Dissecting the Obesity Disease Landscape: Identifying Gene-Gene Interactions that are Highly Associated with Body Mass Index

Rishika De
Department of Genetics
Geisel School of Medicine at Dartmouth
Hanover, NH, USA

Shefali S. Verma
Center for Systems Genomics
The Pennsylvania State University
University Park, PA, USA

IBC BMI Working Group

Michael V. Holmes
University of Pennsylvania
Philadelphia, PA, USA

Folkert W. Asselbergs
University Medical Center Utrecht
Utrecht, The Netherlands

Jason H. Moore
Department of Genetics
Geisel School of Medicine at Dartmouth
Hanover, NH, USA
Jason.H.Moore@dartmouth.edu

Brendan J. Keating
The Children's Hospital of Philadelphia
Philadelphia, PA, USA

Marylyn D. Ritchie
Center for Systems Genomics
The Pennsylvania State University
University Park, PA, USA

Diane Gilbert-Diamond
Department of Community and Family Medicine
Geisel School of Medicine at Dartmouth
Hanover, NH, USA
Diane.Gilbert-Diamond@dartmouth.edu

***Abstract*— Despite heritability estimates of 40-70% for obesity, less than 2% of its variation is explained by Body Mass Index (BMI) associated loci that have been identified so far. Epistasis, or gene-gene interactions are a plausible source to explain portions of the missing heritability of BMI. Using genotypic data from 18,686 individuals across five study cohorts – ARIC, CARDIA, FHS, CHS, MESA – we filtered SNPs (Single Nucleotide Polymorphisms) using two parallel approaches. SNPs were filtered either on the strength of their main effects of association with BMI, or on the number of knowledge sources supporting a specific SNP-SNP interaction in the context of obesity. Filtered SNPs were specifically analyzed for interactions that are highly associated with BMI using QMDR (Quantitative Multifactor Dimensionality Reduction). QMDR is a nonparametric, genetic model-free method that detects non-linear interactions in the context of a quantitative trait. We identified seven novel, epistatic models with a Bonferroni corrected p-value of association < 0.06. Prior experimental evidence helps explain the plausible biological interactions highlighted within our results and their relationship with obesity. We identified interactions between genes involved in mitochondrial dysfunction (POLG2), cholesterol metabolism (SOAT2), lipid metabolism (CYP11B2), cell adhesion (EZR), cell proliferation (MAP2K5), and insulin resistance (IGF1R). This study highlights a novel approach for discovering gene-gene interactions by combining methods such as QMDR with traditional statistics.**

Keywords—obesity; epistasis; gene-gene interaction, multifactor dimensionality reduction, GWAS

I. INTRODUCTION

Obesity is a major risk factor for various diseases such as - heart disease, type 2 diabetes and even certain types of cancer [1], [2]. Approximately, one-third of the adult population in the U.S. is categorized to be obese [3]. Globally, obesity has the potential to affect 1.12 billion individuals by 2030 [4]. In the U.S. alone, the economic burden associated with obesity has been estimated to be around $147 billion/year in healthcare costs and loss of productivity of affected individuals [5]. Moreover, obesity no longer affects only

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

industrialized nations, but it is also making its mark in developing nations, especially among children [3], [6].

Although the current epidemic proportions of obesity can be largely attributed to our lifestyle and food choices, there is also a strong genetic component of obesity. Twin and adoption studies have provided heritability estimates of 40-70% for obesity [7], [8]. Such studies have also found that obesity tends to cluster within families, and that monozygotic twins show greater concordance in Body Mass Index (BMI) and adiposity metrics versus dizygotic twins. Technological advancements in genomics and highly characterized genome-wide reference maps in major populations allow researchers to query a million or more genetic variants by designing genome-wide association studies (GWAS), [9]–[11] and so far, researchers have identified BMI-related signals in 32 loci that are associated with the trait at a genome-wide level [1]. However, these primary associations have been able to explain only about 2% of the variation observed in BMI [1].

The limited success of GWAS has often been attributed to the linear framework employed by these studies. Although, single locus analysis strategies have had success in certain diseases such as age-related macular degeneration and breast cancer [12]–[15], many complex diseases are likely the result of interactions between genetic loci – *epistasis* [9], [11], [16]. The ubiquitous nature of epistasis has been discussed previously, and it has highlighted the importance of designing our studies to embrace the genomic and environmental context of Single Nucleotide Polymorphisms (SNPs), by specifically searching for non-linear interactions between genetic loci [17], [18].

In this study we aimed to identify interactions between SNPs that are associated with BMI using data from 18,686 individuals across five highly characterized National Heart, Lung and Blood Institute (NHLBI) study cohorts.

II. MATERIALS AND METHODS

A. Participants

Figure 1 illustrates the overall study design. Genotype and phenotype information were initially combined for a total of 18,686 individuals of European descent from the following studies: Atherosclerosis Risk in Communities (ARIC) [19]; Coronary Artery Risk Development in Young Adults (CARDIA) [20]; Cardiovascular Health Study (CHS) [21]; Framingham Heart Study (FHS) [22]; and Multi-Ethnic Study of Atherosclerosis (MESA) [23] (Supplementary Table 1).

