBP, CRP, PAI1, IL6, ICAM1, ADIP, and WBCC represent -$\log_{10}p$ values for each trait, hugo –gene symbol, role –SNPs role, diffPosNearGene is a distance in bps from the start or end SNP of the closest gene, while 0 distance when within a gene, newhugo –the closest gene to the reported SNP.

No	Trait combination	rs	chrom	position	meta_nlog10p	BMI	WAIST	HDLC	TG	GLUC	INS	SBP	DBP	CRP	PAI1	IL6	ICAM1	ADIP	WBCC	hugo	role	diffPosNearGene	newhugo
1	1. bwhtgisd_rp	rs1537817	1	39639653	17.71	1.52	3.24	8.94	5.14	2.99	1.28	2.47	1.93	6.33						MACF1	intron-vari	0	MACF1
2	1. bwhtgisd_rp	rs3768302	1	39880319	15.73	1.15	2.72	8.72	4.90	2.95	0.89	2.20	1.65	6.56						KIAA0754	utr-variant	0	KIAA0754
3	6. ht_rpcc	rs1260326	2	27730940	78.71			1.11	132.25					42.26					0.23	GCKR	missense	0	GCKR
4	7. ht_i1ip	rs10184004	2	165508389	18.21			6.98	9.76						2.54			4.52				28106	(GRB14)_beyond
5	7. ht_i1ip	rs10195252	2	165513091	18.33			7.03	9.79						2.65			4.44				-27709	(COBLL1)_beyond
6	7. ht_i1ip	rs2943634	2	227068080	15.59			8.63	7.29						0.96			5.22				22841	(LOC646736)_beyond
7	9. bwhtgisdt2d_rpi1l6m1ipcc	rs13107325	4	103188709	13.27	6.86	3.16	10.14	1.82	0.18	0.40	3.91	4.18	0.36	0.48		1.87	4.13	0.15	SLC39A8	missense	0	SLC39A8
8	9. bwhtgisdt2d_rpi1l6m1ipcc	rs419788	6	31928799	12.72	4.48	2.52	0.07	13.56	0.14	0.71	3.25	0.82	1.65	3.07	0.83	1.06	1.20	3.71	NELFE	upstream-	0	NELFE
9	9. bwhtgisdt2d_rpi1l6m1ipcc	rs437179	6	31929014	12.50	4.41	2.54	0.11	13.46	0.20	0.64	3.28	0.84	1.49	2.89	0.83	1.09	1.07	3.24	SKIV2L	missense	0	SKIV2L
10	9. bwhtgisdt2d_rpi1l6m1ipcc	rs389883	6	31947460	13.49	4.43	2.46	0.24	14.40	0.19	0.90	3.74	0.99	1.43	3.05		0.94	1.16	3.06	STK19	intron-vari	0	STK19
11	5. bw_rpl6cc	rs3857599	6	50938247	15.42	13.58	10.21							3.64					0.54			122468	(TFAP2B)_beyond
12	6. ht_rpcc	rs7811265	7	72934510	37.67			5.92	58.04					7.25					0.70	BAZ1B	intron-vari	0	BAZ1B
13	6. ht_rpcc	rs13233571	7	72971231	35.49			8.54	57.03					7.55					0.07	BCL7B	intron-vari	0	BCL7B
14	6. ht_rpcc	rs11974409	7	72989390	35.79			5.49	57.90					6.94					0.51	TBL2	intron-vari	0	TBL2
15	6. ht_rpcc	rs17145750	7	73026378	36.94			6.82	57.80					6.33					0.56	MLXIPL	intron-vari	0	MLXIPL
16	6. ht_rpcc	rs3289	8	19823192	32.66			26.70	18.94					3.60					1.44	LPL	utr-variant	0	LPL
17	7. ht_i1ip	rs10808546	8	126495818	51.41			18.20	53.42						2.94			4.60				44737	(TRIB1)_beyond
18	9. bwhtgisdt2d_rpi1l6m1ipcc	rs653178	12	112007756	14.55	3.83	3.48	5.80	0.69	0.36	0.26	3.43	6.71	0.44	0.43	2.12	16.50	0.02	1.60	ATXN2	intron-vari	0	ATXN2
19	9. bwhtgisdt2d_rpi1l6m1ipcc	rs11066188	12	112610714	9.16	4.01	3.62	2.69	0.12	0.35	0.13	3.52	5.90	0.33	0.07		11.36	0.17	1.93	HECTD4	intron-vari	0	HECTD4
20	9. bwhtgisdt2d_rpi1l6m1ipcc	rs11066320	12	112906415	8.97	3.83	3.24	2.70	0.24	0.34	0.06	3.70	5.75	0.44	0.22		9.41	0.28	1.19	PTPN11	intron-vari	0	PTPN11
21	7. ht_i1ip	rs12310367	12	124486678	15.55			9.51	7.92						0.14			7.94		ZNF664	intron-vari	0	ZNF664
22	3. whti_ipcc	rs4985155	16	15129459	8.23		5.00	1.66	4.92		0.58							4.11	0.22	PDXDC1	intron-vari	0	PDXDC1
23	4. bwi_rpi1	rs1558902	16	53803574	60.99	61.69	49.38				4.12			5.65	1.41					FTO	intron-vari	0	FTO
24	1. bwhtgisd_rp	rs6567160	18	57829135	24.58	21.74	18.08	7.91	4.75	0.34	1.79	0.75	0.64	3.82								-208947	(MC4R)_beyond
25	6. ht_rpcc	rs2075650	19	45395619	67.63			15.96	18.88					86.52					0.16	TOMM40	intron-vari	0	TOMM40

Table 3. Average correlations and their lower and upper r estimates for 100 replications of simulated metabolic traits and inflammatory biomarkers (immitating 100 sets of 14 cohorts real studies' data, p=17, N=85,523) simulated with missing values (simulation 1, see Methods.3).