B. Genotyping and quality control

Genotyping was performed using the gene-centric ITMAT-Broad-CARe (IBC) array. This array was designed specifically to test over 2,000 loci implicated in various cardiovascular, metabolic and inflammatory phenotypes [24]. The array contains 47, 451 SNPs. Samples with a call rate less than 90% were excluded. Additionally, SNPs with a call rate less than 95%, with an exact test of Hardy-Weinberg equilibrium *p*-value greater than 1.00E-07 or a minor allele frequency (MAF) < 0.05 were also excluded. SNPs were further tested for linkage disequilibrium (LD) – a SNP was removed from each pair of SNPs that had an LD (r^2) $\geq$ 0.6. This reduced our dataset to 17,268 individuals and 28,453 SNPs. Non-founder individuals were also removed from the study population. To check for relatedness between individuals, markers were used for an Identity-by-descent (IBD) analysis using PLINK [25]. For pairs of individuals with a pi-hat ($\hat{\pi}$) value greater than 0.3, one individual was removed. Complete phenotype data was also required for inclusion of an individual in the analysis. This resulted in a final dataset of 15,737 individuals and 28,453 SNPs. To decrease both the computation time and the multiple testing burden two filtering strategies were employed [26]. These are described in more detail below.

C. Marker Selection

1) Main effect filter

As an additional filtering step, SNPs were tested for their independent association with the continuous BMI outcome using linear regression. Upon visual inspection of the distribution of *p*-values, a cut-off value of $p < 0.016$ was chosen, as there was a distinct separation between SNPs exhibiting a stronger main effect and the rest of the SNPs at this cut-off. This resulted in a final list of 498 markers for further analysis [27].

2) Biofilter

As a parallel filtering procedure, SNPs were also analyzed using Biofilter [28]. Biofilter is a knowledge-based approach that enables the analysis of multi-SNP interactions in a large dataset. The software identifies multi-SNP models that exhibit marginal effects on a phenotype, but are also biologically plausible. It combines information from multiple public knowledge sources such as Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Database of Interacting Proteins (DIP) and the Protein Families Database (Pfam) [29]. These sources provide information regarding pairs of genes that may be putative sources of epistasis and relate genes to one another through their mutual participation in biological processes, signaling pathways, protein-protein interactions as well as via the structural similarity between protein motifs. Biofilter measures the strength of the knowledge-based support for a given multi-SNP model with an *implication index*. The implication index is the sum of the number of supporting data sources for each of the genes in a given gene-gene relationship. For our analysis, models with an implication index of five or greater were retained, resulting in a list of 1815 markers (22,644 SNP-SNP models). The selected implication index cut-off was slightly more stringent than those used in previous studies [30].

D. Statistical analyses

1) Covariate adjustment

Baseline BMI values were regressed on Age, Age^2, Sex, the first three principal components of race computed using EIGENSTRAT software [31] and the index SNP rs11642841 in the FTO region. SNPs in the FTO locus are some of the strongest genetic associations identified for obesity risk [32]. Hence, adjustments were made for a SNP in the FTO locus to increase our ability to identify SNP-SNP models that were not

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

primarily driven by the strong main effect of this gene. The residual BMIs from this regression model were then used as the continuous outcome variable in the QMDR analysis.

2) Association analysis – QMDR

SNPs obtained from the two parallel filtering procedures described above, were tested for association with the continuous BMI outcome using Quantitative Multifactor Dimensionality Reduction (QMDR) [33]. QMDR is an extension of the two-class MDR algorithm that can detect and characterize epistatic SNP-SNP interactions in the context of a quantitative trait [34].

The original MDR algorithm was designed as a data reduction approach to identify multi-locus genotype combinations that are associated with high or low risk of disease [34]. Within a given dataset of *m* SNPs, *k* SNPs can be selected to examine a *k*-order interaction. MDR then constructs a contingency table for these *k* SNPs, and calculates case-control ratios for each of the possible multi-locus genotypes. Next, the case-control ratio for each multi-locus genotype is compared to the global case-control ratio for the whole dataset. Accordingly, a genotype is considered *high-risk* if its case-control ratio exceeds the global case-control ratio. Alternatively, it is considered to be *low-risk*.

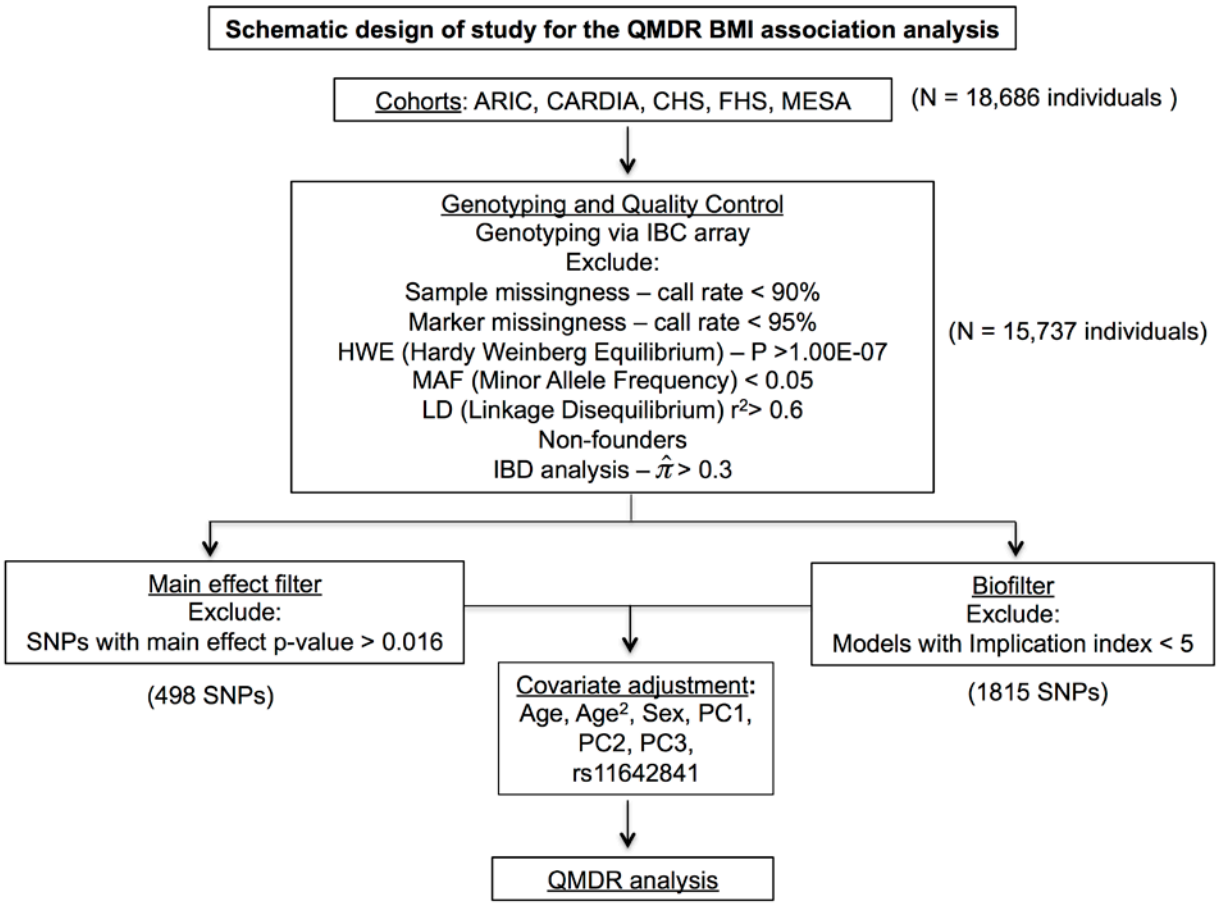

Fig. 1. Schematic design of the QMDR (Quantitative Multifactor Dimensionality Reduction) analysis for identifying SNP-SNP interaction models associated with BMI. Genotyping was performed using the IBC (ITMAT-Broad-CARe) array. The workflow also includes the initial quality control procedures, subsequent association analyses, and covariate adjustment steps performed.

However, when QMDR constructs a similar contingency table for *k* SNPs, it compares the mean value of the phenotype to the overall mean of the phenotypic trait within the dataset. Hence, a genotype combination is considered *high-level* if its mean value is larger than the overall mean of the phenotypic trait within the dataset. Otherwise, it is considered *low-level*. Next, QMDR combines the *'high-level'* and *'low-level'* genotypes into respective groups, and compares the phenotypic outcomes between these two groups using a T-test.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

QMDR also uses a 10-fold cross-validation procedure similar to the original MDR algorithm. The dataset is divided into 10 portions – 9 portions are used as a training dataset, and the remaining portion is used as a testing dataset. Next, the training t-statistic is calculated for each *k*-way interaction in the training dataset. The *k*-way model with the best training score is then used to predict the case-control status in the testing dataset. Ultimately, the best *k*-order interaction model is chosen based upon the training t-statistic and the highest testing t-statistic is used to select the best overall model for the dataset.

In the current analyses, we utilized QMDR to specifically test filtered SNPs for all possible two-way (SNP-SNP) interaction models that are associated with the continuous BMI outcome based on their training T-statistic scores. Amongst these models, we selected the 100 best overall SNP-SNP models using their testing T-statistic scores.

3) Permutation testing to assess statistical significance

Permutation procedures were performed to determine a cut-off threshold for an α=0.05 significance level. A 1000 permutations were performed, and in each permuted dataset the 100 best two-way SNP models were selected based on their T-statistic training and testing values. The null distribution of the 100 best SNP models and T-statistic values obtained from all permutations was utilized to calculate *p*-values for SNP-SNP models. *P*-values were also corrected for multiple testing using standard Bonferroni corrections.

4) Assessing the non-additive nature of identified pairwise interactions

A 1000 permuted datasets were created using the *explicit test of epistasis*, by shuffling genotypic data for each SNP [35]. However, genotype frequencies were maintained so that independent main effects were preserved while non-linear interactions were randomized. Linear regression was used to model the identified statistically significant SNP-SNP interactions in relation to BMI within the original and permuted datasets. Interactions between SNPs were coded as Cartesian products within the regression model. The nine possible two-locus genotypes were coded from 0-8 (Table S2). The null distribution was created using the F-statistic values for the regression models from the 1000 permuted datasets. This was used to calculate the 'explicit epistasis' *p*-value associated with the original pairwise interactions that were identified.

5) Biological evidence for identified pairwise interactions

To identify known biological evidence supporting the statistically significant pairwise interactions, we mapped each SNP to a corresponding gene using information from dbSNP (build 139) and SCANdb (http://www.scandb.org). We also searched for evidence of functional relationships between interacting genes using the Integrated Multi-Species Prediction (IMP) web server [36]. IMP integrates information from a large number of sources including experimentally verified information from gene expression studies, IntAct, MINT, MIPS, and BioGRID databases. It mines empirical data to provide a predictive probability that two genes work together within a given biological process.

III. RESULTS

A. Main effect filter

Using the set of SNPs that emerged from the main effect filter, QMDR analysis identified seven novel SNP-SNP interaction models that were associated with BMI (Bonferroni corrected *p*-value <0.06) (Table 1). These SNP-SNP models also reflect strong epistatic relationships. *P*-values associated with the non-additive nature of these interactions are also presented in Table 1. We also queried the biological and functional context of these interactions using IMP. However, since both *FLJ30838* and *C7orf10* are of unknown function, we gained most insight regarding interactions 3, 5 and 6 (Table 1). *ASTL* and *CYP11B2* were found to interact via two genes – *MEP1B* and *CYP2C9* (Figure 2a). A functional partner of *EZR* was found to interact with MAP2K5 through other participants in the *MAPK* signaling pathway (Fig 2b). Lastly, a member of the *IGF1R* protein complex was found to interact with *CAV3* (Fig 2c).