Correlations of Metabolic traits

bmi	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	1	0.844	-0.336	0.306	0.510	0.282	0.293	0.263
sd	0	0.001	0.003	0.003	0.003	0.004	0.003	0.003
min	1	0.841	-0.344	0.299	0.502	0.272	0.286	0.257
max	1	0.846	-0.328	0.315	0.519	0.293	0.300	0.272
waist	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	0.844	1	-0.336	0.321	0.513	0.279	0.276	0.247
sd	0.001	0	0.003	0.003	0.003	0.004	0.003	0.003
min	0.841	1	-0.344	0.313	0.505	0.269	0.268	0.239
max	0.846	1	-0.328	0.327	0.523	0.288	0.284	0.254
hdlc	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	-0.336	-0.336	1	-0.481	-0.359	-0.175	-0.107	-0.091
sd	0.003	0.003	0	0.004	0.003	0.004	0.004	0.004
min	-0.344	-0.344	1	-0.490	-0.367	-0.183	-0.117	-0.101
max	-0.328	-0.328	1	-0.473	-0.350	-0.168	-0.101	-0.082
tg	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	0.306	0.321	-0.481	1	0.376	0.208	0.202	0.185
sd	0.003	0.003	0.004	0	0.004	0.004	0.004	0.004
min	0.299	0.313	-0.490	1	0.366	0.196	0.193	0.175
max	0.315	0.327	-0.473	1	0.386	0.218	0.210	0.193
ins	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	0.510	0.513	-0.359	0.376	1	0.355	0.209	0.205
sd	0.003	0.003	0.003	0.004	0	0.004	0.004	0.005
min	0.502	0.505	-0.367	0.366	1	0.342	0.199	0.194
max	0.519	0.523	-0.350	0.386	1	0.367	0.223	0.214
gluc	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	0.282	0.279	-0.175	0.208	0.355	1	0.185	0.138
sd	0.004	0.004	0.004	0.004	0.004	0	0.004	0.004
min	0.272	0.269	-0.183	0.196	0.342	1	0.176	0.125
max	0.293	0.288	-0.168	0.218	0.367	1	0.199	0.150
sbp	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	0.293	0.276	-0.107	0.202	0.209	0.185	1	0.742
sd	0.003	0.003	0.004	0.004	0.004	0.004	0	0.001
min	0.286	0.268	-0.117	0.193	0.199	0.176	1	0.737
max	0.300	0.284	-0.101	0.210	0.223	0.199	1	0.745
dbp	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	0.263	0.247	-0.091	0.185	0.205	0.138	0.742	1
sd	0.003	0.003	0.004	0.004	0.005	0.004	0.001	0
min	0.257	0.239	-0.101	0.175	0.194	0.125	0.737	1
max	0.272	0.254	-0.082	0.193	0.214	0.150	0.745	1

Note: correlation mean of 100 simulated replications, sd- standard deviation, min-minimum, and max- maximum of correlation in 100 replications

Highlighted in yellow correlation coefficents ~(0.2 - < 0.4), in orange ~(0.4-<0.6) and in red ≥ 0.6