TABLE I. Results for QMDR association analysis for continuous BMI outcome. Seven signals reached a Bonferroni corrected *p*-value < 0.1. SNPs have been mapped to their corresponding genes using dbSNP (build 139) and SCANdb. SNP1 and SNP2 indicate the individual SNPs within a given SNP-SNP interaction model identified by QMDR. Chromosomal location of SNPs is noted in the following format - Chromosome:Base pair. *P*-values were calculated from a distribution built from 1000 permutations. *P*-values were also corrected using the Bonferroni method. Explicit epistasis *p*-values were calculated from a distribution built from 1000 permutations using the 'explicit test of epistasis'.

Rank	Model	SNP1	Chr:bp	Gene1	SNP2	Chr:bp	Gene2	Permuted P-Value	Bonferroni Corrected P-value	Explicit Epistasis P-value
1	rs17171686,rs1427463	rs17171686	7:40335451	C7orf10	rs1427463	17:59923044	POLG2	< 0.00011	0.01	0.000
2	rs12617233,rs1427463	rs12617233	2:58893502	FLJ30838	rs1427463	17:59923044	POLG2	< 0.00012	0.01	0.012
3	rs749457,rs1799998	rs749457	2:96159671	ASTL	rs1799998	8:143996602	CYP11B2	< 0.00026	0.03	0.000
4	rs12617233,rs12210959	rs12617233	2:58893502	FLJ30838	rs12210959	6:6121143	F13A1	< 0.00038	0.04	0.003
5	rs3102976,rs997295	rs3102976	6:159110007	EZR	rs997295	15:65803397	MAP2K5	< 0.00046	0.05	0.000
6	rs2268484,rs8038415	rs2268484	3:8748950	CAV3	rs8038415	15:97316957	IGF1R	< 0.00046	0.05	0.009
7	rs12617233,rs822682	rs12617233	2:58893502	FLJ30838	rs822682	12:51798711	SOAT2	< 0.00061	0.06	0.018

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

B. Biofilter

Using the set of SNPs that emerged from the Biofilter procedure, QMDR analysis did not identify any significant SNP-SNP interaction models that were associated with BMI.

IV. DISCUSSION

In this study, we analyzed the genetic and phenotypic information for a total of 15,737 individuals combined from five study cohorts – ARIC, CARDIA, CHS, FHS and MESA. SNPs were either filtered based on the strength of their independent effects or on the number of independent sources of biological knowledge supporting them. Filtered SNPs were then specifically tested for SNP-SNP interactions.

Historically, GWA studies have employed a linear modeling framework that tests single SNPs one at a time, for its association with a given phenotype. Unfortunately, such an approach does not consider the genomic context of a SNP. Moreover, Hirschhorn *et al* have shown that positive results from studies employing such an approach typically cannot be replicated across independent studies [37]. This has highlighted the need for embracing the complexity of a genotype-phenotype relationship by focusing on gene-gene interactions [38]. However, detecting gene-gene interactions in a GWAS presents a considerable computational and statistical challenge. Moore and Ritchie describe the need for designing new computational methods for detecting high-order non-linear interactions since traditional approaches such as logistic regression have limited power when modeling such interactions in high-dimensional data [39], [40]. They also stress upon the importance of filtering methods for the selection of SNPs to be included in an analysis. The exhaustive search of all possible combinations of thousands of SNPs is computationally very expensive. Our approach addresses both of these challenges. QMDR – an extension of the original MDR developed for quantitative traits – is a non-parametric method that does not assume any genetic model. Most importantly, MDR greatly reduces the degrees of freedom required for modeling interactions. We also address the SNP-selection problem by applying two parallel filtering approaches, thereby effectively reducing our search space for detecting meaningful interactions.

We identified seven novel interactions that are highly associated with BMI. These seven interactions were also explicitly tested for the presence of epistasis. All the identified interactions exhibited an epistatic component. Four of these interactions stand out since SNPs within these interactions have previously been identified as independent signals associated with BMI [27]. These SNPs are – rs12617233 in *FLJ30838* and rs997295 in *MAP2K5* – within interactions 2, 4, 5 and 7 (Table 1). *FLJ30838* is a long intergenic non-coding RNA (lincRNA) of unknown function. It was found to interact with rs1427463 in *POLG2*, rs12210959 in *F13A1*, and rs822682 in *SOAT2*. Incidentally, none of these other SNPs have been implicated in obesity before. The rs1427463 variant has been associated with height previously in an African ancestry population, which obviously factors into BMI calculations [41].

In the first two interactions, rs1427463 in *POLG2* interacts with rs17171686 in *C7orf10* and rs12617233 in *FLJ30838* respectively. Although, not much is known regarding the functions of *C7orf10* and *FLJ3083*, we can gain some understanding regarding these interactions from *POLG2*. *POLG2* encodes for a subunit of the mitochondrial DNA polymerase gamma. It is largely involved in metabolic pathways and the transcriptional activation of mitochondrial biogenesis [42], [43]. Researchers have shown that an increase in mitochondrial biogenesis was able to prevent the development of obesity in mice [44]. Conversely, genetically engineered mice with reduced expression of genes involved in mitochondrial respiration eventually developed obesity [45]. Consequently, the involvement of mitochondrial dynamics in obesity has gained a lot of support [46].

The SNP rs12617233 in *FLJ30838* also interacts with the SNPs rs12210959 in *F13A1*, and rs822682 in *SOAT2* respectively. As mentioned before, not much is known about FLJ30838 or about the functional role of the SNPs in *F13A1* and *SOAT2*. *F13A1* encodes for the A subunit of the coagulation factor XIII [47]. Several SNPs on this gene were found to be highly associated with BMI in a study utilizing gene expression data from monozygotic twins to deeply interrogate GWAS data [48]. Interestingly, the SNP identified in our study is independent of the *F13A1* signals identified by Naukkarinen *et al*. Moreover, this gene is also a part of the pathway involved in the formation of the fibrin clot [42], [43]. Several studies in obese individuals and rodent models of obesity have also reported increased levels of coagulation factors [49], [50] . However, the exact mechanism by which this gene and other coagulation factors may impact obesity is unknown.

SOAT2, also known as sterol O-acyltransferase 2, has been found to be a major regulator of cholesterol metabolism and absorption in the small intestine and liver of mice on a high-cholesterol and high-fat diet [51]. Impaired cholesterol absorption has been linked to high BMI and obesity [52], [53]. Although the exact mechanism is unclear, researchers have suggested that increased cholesterol synthesis and secretion combined with dietary cholesterol being consumed, may ultimately affect cholesterol absorption efficiency in obese individuals [54].

We also found a significant association between rs749457 in *ASTL* and rs1799998 in *CYP11B2* related to BMI. The variant rs1799998 has been associated with insulin resistance, diabetes, and metabolic syndrome in humans, but it has not been shown to have an independent association with BMI [55]–[57]. Little is known regarding the function of *ASTL* in humans, a specific protease that uses metals in catalytic processes [58]. However, there is moderate support connecting *ASTL* to a functional partner of *CYP11B2* (Figure 2a). *ASTL* shares a strong sequence similarity and a common genetic ancestor with *MEP1A*. Both *MEP1A* and *MEP1B* are distinct yet evolutionarily related subunits of meprins – proteases that are involved with metals [59]. *MEP1B* shares a transcription factor binding site with and is part of the same gene expression signature as *CYP2C9*. Both *CYP11B2* and *CYP2C9* are functionally related by their roles in lipid metabolism [36]. *CYP11B2* is specifically involved in

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

mineralocorticoid biosynthesis [42], [43]. Incidentally, the mineralocorticoid receptor has been shown to play an important role in the positive control of adipogenesis and thus, in the development of obesity [60].

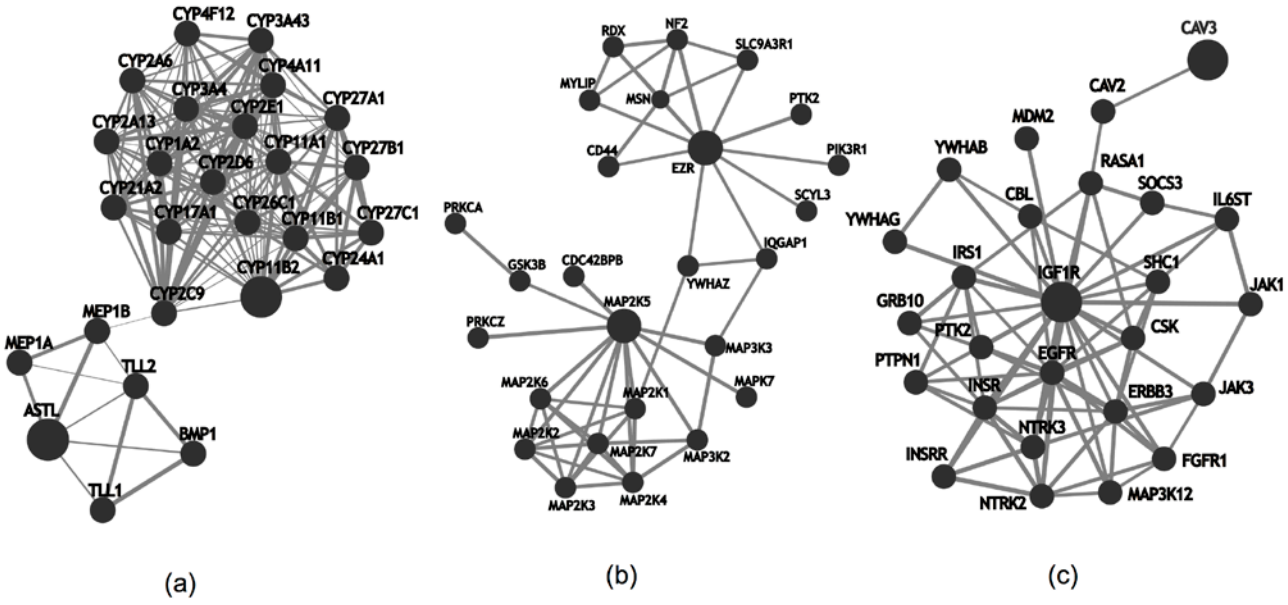

Fig. 2. Functional relationship networks generated from Integrated Multi-Species Prediction (IMP) from identified SNP- SNP interactions that are highly associated with BMI. Identified SNPs were mapped to their respective genes. Gene pairs were used to query IMP to make functional connections between them. IMP is a web-based tool that mines empirical data to provide a predictive probability that two genes have a functional relationship. Nodes in the network represent genes. Query genes are represented with larger nodes. Edges between nodes represent a functional relationship between two genes. Shown are interactions between (a) rs749457 in *ASTL* and rs1799998 in *CYP11B2* (b) rs3102976 in *EZR* and rs997295 in *MAP2K5* (c) rs2268484 in *CAV3* and rs8038415 in *IGF1R*.