Correlations of Metabolic traits and Inflammatory markers

fib	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.254**	**0.253**	**-0.200**	**0.105**	**0.147**	**0.079**	**0.105**	**0.073**
sd	0.004	0.004	0.004	0.005	0.004	0.005	0.004	0.004
min	**0.244**	**0.243**	**-0.209**	**0.093**	**0.137**	**0.067**	**0.098**	**0.064**
max	**0.263**	**0.264**	**-0.189**	**0.116**	**0.159**	**0.090**	**0.117**	**0.087**
crp	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.406**	**0.380**	**-0.196**	**0.275**	**0.259**	**0.154**	**0.191**	**0.156**
sd	0.003	0.003	0.004	0.010	0.005	0.004	0.003	0.004
min	**0.399**	**0.374**	**-0.204**	**0.255**	**0.249**	**0.141**	**0.184**	0.148
max	**0.415**	**0.387**	**-0.187**	**0.290**	**0.273**	**0.166**	**0.199**	**0.166**
pai1	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.444**	**0.447**	**-0.372**	**0.389**	**0.497**	**0.332**	**0.176**	**0.152**
sd	0.008	0.008	0.009	0.008	0.008	0.009	0.009	0.009
min	**0.421**	**0.430**	**-0.391**	**0.372**	**0.480**	**0.304**	**0.159**	**0.129**
max	**0.470**	**0.473**	**-0.345**	**0.405**	**0.520**	**0.349**	**0.204**	**0.173**
il6	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.285**	**0.290**	**-0.175**	**0.140**	**0.192**	**0.126**	**0.129**	**0.090**
sd	0.008	0.008	0.007	0.008	0.007	0.008	0.009	0.008
min	**0.269**	**0.276**	**-0.192**	**0.121**	**0.176**	**0.108**	**0.111**	**0.070**
max	**0.308**	**0.312**	**-0.158**	**0.157**	**0.212**	**0.146**	**0.149**	**0.114**
tnfa	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.098**	**0.094**	**-0.178**	**0.135**	**0.105**	**0.072**	**0.042**	**0.047**
sd	0.011	0.011	0.009	0.010	0.009	0.009	0.009	0.010
min	**0.074**	**0.069**	**-0.208**	**0.110**	**0.085**	**0.051**	**0.014**	**0.022**
max	**0.125**	**0.121**	**-0.154**	**0.159**	**0.130**	**0.097**	**0.068**	**0.065**
icam1	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.163**	**0.176**	**-0.210**	**0.171**	**0.197**	**0.106**	**0.101**	**0.077**
sd	0.005	0.004	0.005	0.007	0.007	0.007	0.005	0.005
min	**0.154**	**0.165**	**-0.228**	**0.154**	**0.179**	**0.091**	**0.084**	**0.066**
max	**0.174**	**0.189**	**-0.201**	**0.183**	**0.219**	**0.126**	**0.113**	**0.091**
il10	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.001**	**0.032**	**-0.069**	**-0.011**	**0.041**	**0.019**	**-0.001**	**0.011**
sd	0.017	0.018	0.016	0.017	0.014	0.014	0.015	0.016
min	**-0.041**	**-0.005**	**-0.110**	**-0.048**	**0.008**	**-0.011**	**-0.036**	**-0.023**
max	**0.055**	**0.087**	**-0.026**	**0.049**	**0.087**	**0.055**	**0.031**	**0.050**
adip	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**-0.233**	**-0.244**	**0.399**	**-0.331**	**-0.319**	**-0.195**	**-0.075**	**-0.055**
sd	0.006	0.006	0.005	0.005	0.006	0.007	0.007	0.006
min	**-0.250**	**-0.255**	**0.385**	**-0.344**	**-0.332**	**-0.210**	**-0.091**	**-0.066**
max	**-0.219**	**-0.226**	**0.413**	**-0.317**	**-0.300**	**-0.177**	**-0.060**	**-0.039**
wbbc	BMI	WAIST	HDLC	TG	INS	GLUC	SBP	DBP
mean	**0.146**	**0.175**	**-0.195**	**0.236**	**0.168**	**0.095**	**0.120**	**0.067**
sd	0.004	0.004	0.004	0.004	0.005	0.004	0.005	0.005
min	**0.133**	**0.165**	**-0.207**	**0.227**	**0.156**	**0.086**	**0.104**	**0.056**
max	**0.153**	**0.184**	**-0.182**	**0.247**	**0.181**	**0.104**	**0.134**	**0.079**

Correlations of Inflammatory markers

fib	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	1	0.442	0.150	0.331	0.101	0.229	0.099	-0.126	0.291
sd	0	0.004	0.010	0.007	0.009	0.005	0.014	0.009	0.005
min	1	0.433	0.128	0.317	0.075	0.219	0.065	-0.147	0.279
max	1	0.452	0.171	0.347	0.119	0.239	0.145	-0.102	0.309
crp	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	0.442	1	0.268	0.416	0.135	0.266	0.200	-0.082	0.319
sd	0.004	0	0.008	0.008	0.010	0.004	0.015	0.006	0.005
min	0.433	1	0.247	0.398	0.110	0.257	0.157	-0.097	0.306
max	0.452	1	0.287	0.439	0.157	0.274	0.239	-0.064	0.333
pai1	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	0.150	0.268	1	0.141	0.160	0.210	0.060	-0.353	0.160
sd	0.010	0.008	0	0.029	0.017	0.010	0.017	0.014	0.008
min	0.128	0.247	1	0.070	0.123	0.188	0.022	-0.384	0.142
max	0.171	0.287	1	0.202	0.210	0.232	0.121	-0.309	0.180
il6	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	0.331	0.416	0.141	1	0.251	0.247	.	-0.130	0.234
sd	0.007	0.008	0.029	0	0.011	0.010	.	0.009	0.015
min	0.317	0.398	0.070	1	0.223	0.220	.	-0.155	0.207
max	0.347	0.439	0.202	1	0.274	0.271	.	-0.112	0.280
tnfa	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	0.101	0.135	0.160	0.251	1	0.253	0.099	-0.060	0.058
sd	0.009	0.010	0.017	0.011	0	0.009	0.015	0.010	0.016
min	0.075	0.110	0.123	0.223	1	0.225	0.054	-0.088	0.024
max	0.119	0.157	0.210	0.274	1	0.277	0.134	-0.036	0.099
icam1	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	**0.229**	0.266	0.210	0.247	0.253	1	0.179	-0.050	0.173
sd	0.005	0.004	0.010	0.010	0.009	0	0.014	0.007	0.009
min	0.219	0.257	0.188	0.220	0.225	1	0.140	-0.064	0.150
max	0.239	0.274	0.232	0.271	0.277	1	0.209	-0.035	0.195
il10	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	**0.099**	0.200	0.060	.	0.099	0.179	1	-0.019	0.049
sd	0.014	0.015	0.017	.	0.015	0.014	0	0.015	0.018
min	0.065	0.157	0.022	.	0.054	0.140	1	-0.052	0.005
max	0.145	0.239	0.121	.	0.134	0.209	1	0.019	0.088
adip	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	-0.126	-0.082	-0.353	-0.130	-0.060	-0.050	-0.019	1	-0.219
sd	0.009	0.006	0.014	0.009	0.010	0.007	0.015	0	0.015
min	-0.147	-0.097	-0.384	-0.155	-0.088	-0.064	-0.052	1	-0.246
max	-0.102	-0.064	-0.309	-0.112	-0.036	-0.035	0.019	1	-0.181
wbbc	FIB	CRP	PAI1	IL6	TNFA	ICAM1	IL10	ADIP	WBCC
mean	0.291	0.319	0.160	0.234	0.058	0.173	0.049	-0.219	1
sd	0.005	0.005	0.008	0.015	0.016	0.009	0.018	0.015	0
min	0.279	0.306	0.142	0.207	0.024	0.150	0.005	-0.246	1
max	0.309	0.333	0.180	0.280	0.099	0.195	0.088	-0.181	1