We also observed an interaction between rs997295 in *MAP2K5* and rs3102976 in *EZR* related to BMI. MAP2K5 is a part of the MAPK signaling pathway involved in growth factor stimulated cell proliferation. *EZR*, or ezrin, encodes a cytoplasmic peripheral membrane protein that acts as a linkage between the plasma membrane and the actin cytoskeleton. Consequently, ezrin plays an important role in cell adhesion, migration and organization. There is prior experimental evidence supporting the physical and functional connection between these two genes (Figure 2b). MAP2K5 is known to physically interact and participate with MAP3K3 in the MAPK signaling pathway. MAP3K3 also interacts with the GTPase activating protein IQGAP1. IQGAP1 interacts with EZR and is also involved in the regulation of the actin cytoskeleton [36]. One can imagine the strong need for regulating the actin cytoskeleton during dynamic processes such as adipogenesis.

Lastly, we observed an interaction between rs2268484 in *CAV3* and rs8038415 in *IGF1R* related to BMI. *CAV3* encodes for a muscle-specific form of the caveolin family of proteins. Researchers have found that *CAV3*-knockout mice developed insulin resistance in their skeletal muscles [61] and that adenovirus-mediated gene transfer of *CAV3* was able to increase glycogen synthesis in the liver as well as improve insulin signaling in diabetic obese mice [62]. *IGF1R* codes for a receptor of insulin-like growth factor, a polypeptide hormone that plays an important role in growth. IGF1 also regulates pancreatic β-cell mass and thus plays a crucial role in insulin signaling. Hence, impaired *IGF1* signaling may alter insulin secretion by β-cells and negatively impact the hypothalamus – a region of the brain associated with food intake – ultimately causing weight gain [63]. Interestingly, a member of the *IGF1R* protein complex assembly (*RAS1*) is known to interact with a functional partner of *CAV3* in a number of processes such as signal transduction, endocytosis and focal adhesion (Figure 2c) [36].

The use of the IBC array in this study highlights the strengths of this custom array in detecting potentially disease-causing loci that are also supported by a substantial amount of biological evidence. However, while the array has dense coverage in gene-centric regions, it only includes 2000 loci. This limitation of the array was highlighted by the inability of BioFilter to identify any statistically significant SNP-SNP models. The use of methods such as BioFilter may be more suited for larger GWAS datasets including more loci.

This study identifies gene-gene interactions that are potentially associated with obesity. Prior experimental evidence suggests the plausible biological relevance of several of the identified loci. However, we also identified a few loci of unknown function. Unfortunately, the inferences that can be drawn from our results are limited by a functional annotation bias – well-studied genes are assigned many annotations while understudied genes often lack annotations. One could

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

speculate that the genes involved in these interactions are multi-functional, thereby connecting various biological processes and pathways. Ultimately, further biological validations will be necessary to determine whether the identified interactions play a role in the complex genetic architecture of obesity.

V. CONCLUSION

Main effects analyses have explained little of the genetic heritability of obesity. The use of methods such as QMDR in conjunction with traditional statistical analyses can unravel this complex network by identifying gene-gene interactions that play key roles in the etiology of obesity. Our QMDR analysis of genotypic data from 5 study cohorts identified novel interactions between genetic variants that are highly associated with BMI. Future studies are necessary to verify the observed associations.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health grants – NLM R01 grants (LM010098, LM011360, LM009012), and GMS P20 grants (GM103506, GM103534 and GM104416). CARe acknowledges the support of the National Heart, Lung and Blood Institute and the contributions of the research institutions, study investigators, field staff, and study participants in creating this resource for biomedical research (NHLBI contract number HHSN268200960009C).

REFERENCES

[1] E. K. Speliotes et al., “Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.,” *Nat. Genet.*, vol. 42, no. 11, pp. 937–948, Nov. 2010.

[2] E. E. Calle and R. Kaaks, “Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms.,” *Nat. Rev. Cancer*, vol. 4, no. 8, pp. 579–591, Aug. 2004.

[3] C. L. Ogden, M. D. Carroll, B. K. Kit, and K. M. Flegal, “Prevalence of childhood and adult obesity in the United States, 2011-2012.,” *JAMA*, vol. 311, no. 8, pp. 806–814, Feb. 2014.

[4] T. Kelly, W. Yang, C. S. Chen, K. Reynolds, and J. He, “Global burden of obesity in 2005 and projections to 2030.,” *Int. J. Obes.*, vol. 32, no. 9, pp. 1431–1437, Sep. 2008.

[5] A. Scuteri et al., “Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits.,” *PLoS Genet.*, vol. 3, no. 7, p. e115, Jul. 2007.

[6] J. Zhao et al., “Role of BMI-associated loci identified in GWAS meta-analyses in the context of common childhood obesity in European Americans.,” *Obesity*, vol. 19, no. 12, pp. 2436–2439, Dec. 2011.

[7] A. J. Stunkard, T. T. Foch, and Z. Hrubec, “A Twin Study of Human Obesity,” *J. Am. Med. Assoc.*, vol. 256, no. 1, pp. 51–54, 1986.

[8] H. H. Maes, M. C. Neale, and L. J. Eaves, “Genetic and environmental factors in relative body weight and human adiposity.,” *Behav. Genet.*, vol. 27, no. 4, pp. 325–351, Jul. 1997.

[9] E. E. Eichler et al., “Missing heritability and strategies for finding the underlying causes of complex disease.,” *Nat. Rev. Genet.*, vol. 11, no. 6, pp. 446–450, Jun. 2010.

[10] J. H. Moore, F. W. Asselbergs, and S. M. Williams, “Bioinformatics challenges for genome-wide association studies.,” *Bioinformatics*, vol. 26, no. 4, pp. 445–455, Feb. 2010.

[11] T. A. Manolio et al., “Finding the missing heritability of complex diseases.,” *Nature*, vol. 461, no. 7265, pp. 747–753, Oct. 2009.