Figure 1. Prevalence of MetS and its components and mean levels of inflammatory biomarkers in individuals classified with and without MetS.
Footnote: Top histogram numbers represent prevalence (%) of MetS, T2D and MetS components. Bottom numbers represent number of participants for a particular trait. The biomarker boxplot graph comparisons were built by using "rnorm" function in R with mean, standard deviation and sample size corresponding to subgroups with and without MetS from original (B) data. Overall, they represent 53 tests of biomarkers per MetS strata, summarized in Supplemental Figure 1(a-g). The number within each pair of boxplots marked by "D=" is the difference of two means of an inflammatory biomarker in groups of participants classified with versus without MetS (M_1 vs. M_0). The light yellow boxed number at the bottom of the same graph marked with "p_t=" represents a p-value calculated by pooled t-test for testing if their means (M_1 vs. M_0) are different. In case the color of p_t-value box is gray, then the p-value does not pass the Bonferroni threshold p=9.43e-04.

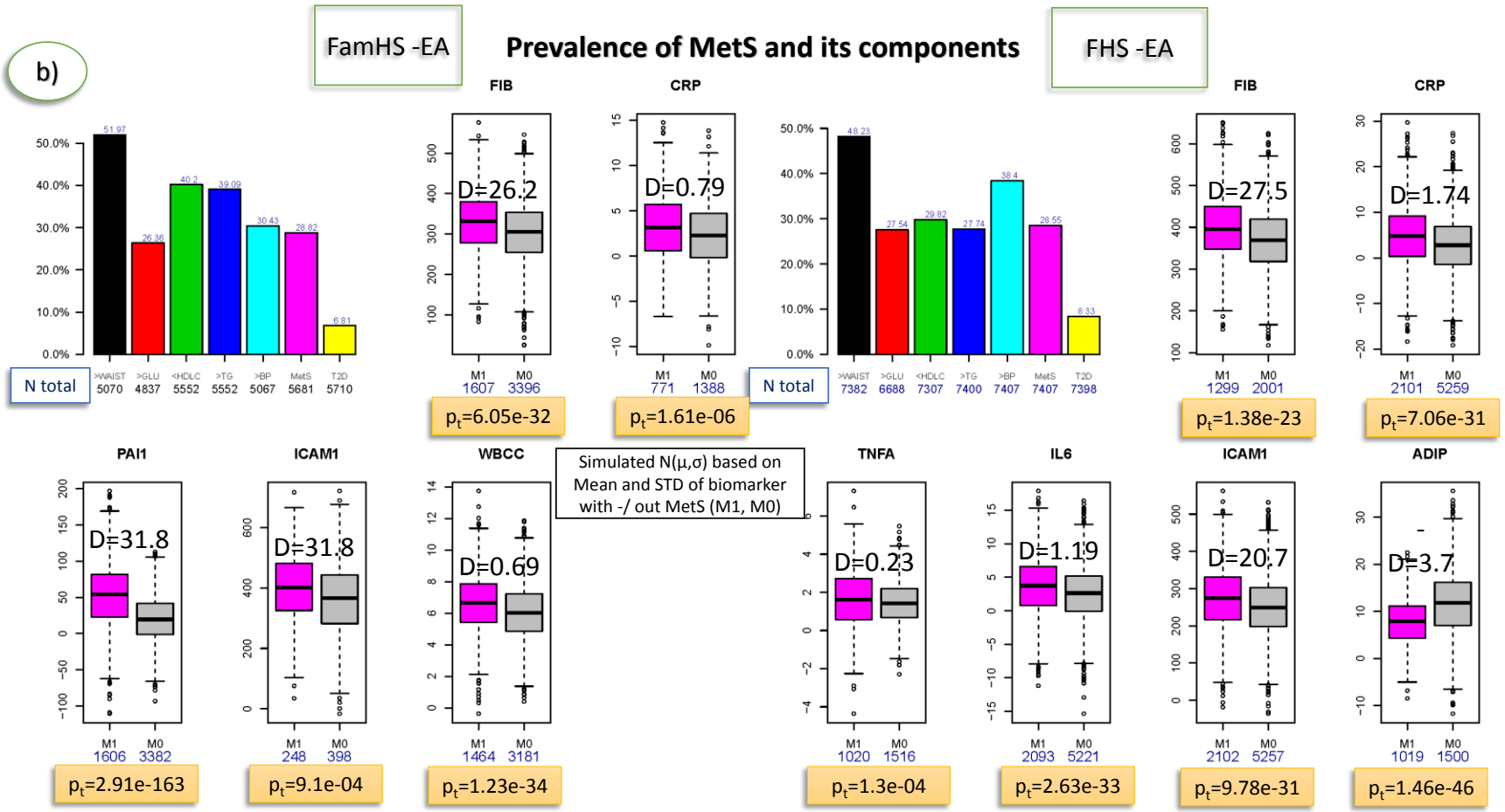
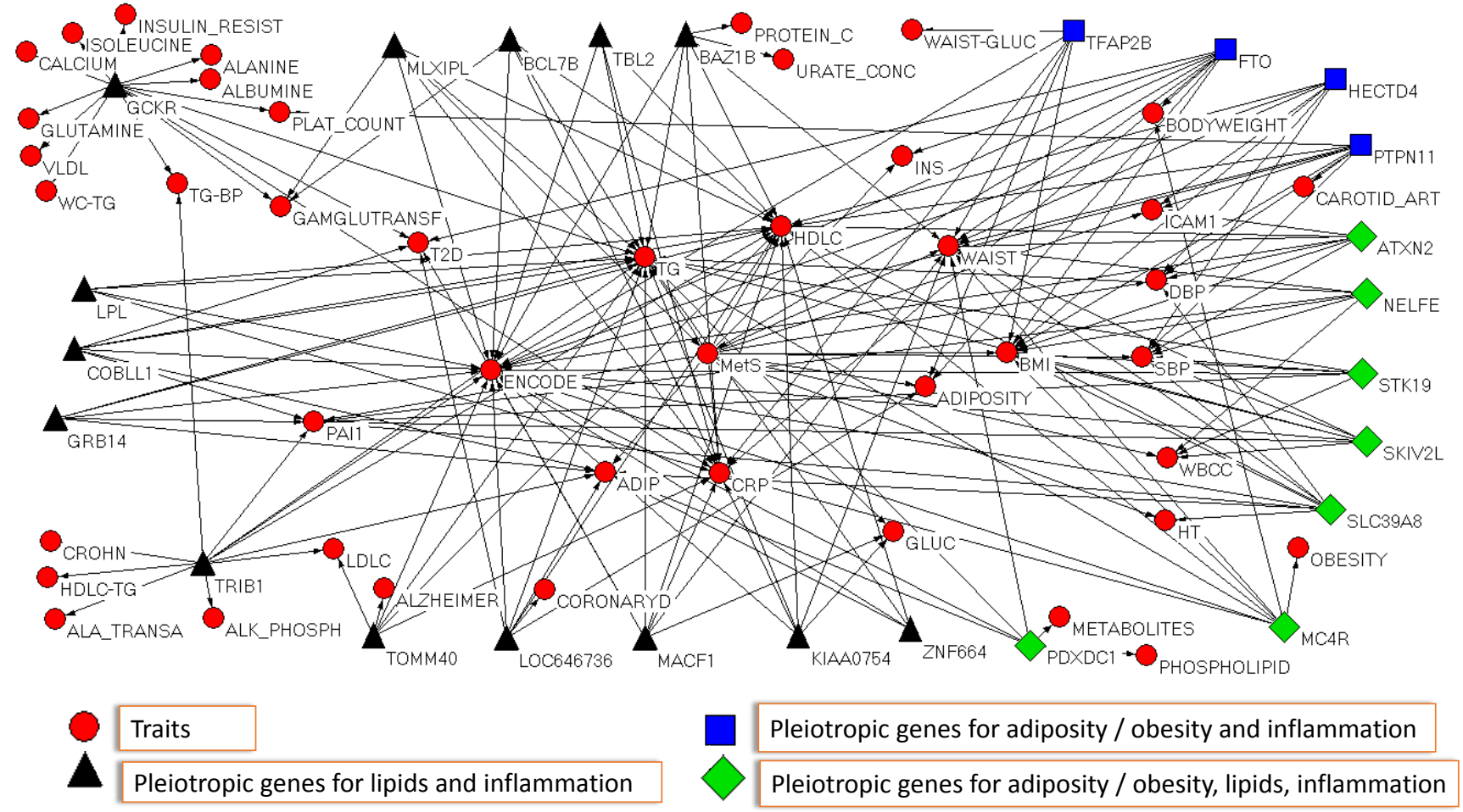

b)
FamHS -EA
Prevalence of MetS and its components
FHS -EA
FIB
CRP
D=26.2
D=0.79
N total
>WAIST >GLU <HDLC >TG >BP MetS T2D
5070 4837 5552 5552 5067 5681 5710
M1 M0
1607 3396
771 1388
p_t=6.05e-32
p_t=1.61e-06
N total
7382 6688 7307 7400 7407 7407 7398
D=27.5
D=1.74
1299 2001
2101 5259
p_t=1.38e-23
p_t=7.06e-31
PAI1
ICAM1
WBCC
D=31.8
D=31.8
D=0.69
1606 3382
248 398
1464 3181
p_t=2.91e-163
p_t=9.1e-04
p_t=1.23e-34
Simulated N(μ,σ) based on Mean and STD of biomarker with -/ out MetS (M1, M0)
TNFA
IL6
ICAM1
ADIP
D=0.23
D=1.19
D=20.7
D=3.7
1020 1516
2093 5221
2102 5257
1019 1500
p_t=1.3e-04
p_t=2.63e-33
p_t=9.78e-31
p_t=1.46e-46

Figure 2. A network of 25 pleiotropic genes with hypothetical contributions to MetS, including inflammation.
Footnote: In the figure they connect by GWAS phenotypic evidence and whether selected SNPs show any regulatory features based on the ENCODE database as implemented via HaploReg [32] / regulomeDB [33] software (All phenotypic labels correspond to associations reported in the Results, Discussion, Box 2 and Suppplemental Table 5).