[12] A. O. Edwards, R. Ritter, K. J. Abel, A. Manning, C. Panhuysen, and L. A. Farrer, “Complement factor H polymorphism and age-related macular degeneration.,” *Science*, vol. 308, no. 5720, pp. 421–424, Apr. 2005.

[13] R. J. Klein et al., “Complement factor H polymorphism in age-related macular degeneration.,” *Science*, vol. 308, no. 5720, pp. 385–389, Apr. 2005.

[14] J. L. Haines et al., “Complement factor H variant increases the risk of age-related macular degeneration.,” *Science*, vol. 308, no. 5720, pp. 419–421, Apr. 2005.

[15] D. F. Easton et al., “Genome-wide association study identifies novel breast cancer susceptibility loci,” *Nature*, vol. 447, no. 7148, pp. 1087–1093, 2009.

[16] N. Zaitlen and P. Kraft, “Heritability in the genome-wide association era.,” *Hum. Genet.*, vol. 131, no. 10, pp. 1655-1664, Jul. 2012.

[17] J. H. Moore, “The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases,” *Hum. Hered.*, vol. 56, no. 1–3, pp. 73–82, 2003.

[18] T. F. C. Mackay, “Epistasis and quantitative traits: using model organisms to study gene-gene interactions.,” *Nat. Rev. Genet.*, vol. 15, no. 1, pp. 22–33, Jan. 2014.

[19] C. Hill et al., “The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives.,” *Am. J. Epidemiol.*, vol. 129, no. 4, pp. 687–702, Apr. 1989.

[20] G. D. Friedman et al., “CARDIA: study design, recruitment, and some characteristics of the examined subjects.,” *J. Clin. Epidemiol.*, vol. 41, no. 11, pp. 1105–16, Jan. 1988.

[21] L. P. Fried et al., “The Cardiovascular Health Study: design and rationale.,” *Ann. Epidemiol.*, vol. 1, no. 3, pp. 263–76, Feb. 1991.

[22] T. R. Dawber, G. F. Meadors, and F. E. Moore, “Epidemiological approaches to heart disease: the Framingham Study.,” *Am. J. Public Health Nations Health*, vol. 41, no. 3, pp. 279–281, Mar. 1951.

[23] D. E. Bild et al., “Multi-ethnic study of atherosclerosis: objectives and design.,” *Am. J. Epidemiol.*, vol. 156, no. 9, pp. 871–881, Nov. 2002.

[24] B. J. Keating et al., “Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies.,” *PLoS One*, vol. 3, no. 10, p. e3583, Jan. 2008.

[25] S. Purcell et al., “PLINK: a tool set for whole-genome association and population-based linkage analyses.,” *Am. J. Hum. Genet.*, vol. 81, no. 3, pp. 559–575, Sep. 2007.

[26] X. Sun, Q. Lu, S. Mukheerjee, P. K. Crane, R. Elston, and M. D. Ritchie, “Analysis pipeline for the epistasis search - statistical versus biological filtering.,” *Front. Genet.*, vol. 5, no. April, p. 106, Jan. 2014.

[27] Y. Guo, M. B. Lanktree, K. C. Taylor, H. Hakonarson, L. A. Lange, and B. J. Keating, “Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals.,” *Hum. Mol. Genet.*, vol. 22, no. 1, pp. 184–201, Jan. 2013.

[28] W. S. Bush, S. M. Dudek, and M. D. Ritchie, “Biofilter: A Knowledge-Integration System for the Multi-Locus Analysis of Genome-Wide Association Studies,” *Pacific Symp. Biocomput.*, pp. 368–379, 2009.

[29] S. a Pendergrass et al., “Genomic analyses with biofilter 2.0: knowledge driven filtering, annotation, and model development.,” *BioData Min.*, vol. 6, no. 1, p. 25, Jan. 2013.

[30] S. D. Turner et al., “Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent EMR-linked biobanks.,” *PLoS One*, vol. 6, no. 5, p. e19586, Jan. 2011.

[31] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, “Principal components analysis corrects for stratification in genome-wide association studies.,” *Nat. Genet.*, vol. 38, no. 8, pp. 904–909, Aug. 2006.

[32] T. M. Frayling et al., “A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.,” *Science*, vol. 316, no. 5826, pp. 889–894, May 2007.

[33] J. Gui et al., “A Simple and Computationally Efficient Approach to Multifactor Dimensionality Reduction Analysis of Gene-Gene Interactions for Quantitative Traits.,” *PLoS One*, vol. 8, no. 6, p. e66545, Jan. 2013.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