References

1. Grundy, S.M. *et al.* Clinical management of metabolic syndrome: report of the American Heart Association/National Heart, Lung, and Blood Institute/American Diabetes Association conference on scientific issues related to management. *Circulation* **109**, 551-6 (2004).
2. Galassi, A., Reynolds, K. & He, J. Metabolic syndrome and risk of cardiovascular disease: a meta-analysis. *Am J Med* **119**, 812-9 (2006).
3. Monda, K.L. *et al.* The genetics of obesity and the metabolic syndrome. *Endocr Metab Immune Disord Drug Targets* **10**, 86-108 (2010).
4. Eckel, R.H., Grundy, S.M. & Zimmet, P.Z. The metabolic syndrome. *Lancet* **365**, 1415-28 (2005).
5. Lusis, A.J., Attie, A.D. & Reue, K. Metabolic syndrome: from epidemiology to systems biology. *Nat Rev Genet* **9**, 819-30 (2008).
6. Djousse, L., Padilla, H., Nelson, T.L., Gaziano, J.M. & Mukamal, K.J. Diet and metabolic syndrome. *Endocr Metab Immune Disord Drug Targets* **10**, 124-37 (2010).
7. Katzmarzyk, P.T. *et al.* Targeting the metabolic syndrome with exercise: evidence from the HERITAGE Family Study. *Med Sci Sports Exerc* **35**, 1703-9 (2003).
8. Johnson, L.W. & Weinstock, R.S. The metabolic syndrome: concepts and controversy. *Mayo Clin Proc* **81**, 1615-20 (2006).
9. Kahn, R. *et al.* The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* **28**, 2289-304 (2005).
10. Dallmeier, D. *et al.* Metabolic syndrome and inflammatory biomarkers: a community-based cross-sectional study at the Framingham Heart Study. *Diabetol Metab Syndr* **4**, 28 (2012).
11. Henneman, P. *et al.* Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: the Erasmus Rucphen Family study. *J Med Genet* **45**, 572-7 (2008).
12. Kraja, A.T. *et al.* Do inflammation and procoagulation biomarkers contribute to the metabolic syndrome cluster? *Nutr Metab (Lond)* **4**, 28 (2007).
13. Wannamethee, S.G. *et al.* The metabolic syndrome and insulin resistance: relationship to haemostatic and inflammatory markers in older non-diabetic men. *Atherosclerosis* **181**, 101-8 (2005).
14. Ridker, P.M., Wilson, P.W. & Grundy, S.M. Should C-reactive protein be added to metabolic syndrome and to assessment of global cardiovascular risk? *Circulation* **109**, 2818-25 (2004).
15. Hung, J., McQuillan, B.M., Thompson, P.L. & Beilby, J.P. Circulating adiponectin levels associate with inflammatory markers, insulin resistance and metabolic syndrome independent of obesity. *Int J Obes (Lond)* **32**, 772-9 (2008).
16. Matsushita, K. *et al.* Comparison of circulating adiponectin and proinflammatory markers regarding their association with metabolic syndrome in Japanese men. *Arterioscler Thromb Vasc Biol* **26**, 871-6 (2006).
17. Sakkinen, P.A., Wahl, P., Cushman, M., Lewis, M.R. & Tracy, R.P. Clustering of procoagulation, inflammation, and fibrinolysis variables with metabolic factors in insulin resistance syndrome. *Am J Epidemiol* **152**, 897-907 (2000).
18. Espinola-Klein, C., Gori, T., Blankenberg, S. & Munzel, T. Inflammatory markers and cardiovascular risk in the metabolic syndrome. *Front Biosci (Landmark Ed)* **16**, 1663-74 (2011).
19. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
20. Ingelsson, E. *et al.* Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. *Diabetes* **59**, 1266-75 (2010).

21. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* **41**, 666-76 (2009).
22. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
23. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-13 (2010).
24. Dehghan, A. *et al.* Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* **123**, 731-8 (2011).
25. Huang, J. *et al.* Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. *Blood* **120**, 4873-81 (2012).
26. Nalls, M.A. *et al.* Multiple loci are associated with white blood cell phenotypes. *PLoS Genet* **7**, e1002113 (2011).
27. Pare, G. *et al.* Genome-wide association analysis of soluble ICAM-1 concentration reveals novel associations at the NFKBIK, PNPLA3, RELA, and SH2B3 loci. *PLoS Genet* **7**, e1001374 (2011).
28. Richards, J.B. *et al.* A genome-wide association study reveals variants in ARL15 that influence adiponectin levels. *PLoS Genet* **5**, e1000768 (2009).
29. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* **41**, 25-34 (2009).
30. Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet* **8**, e1002607 (2012).
31. Tekola Ayele, F. *et al.* Genome-wide associated loci influencing interleukin (IL)-10, IL-1Ra, and IL-6 levels in African Americans. *Immunogenetics* **64**, 351-9 (2012).
32. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
33. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
34. Kodama, A., Karakesisoglou, I., Wong, E., Vaezi, A. & Fuchs, E. ACF7: an essential integrator of microtubule dynamics. *Cell* **115**, 343-54 (2003).
35. Lin, C.M., Chen, H.J., Leung, C.L., Parry, D.A. & Liem, R.K. Microtubule actin crosslinking factor 1b: a novel plakin that localizes to the Golgi complex. *J Cell Sci* **118**, 3727-38 (2005).
36. International Consortium for Blood Pressure Genome-Wide Association, S. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
37. Febbo, P.G. *et al.* Literature Lab: a method of automated literature interrogation to infer biology from microarray analysis. *BMC Genomics* **8**, 461 (2007).
38. Waterworth, D.M. *et al.* Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* **30**, 2264-76 (2010).
39. Tang, W. *et al.* Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood* **116**, 5032-6 (2010).
40. Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-54 (2013).
41. Kim, Y.J. *et al.* Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet* **43**, 990-5 (2011).
42. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**, 189-97 (2008).
43. Willer, C.J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* **40**, 161-9 (2008).

44. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet* **44**, 269-76 (2012).
45. Kristiansson, K. *et al.* Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. *Circ Cardiovasc Genet* **5**, 242-9 (2012).
46. Kooner, J.S. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* **40**, 149-51 (2008).
47. Chambers, J.C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* **43**, 1131-8 (2011).
48. Benhamed, F. *et al.* The lipogenic transcription factor ChREBP dissociates hepatic steatosis from insulin resistance in mice and humans. *J Clin Invest* **122**, 2176-94 (2012).
49. Herman, M.A. *et al.* A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism. *Nature* **484**, 333-8 (2012).
50. Iizuka, K., Bruick, R.K., Liang, G., Horton, J.D. & Uyeda, K. Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. *Proc Natl Acad Sci U S A* **101**, 7281-6 (2004).
51. Dentin, R. *et al.* Liver-specific inhibition of ChREBP improves hepatic steatosis and insulin resistance in ob/ob mice. *Diabetes* **55**, 2159-70 (2006).
52. Donnelly, K.L. *et al.* Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J Clin Invest* **115**, 1343-51 (2005).
53. Jeong, Y.S. *et al.* Integrated expression profiling and genome-wide analysis of ChREBP targets reveals the dual role for ChREBP in glucose-regulated gene expression. *PLoS One* **6**, e22544 (2011).
54. Albrechtsen, A. *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56**, 298-310 (2013).
55. Fassett, J.T. *et al.* Microtubule Actin Cross-Linking Factor 1 Regulates Cardiomyocyte Microtubule Distribution and Adaptation to Hemodynamic Overload. *PLoS One* **8**, e73887 (2013).
56. Rees, M.G. *et al.* Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia* **55**, 114-22 (2012).
57. Cariou, B. *et al.* Increased adipose tissue expression of Grb14 in several models of insulin resistance. *FASEB J* **18**, 965-7 (2004).
58. Carroll, E.A. *et al.* Cordon-bleu is a conserved gene involved in neural tube formation. *Dev Biol* **262**, 16-31 (2003).
59. Cooney, G.J. *et al.* Improved glucose homeostasis and enhanced insulin signalling in Grb14-deficient mice. *EMBO J* **23**, 582-93 (2004).
60. Depetris, R.S. *et al.* Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. *Mol Cell* **20**, 325-33 (2005).
61. Gieger, C. *et al.* New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**, 201-8 (2011).
62. Kraja, A.T. *et al.* A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. *Diabetes* **60**, 1329-39 (2011).
63. O'Seaghdha, C.M. *et al.* Meta-analysis of genome-wide association studies identifies six new Loci for serum calcium concentrations. *PLoS Genet* **9**, e1003796 (2013).
64. Rung, J. *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* **41**, 1110-5 (2009).
65. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* **357**, 443-53 (2007).

66. Santoro, N. *et al.* Variant in the glucokinase regulatory protein (GCKR) gene is associated with fatty liver in obese children and adolescents. *Hepatology* **55**, 781-9 (2012).
67. Stancakova, A. *et al.* Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* **61**, 1895-902 (2012).
68. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet* **44**, 260-8 (2012).
69. Aulchenko, Y.S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* **41**, 47-55 (2009).
70. Chambers, J.C. *et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* **40**, 716-8 (2008).
71. Diabetes Genetics Initiative of Broad Institute of, H. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-6 (2007).
72. Heid, I.M. *et al.* Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circ Cardiovasc Genet* **1**, 10-20 (2008).
73. Johansen, C.T. *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* **42**, 684-7 (2010).
74. Kathiresan, S. *et al.* A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet* **8 Suppl 1**, S17 (2007).
75. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* **41**, 56-65 (2009).
76. Kilpelainen, T.O. *et al.* Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet* **43**, 753-60 (2011).
77. Middelberg, R.P. *et al.* Genetic variants in LPL, OASL and TOMM40/APOE-C1-C2-C4 genes are associated with multiple cardiovascular-related traits. *BMC Med Genet* **12**, 123 (2011).
78. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* **41**, 35-46 (2009).
79. Tan, A. *et al.* A genome-wide association and gene-environment interaction study for serum triglycerides levels in a healthy Chinese male population. *Hum Mol Genet* **21**, 1658-64 (2012).
80. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* **42**, 579-89 (2010).
81. Zabaneh, D. & Balding, D.J. A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS One* **5**, e11961 (2010).
82. Akira, S., Misawa, T., Satoh, T. & Saitoh, T. Macrophages control innate inflammation. *Diabetes Obes Metab* **15 Suppl 3**, 10-8 (2013).
83. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* **40**, 955-62 (2008).
84. Demirkan, A. *et al.* Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet* **8**, e1002490 (2012).
85. Fox, C.S. *et al.* Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet* **8**, e1002695 (2012).
86. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41**, 1088-93 (2009).
87. Heard-Costa, N.L. *et al.* NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet* **5**, e1000539 (2009).
88. International Multiple Sclerosis Genetics, C. *et al.* Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* **357**, 851-62 (2007).

89. Lemaitre, R.N. *et al.* Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet* **7**, e1002193 (2011).
90. Lindgren, C.M. *et al.* Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* **5**, e1000508 (2009).
91. Liu, M.J. *et al.* ZIP8 regulates host defense through zinc-mediated inhibition of NF-kappaB. *Cell Rep* **3**, 386-400 (2013).
92. Liuzzi, J.P. *et al.* Interleukin-6 regulates the zinc transporter Zip14 in liver and contributes to the hypozincemia of the acute-phase response. *Proc Natl Acad Sci U S A* **102**, 6843-8 (2005).
93. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* **41**, 157-9 (2009).
94. O'Donnell, C.J. *et al.* Genome-wide association study for subclinical atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study. *BMC Med Genet* **8 Suppl 1**, S4 (2007).
95. Reiner, A.P. *et al.* Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am J Hum Genet* **82**, 1193-201 (2008).
96. Seshadri, S. *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832-40 (2010).
97. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* **41**, 1182-90 (2009).
98. Staubert, C. *et al.* Evolutionary aspects in evaluating mutations in the melanocortin 4 receptor. *Endocrinology* **148**, 4642-8 (2007).
99. Stocks, T. *et al.* TFAP2B influences the effect of dietary fat on weight loss under energy restriction. *PLoS One* **7**, e43212 (2012).
100. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54-60 (2011).
101. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* **41**, 18-24 (2009).
102. Yang, Q., Kathiresan, S., Lin, J.P., Tofler, G.H. & O'Donnell, C.J. Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. *BMC Med Genet* **8 Suppl 1**, S12 (2007).
103. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336-41 (2007).
104. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-8 (2010).
105. Ye, R. & Scherer, P.E. Adiponectin, driver or passenger on the road to insulin sensitivity? *Mol Metab* **2**, 133-141 (2013).
106. Digby, J.E. *et al.* Anti-inflammatory effects of nicotinic acid in human monocytes are mediated by GPR109A dependent mechanisms. *Arterioscler Thromb Vasc Biol* **32**, 669-76 (2012).
107. Ory, V. *et al.* c-mip down-regulates NF-kappaB activity and promotes apoptosis in podocytes. *Am J Pathol* **180**, 2284-92 (2012).
108. Kitchener, R.L. & Grunden, A.M. Prolidase function in proline metabolism and its medical and biotechnological applications. *J Appl Microbiol* **113**, 233-47 (2012).
109. Duong, H.S. *et al.* Elevated prolidase activity in keloids: correlation with type I collagen turnover. *Br J Dermatol* **154**, 820-8 (2006).
110. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-63 (2009).
111. Freathy, R.M. *et al.* Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes* **57**, 1419-26 (2008).

112. Fall, T. *et al.* The role of adiposity in cardiometabolic traits: a mendelian randomization analysis. *PLoS medicine* **10**, e1001474 (2013).
113. Province, M.A. *et al.* A meta-analysis of genome-wide linkage scans for hypertension: the National Heart, Lung and Blood Institute Family Blood Pressure Program. *American journal of hypertension* **16**, 144-7 (2003).
114. Borecki, I.B. & Province, M.A. Genetic and genomic discovery using family studies. *Circulation* **118**, 1057-63 (2008).
115. Grundy, S.M. *et al.* Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation* **109**, 433-8 (2004).
116. Wu, J. *et al.* A summary of the effects of antihypertensive medications on measured blood pressure. *Am J Hypertens* **18**, 935-42 (2005).
117. Wu, J. *et al.* An investigation of the effects of lipid-lowering medications: genome-wide linkage analysis of lipids in the HyperGEN study. *BMC Genet* **8**, 60 (2007).
118. Ripley, B.D. *Stochastic simulation*, (Wiley, New York ; Chichester, 1987).
119. Cureton, E.E. & D'Agostino, R.B. *Factor analysis, an applied approach*, xxi, 457 p. (L. Erlbaum Associates, Hillsdale, N.J., 1983).
120. Pearson, K. *On the theory of contingency and its relation to association and normal correlation*, (London, 1904).
121. Fisher, R.A. *Statistical Methods for Research Workers, Thirteenth edition*, (Oliver and Loyd, Ltd., London, England, 1925).
122. Province, M.A. The significance of not finding a gene. *Am J Hum Genet* **69**, 660-3 (2001).
123. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).