[34] M. D. Ritchie et al., “Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer.,” *Am. J. Hum. Genet.*, vol. 69, no. 1, pp. 138–147, Jul. 2001.
[35] C. S. Greene et al., “Enabling personal genomics with an explicit test of epistasis.,” *Pac. Symp. Biocomput.*, pp. 327–336, Jan. 2010.
[36] A. K. Wong, C. Y. Park, C. S. Greene, L. a Bongo, Y. Guan, and O. G. Troyanskaya, “IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks.,” *Nucleic Acids Res.*, vol. 40, no. Web Server issue, pp. W484–90, Jul. 2012.
[37] J. N. Hirschhorn, K. Lohmueller, E. Byrne, and K. Hirschhorn, “A comprehensive review of genetic association studies.,” *Genet. Med.*, vol. 4, no. 2, pp. 45–61, 2002.
[38] T. F. Mackay and J. H. Moore, “Why epistasis is important for tackling complex human disease genetics.,” *Genome Med.*, vol. 6, no. 6, p. 42, Jan. 2014.
[39] J. H. Moore and S. M. Williams, “New strategies for identifying gene-gene interactions in hypertension.,” *Ann. Med.*, vol. 34, no. 2, pp. 88–95, Jan. 2002.
[40] J. Moore and M. Ritchie, “The Challenges of Whole-Genome Approaches to Common Disease,” *J. Am. Med. Assoc.*, vol. 291, no. 13, pp. 1642–1643, 2004.
[41] C. L. Carty et al., “Genome-wide association study of body height in African Americans: the Women’s Health Initiative SNP Health Association Resource (SHARe).,” *Hum. Mol. Genet.*, vol. 21, no. 3, pp. 711–720, Feb. 2012.
[42] D. Croft et al., “The Reactome pathway knowledgebase.,” *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D472–7, Jan. 2014.
[43] M. Kanehisa and S. Goto, “KEGG: kyoto encyclopedia of genes and genomes.,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000.
[44] Y. X. Wang et al., “Regulation of muscle fiber type and running endurance by PPARdelta.,” *PLoS Biol.*, vol. 2, no. 10, p. e294, Oct. 2004.
[45] M. Schuler et al., “PGC1alpha expression is controlled in skeletal muscles by PPARbeta, whose ablation results in fiber-type switching, obesity, and type 2 diabetes.,” *Cell Metab.*, vol. 4, no. 5, pp. 407–414, Nov. 2006.
[46] A. Zorzano, M. Liesa, and M. Palacín, “Role of mitochondrial dynamics proteins in the pathophysiology of obesity and type 2 diabetes.,” *Int. J. Biochem. Cell Biol.*, vol. 41, no. 10, pp. 1846–1854, Oct. 2009.
[47] A. Ichinose and E. W. Davie, “Characterization of the gene for the a subunit of human factor XIII (plasma transglutaminase), a blood coagulation factor.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, no. 16, pp. 5829–33, Aug. 1988.
[48] J. Naukkarinen et al., “Use of genome-wide expression data to mine the ‘Gray Zone’ of GWA studies leads to novel candidate obesity genes.,” *PLoS Genet.*, vol. 6, no. 6, p. e1000976, Jun. 2010.
[49] T. Skurk and H. Hauner, “Obesity and impaired fibrinolysis: role of adipose production of plasminogen activator inhibitor-1.,” *Int. J. Obes. Relat. Metab. Disord.*, vol. 28, no. 11, pp. 1357–1364, Nov. 2004.
[50] D. C. W. Lau, B. Dhillon, H. Yan, P. E. Szmitko, and S. Verma, “Adipokines: molecular links between obesity and atheroslcerosis.,” *Am. J. Physiol. Heart Circ. Physiol.*, vol. 288, no. 5, pp. H2031–2041, May 2005.
[51] K. K. Buhman et al., “Resistance to diet-induced hypercholesterolemia and gallstone formation in ACAT2-deficient mice.,” *Nat. Med.*, vol. 6, no. 12, pp. 1341–1347, Dec. 2000.
[52] T. a Miettinen and H. Gylling, “Cholesterol absorption efficiency and sterol metabolism in obesity.,” *Atherosclerosis*, vol. 153, no. 1, pp. 241–248, Nov. 2000.
[53] P. Simonen, H. Gylling, A. N. Howard, and T. A. Miettinen, “Introducing a new component of the metabolic syndrome :,” pp. 82–88, 2000.
[54] H. Gylling and T. a Miettinen, “Cholesterol absorption: influence of body weight and the role of plant sterols.,” *Curr. Atheroscler. Rep.*, vol. 7, no. 6, pp. 466–471, Nov. 2005.
[55] P. Russo et al., “-344C/T Variant in the promoter of the aldosterone synthase gene (CYP11B2) is associated with metabolic syndrome in men.,” *Am. J. Hypertens.*, vol. 20, no. 2, pp. 218–222, Feb. 2007.
[56] K. Ranade et al., “Genetic variation in aldosterone synthase predicts plasma glucose levels.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 23, pp. 13219–13224, Nov. 2001.
[57] N. M. Bellili et al., “Associations of the -344 T>C and the 3097 G>A polymorphisms of CYP11B2 gene with hypertension, type 2 diabetes, and metabolic syndrome in a French population.,” *Am. J. Hypertens.*, vol. 23, no. 6, pp. 660–667, Jun. 2010.
[58] V. Quesada, L. M. Sánchez, J. Álvarez, and C. López-Otín, “Identification and Characterization of Human and Mouse Ovastacin: a novel metalloproteinase similar to hatching enzymes from arthropods, birds, amphibians, and fish ,” *J. Biol. Chem.* , vol. 279 , no. 25 , pp. 26627–26634, Jun. 2004.
[59] J. S. Bond, K. Rojas, J. Overhauser, H. Y. Zoghbi, and W. Jiang, “The structural genes, MEP1A and MEP1B, for the alpha and beta subunits of the metalloendopeptidase meprin map to human chromosomes 6p and 18q, respectively.,” *Genomics*, vol. 25, no. 1, pp. 300–3, Jan. 1995.
[60] M. Caprio, B. Fève, A. Claës, S. Viengchareun, M. Lombès, and M.-C. Zennaro, “Pivotal role of the mineralocorticoid receptor in corticosteroid-induced adipogenesis.,” *FASEB J.*, vol. 21, no. 9, pp. 2185–2194, Jul. 2007.
[61] J. Oshikawa et al., “Insulin resistance in skeletal muscles of caveolin-3-null mice.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 34, pp. 12670–12675, Aug. 2004.
[62] K. Otsu et al., “Caveolin gene transfer improves glucose metabolism in diabetic mice,” pp. 450–456, 2010.
[63] S. E. Kahn, R. L. Hull, and K. M. Utzschneider, “Mechanisms linking obesity to insulin resistance and type 2 diabetes.,” *Nature*, vol. 444, no. 7121, pp. 840–846, Dec. 2006.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE