BMC Proceedings

PROCEEDINGS

Open Access

Analysis of genetic and nongenetic factors influencing triglycerides-lowering drug effects based on paired observations

Zheng Xu[1,2*], Qing Duan[3], Juan Cui[4], Yumou Qiu[1], Qidong Jia[5], Cong Wu[6] and Jennifer Clarke[1,2]

Abstract

Obesity is a risk factor for heart disease, stroke, diabetes, high blood pressure, and other chronic diseases. Some drugs, including fenofibrate, are used to treat obesity or excessive weight by lowering the level of specific triglycerides. However, different groups have different drug sensitivities and, consequently, there are differences in drug effects. In this study, we assessed both genetic and nongenetic factors that influence drug responses and stratified patients into groups based on differential drug effect and sensitivity. Our methodology of investigating genetic factors and nongenetic factors is applicable to studying differential effects of other drugs, such as statins, and provides an approach to the development of personalized medicine.

Background

Obesity and excessive weight (body mass index > 25) are highly prevalent among US adults and youth [1]. Obesity puts a person at a higher risk for heart disease, stroke, diabetes, high blood pressure, and other medical ailments. Consequently, effective treatment strategies for obesity and excessive weight designed to improve an individual's health and quality of life are highly desired.

Genetic and nongenetic factors jointly influence the likelihood of obesity and being overweight [1, 2]. Obesity is associated with changes in blood lipid levels, which can increase the risk of cardiovascular diseases. Fenofibrate is recommended because of its triglyceride-lowering effect. However, many genetic and nongenetic factors may influence the effects of such medications. Instead of the same types and amounts of drugs for all patients, personalized medicine considers the differences in drug effects between individuals and recommends the optimal treatment strategy for each patient individually.

Our goal was to provide a better understanding of drug mechanisms and to contribute to the development of precision medicine by studying the genetic and nongenetic factors that influence the effects of fenofibrate in treatment of obesity. We identified groups of individuals with differential drug effects. In addition, our methodology has the potential to be applied in the study of differences in drug effects and personalized medicine based on other medicines such as statins.

Methods

GAW20 provided the data. We are interested in understanding factors that may influence the drug effects of fenofibrate on triglyceride levels.

Our proposed methodology can work for the *paired observation* situation in which the research problem is how multiple factors influence the differences in drug effects. We measured the phenotypes of the *same individuals before and after* treatment with fenofibrate, and genotypes and nongenetic data on these individuals are available.

For a data set with paired observations, denote the response or phenotype of interest (drug effect in this context) before and after treatment with a drug as Y_{pre} and Y_{post}. The change $\Delta Y = Y_{post} - Y_{pre}$ is the "drug effect." In this

* Correspondence: zheng.xu@unl.edu
[1]Department of Statistics, University of Nebraska, 340 Hardin Hall North Wing, Lincoln, NE 68588, USA
[2]Department of Food Science and Technology, University of Nebraska, 1901 N. 21st Street, Lincoln, NE 68588, USA
Full list of author information is available at the end of the article

Xu *et al. BMC Proceedings* 2018, **12**(Suppl 9):46

Page 250 of 258
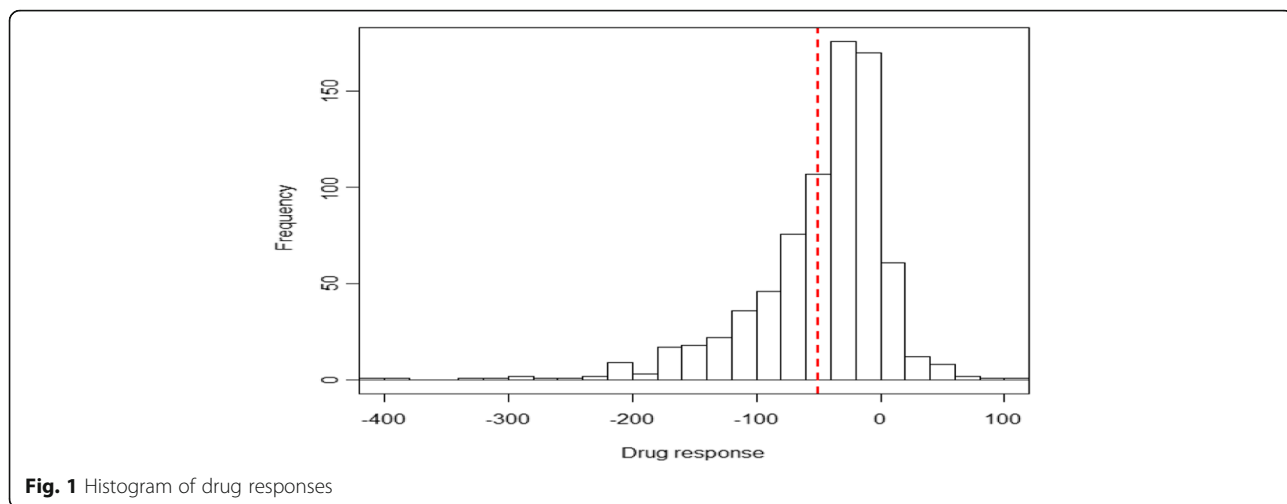
**Fig. 1** Histogram of drug responses

study, our interest is the raw difference in the level of tri-glycerides at visit 4 (after the treatment) minus at visit 2 (before the treatment).

We first inspected the distribution of the drug effect (ie, response) ΔY to characterize its average and variability. Then, among s nongenetic factors $C_1, C_2, ..., C_s$ and the top 10 principal components (PCs) of the genotypes $PC_1, ..., PC_{10}$ (ie ancestry variables), we checked the association of these variables with drug effect in a multiple regression framework. When there are different genetic marker frequencies and different drug effects, there are drug–ancestry interactions that influence the drug effects.

It was noted that factors influencing phenotype Y may not be the factors influencing drug response ΔY. For example, assume that before and after the treatment, factor Z influences Y. We can model $Y_{pre} = \beta_{0,\ pre} + \beta_{Z,\ pre}Z + ...$ and $Y_{post} = \beta_{0,\ post} + \beta_{Z,\ post}Z + ...$ Then we have $\Delta Y = Y_{post} - Y_{pre} = (\beta_{0,\ post} - \beta_{0,\ pre}) + (\beta_{Z,\ post} - \beta_{Z,\ pre})Z + ...$

For Z to be a factor influencing ΔY, the factor Z has to have different magnitude in its effect on Y_{post} and Y_{pre}, which is the drug-by-factor-Z interaction effect in the expression of Y. This interaction effect captures, for example, the situation in which Z has no effect on Y_{pre} but does affect Y_{post}.

We assessed genetic variants genome-wide to find single-nucleotide polymorphisms (SNPs) associated with ΔY. We divided the sets of SNPs into common SNPs and rare SNPs. For common SNPs (minor allele frequency [MAF] ≥5%), we conducted genome-wide association studies based on familial data, controlling for covariates. For rare SNPs (1% ≤ MAF < 5%), we conducted both gene-based and region-based rare variant tests based on familial data, using the fast family-based sequencing kernel association testing method (FFBSKAT), which is a specific method to extend the sequence kernel association test (SKAT) for unrelated individuals to familial

Fig. 2 Histogram of MAFs

Xu *et al. BMC Proceedings* 2018, **12**(Suppl 9):46

Page 251 of 258
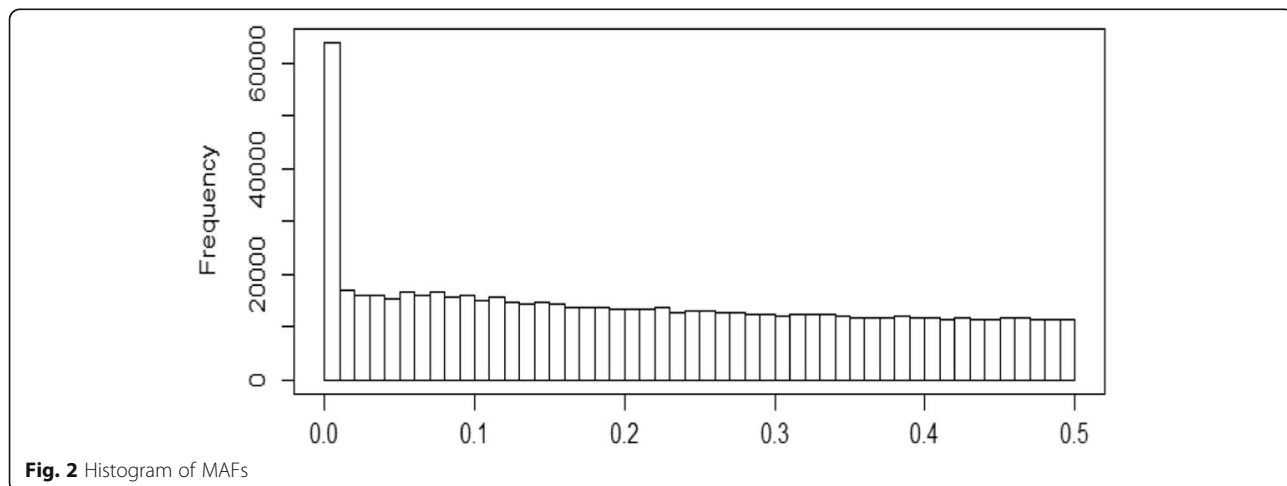
Fig. 3 Manhattan plot for common SNPs
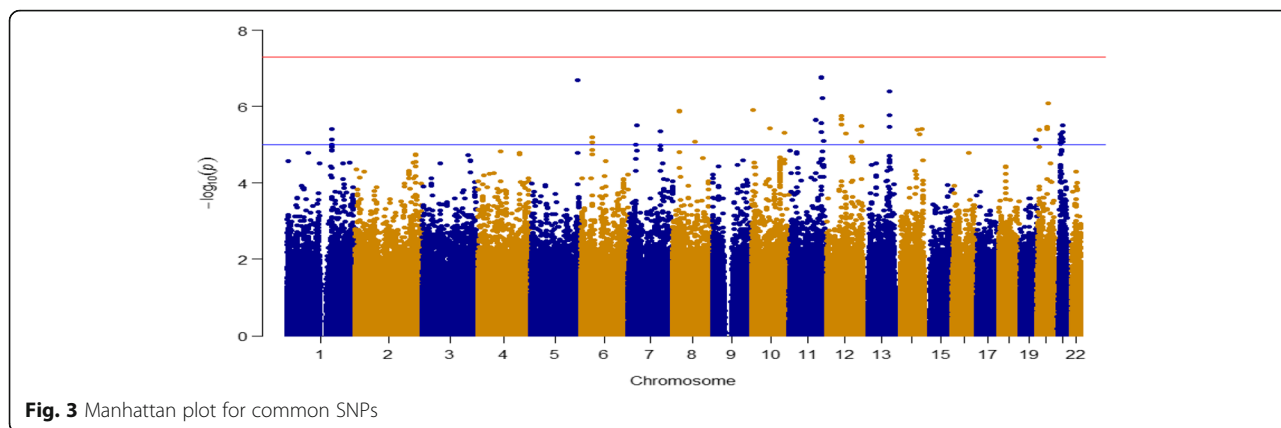
data [3]. The FFBSKAT method was implemented using the Family REGional Association Tests (*FREGAT*) R package [4].

Results

Drug effects are triglyceride-lowering on average but with big variations

There are 1105 participants with phenotypes and covariates available, 4151 participants with pedigree information available, and 822 participants with genotypes available for a dense set of 718,542 SNPs from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study in the GAW20 data [5]. Our quality control (QC) step filtered out SNPs and individuals with a success rate of less than 97%, leaving 822 persons and 700,763 SNPs after QC. Note that the maximum genotype missing rate for an individual is 2.93% so that all 822 participants passed QC. The intersection of the 822 participants with genotypes, the 1105 participants with phenotype and covariates, and the 4151 participants with pedigree information includes 821 common individuals. Note that these 821 individuals have missing values in genotypes, phenotypes, and covariates, and we *did not* restrict our analysis only to the individuals with complete data. The 821 individuals are from 173 families. Thus, they are related individuals (familial data). We conducted analysis using a linear mixed model considering the relatedness within families. Nongenetic covariates include gender, age, field center (Minnesota and Utah), smoking status (never, past, and current smoker), metabolic syndrome defined by the adult treatment panel (ATP), and metabolic syndrome defined by the International Diabetes Federation (IDF) in the GAW20 [5].

Figure 1 shows a histogram of drug response ΔY (ie, changes in the level of triglycerides). We found there was a 50.37 mg/dL decrease on average, indicating the overall drug effect is triglyceride-lowering. However, there was a big variation in drug response ΔY, implying differential drug effects.

Nongenetic factors and ancestry

We next studied the effects of nongenetic factors and ancestry, that is, population structure. Ancestry was represented by the top 10 principal components (PCs) of genotypes at independent SNPs. We first used PLINK to prune SNPs based on linkage disequilibrium to generate a set of independent SNPs using the default setting, that is, squared correlation of < 0.1 [6]. Then we used the eigenstate software to calculate PCs of genotypes [7].

We conducted *multiple regressions* with more than one covariate considering that pairwise analyses (eg, the analysis of a drug response and only one covariate) suffer from confounding effects, cannot control for other covariates, and are less reliable. The regression of drug effects on nongenetic factors (age, center, gender, smoking, IDF, and ATP) and ancestry (PC_1 to PC_{10}) was conducted. Linear mixed-model–based testing was used because of the relatedness of individuals in familial data that was implemented using the *FREGAT* package of R software. A theoretical kinship matrix was calculated from pedigree information using the R (version 3.3.1) package *kinship2* [8]. We found statistical significance in center (p value = 0.013) and ATP (p value = 1.88×10^{-5}), but no significance

Table 1 Top 10 SNPs in GWAS for common SNPs

Chr	Base pair	SNP name	p value	q value	FDR < 5%
11	116,154,127	rs964184	1.73E-07	3.74E-02	Yes
11	116,208,850	rs5128	1.76E-07	3.74E-02	Yes
5	170,899,076	rs919758	2.02E-07	3.74E-02	Yes
13	96,316,286	rs9516776	3.91E-07	4.84E-02	Yes
11	119,259,136	rs503175	6.05E-07	5.44E-02	No
20	37,617,241	rs4812401	8.06E-07	5.77E-02	No
10	4,516,700	rs10795173	1.22E-06	6.14E-02	No
8	25,544,459	rs10081452	1.29E-06	6.19E-02	No
8	25,549,340	rs1425739	1.32E-06	6.21E-02	No
13	96,315,120	rs7332653	1.68E-06	6.37E-02	No

Xu *et al. BMC Proceedings* 2018, **12**(Suppl 9):46

Page 252 of 258
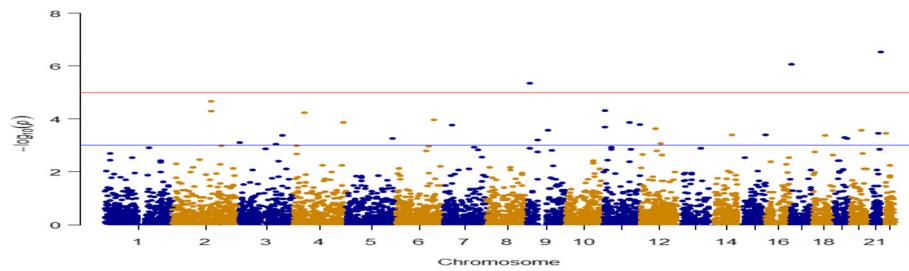
Fig. 4 Manhattan plot for gene-based rare-variant testing

in age (p value = 0.053), gender (p value = 0.126), smoking (p value = 0.067), IDF (p value = 0.137), or PC_1 to PC_{10}. These findings, especially the insignificance of PC_1 to PC_{10}, were consistent with the results of other GAW20 groups, even though different analysis frameworks were used.

Genome-wide association study of common SNPs
We divided SNPs with a MAF ≥ 0.01 into common and rare SNPs. Figure 2 is a histogram of MAFs. Because we have only 821 related individuals from 173 families, making the effective sample size smaller than 821, we considered SNPs with a MAF ≥ 0.05 as common, and SNPs with 0.01 ≤ MAF < 0.05 as rare [9]. There are 574,602 common SNPs in our analysis. We conducted a genome-wide association study (GWAS) for common SNPs for familial data, implemented using the R package GWAF [10]. We converted p values into false discovery rate (FDR) q values using Benjamini and Hochberg's method, implemented in the R package fdrtool [11, 12]. We controlled for an FDR ≤ 0.05 in our report. Figure 3 shows the Manhattan plot and Table 1 lists the top 10 SNPs.

Gene-based and region-based rare-variant association testing results
We used the R package *FREGAT* to conduct gene-based and region-based rare-variant association testing based on familial data [3, 4]. There are 63,689 rare SNPs (ie, 0.01 ≤ MAF < 0.05) in our analysis. We downloaded gene annotations (UCSC build hg19) including 57,816 genes. SNPs that lie within 1 kb of the flanking region upstream and downstream of each gene were considered as promoters are usually within 1 kb of the associated gene transcription start site; 6138 genes included at least one rare SNP for testing. Figure 4 shows a Manhattan plot using the midpoints of genes as base pair locations and Table 2 lists the top 10 genes. There are 6 genes with an FDR < 0.05, namely, *DNMT3L, SPATA22, RP11-403H13.1, AC010740.1, OR52N4,* and *LRP1B.*
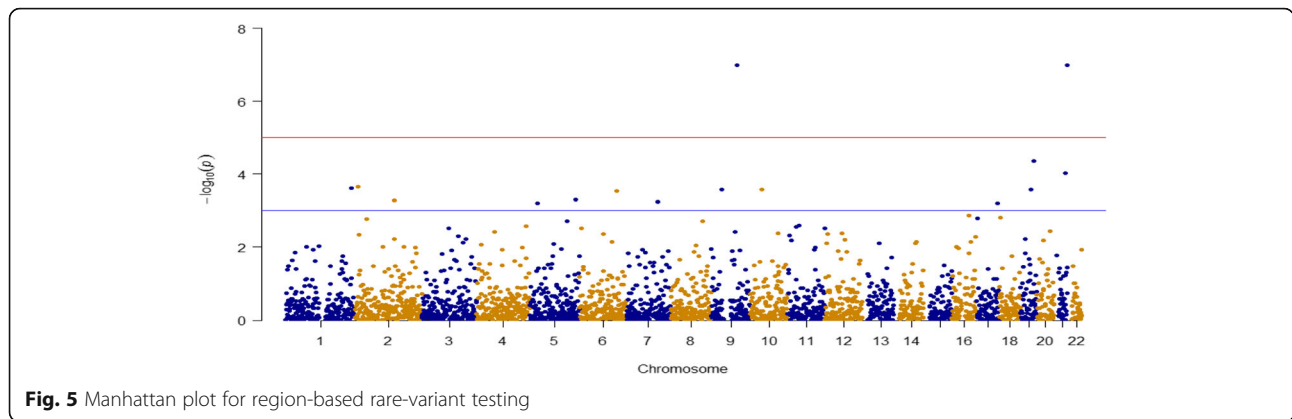
We also conducted a region-based rare-variant test with every 1 Mb as 1 region. The whole genome was divided into 2686 regions. Figure 5 shows a Manhattan plot using the midpoint of regions as base pair locations and Table 3 lists the top 10 regions. There are 3 regions with an FDR ≤ 0.05. They are Chr9: 94 M–95 M, Chr21: 45 M–46 M, and Chr19: 49 M–50 M. It is encouraging that our previously discovered gene *DNMT3L* (Chr21: 45,666,222-45,682,099) also lies in our reported range Chr21: 45 M–46 M.

Discussion
A lot of factors have the potential to influence drug responses, shown as a change in triglyceride levels, to fenofibrate treatment. This study is based on a linear

Table 2 Top 10 genes in gene-based rare-variant test

Gene	Chr	Base pair of midpoint	p value	q values	FDR < 5%
DNMT3L	21	45,674,160	2.97E-07	1.81E-03	Yes
SPATA22	17	3,380,229	8.71E-07	2.66E-03	Yes
RP11-403H13.1	9	6,940,764	4.31E-06	8.76E-03	Yes
AC010740.1	2	141,656,333	2.14E-05	3.06E-02	Yes
OR52N4	11	5,776,441	4.74E-05	4.67E-02	Yes
LRP1B	2	141,939,131	4.94E-05	4.75E-02	Yes
RP11-722 M1.1	4	36,570,840	5.84E-05	5.09E-02	No
AHI1	6	135,711,792	1.07E-04	7.10E-02	No
FAM76B	11	95,512,839	1.33E-04	7.81E-02	No
NEIL3	4	178,257,543	1.35E-04	7.88E-02	No

Xu *et al. BMC Proceedings* 2018, **12**(Suppl 9):46

Page 253 of 258

Fig. 5 Manhattan plot for region-based rare-variant testing

mixed-model multiple regression (1) for nongenetic factors and ancestry variables, (2) for common SNPs, and (3) for rare SNPs. These analyses were performed using the R packages *FREGAT* and *GAWF*, with sample relatedness represented by theoretical kinship matrix, calculated using the R package *kinship2*.

Our analysis is based on 821 persons from 173 families. The effective sample size of this familial data is smaller than 821; consequently, there may not be enough power to identify associated variants. In addition, our GWAS analysis for common SNPs was based on a set of 574,602 SNPs without imputation. Imputation of genotype or summary statistics may uncover more associated SNPs, thereby increasing power [13, 14].

Despite a relatively small effective sample size, we still found that for nongenetic factors, some variables had *p* values < 5%, some had *p* values between 5 and 10%, and other variables had *p* values > 10%; and that for SNPs, there are 4 SNPs, 6 genes, and 3 regions of 1 Mb reported with an FDR controlled at 5%.

The roles of the top SNP, rs964181, and the top gene, *DNMT3L*, were also found in other published studies of obesity and triglyceride levels. The top SNP, rs964184, was found to be associated with hypertriglyceridemia

Table 3 Top 10 regions in region-based rare-variant test

Chr	Start base pair	End base pair	*p* value	*q* value	FDR < 5%
9	94M	95M	1.01E-07	1.36E-04	Yes
21	45M	46M	1.02E-07	1.36E-04	Yes
19	49M	50M	4.40E-05	3.66E-02	Yes
21	39M	40M	9.59E-05	5.53E-02	No
2	11M	12M	2.17E-04	7.30E-02	No
1	238M	239M	2.45E-04	7.52E-02	No
10	42M	43M	2.64E-04	7.64E-02	No
9	38M	39M	2.67E-04	7.66E-02	No
9	37M	38M	2.68E-04	7.67E-02	No

[15], as well as with a lipid-lowering response to another medicine, statins [16]. The top gene, *DNMT3L*, is an enzymatically inactive regulatory factor, regulates DNA methylation activity, and is closely associated with epigenetic functions influencing obesity from epigenetic and regulation evidence [17]. *DNMT3L* encodes a DNA (cytosine-5)-methyltransferase 3–like enzyme, and an increased expression of DNA methyltransferase is found in obese adipose tissue [18]. A DNA methylation study revealed differential modification of many obesity genes before and after gastric bypass and weight loss, providing a model to investigate obesity and weight loss in humans [19].

The above association results only suggest and prioritize potential factors for future biological verification. Some reported significant variables may be just false positives. Following statistical analyses, functional analyses via biologically experimental verification and additional support from the published literature are needed. Integrative genome browsers with the database of GWAS catalog, gene annotations, and epigenetic and regulatory information can be used for this purpose [20, 21].

Conclusions

We conducted an assessment of nongenetic and genetic factors that impact the drug response, shown as a change in triglyceride level, to fenofibrate treatment based on the GOLDN study data, and identified groups of participants with different drug sensitivities. We report significant associations of drug response with center and ATP variables with *p* values less than 5%, and 4 common SNPs (rs964184, rs5128, rs919758, and rs9516776), 6 genes (*DNMT3L*, *SPATA22*, *RP11-403H13.1*, *AC010740.1*, *OR52N4*, and *LRP1B*) and 3 regions of 1 Mb (Chr9: 94 M– 95 M, Chr21: 45 M–46 M, and Chr19: 49 M–50 M) at an FDR controlled at 0.05. It is also encouraging that the reported gene *DNMT3L* (Chr21: 45,666,222-45,682,099, from a gene-based test) also lies in our reported range of Chr21:

Xu *et al. BMC Proceedings* 2018, **12**(Suppl 9):46

Page 254 of 258

45 M-46 Mb (from a range-based test). The roles of the top SNP, rs964184, and the top gene, *DNMT3L*, were also found in other studies on obesity and triglycerides. Both gene-based and region-based tests implied that *DNMT3L* plays a crucial role in influencing the mechanism and effects of triglyceride-lowering drugs treating obesity. Our methodology can be applied to studying other drugs, such as statins, and provides an approach to the development of personalized medicine.

Availability of data and materials
The data that support the findings of this study are available from the Genetic Analysis Workshop (GAW), but restrictions apply to the availability of these data, which were used under license for the current study. Qualified researchers may request these data directly from GAW.

About this supplement
This article has been published as part of BMC Proceedings Volume 12 Supplement 9, 2018: Genetic Analysis Workshop 20: envisioning the future of statistical genetics by exploring methods for epigenetic and pharmacogenomic data. The full contents of the supplement are available online at https://bmcproc.biomedcentral.com/articles/supplements/volume-12-supplement-9.

Authors' contributions
ZX designed the whole analysis framework, ZX, QD, YQ and CW performed the data analysis and ZX, JCui, YQ, QJ, CW, JClarke wrote the manuscript. All the authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Statistics, University of Nebraska, 340 Hardin Hall North Wing, Lincoln, NE 68588, USA. [2]Department of Food Science and Technology, University of Nebraska, 1901 N. 21st Street, Lincoln, NE 68588, USA. [3]Department of Genetics, University of North Carolina, 120 Mason Farm Road, Chapel Hill, NC 27599, USA. [4]Department of Computer Science and Engineering, University of Nebraska, 256 Avery Hall, Lincoln, NE 68508, USA. [5]Center for Biotechnology, University of Nebraska, 1901 Vine Street, Lincoln, NE 68588, USA. [6]Jiangxi Provincial Key Laboratory for Animal Health and Institute of Animal Population Health, College of Animal Science and Technology, Jiangxi Agricultural University, Nanchang 330045, Jiangxi, China.

Published: 17 September 2018

References
1. Ogden CL, Carroll MD, Fryar CD, Flegal KM. Prevalence of obesity among adults and youth: United States, 2011-2014. NCHS Data Brief. 2015;219:1–8.
2. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. JAMA. 2012;307(5):483–90.
3. Svishcheva GR, Belonogova NM, Axenovich TI. FFBSKAT: fast family-based sequence kernel association test. PLoS One. 2014;9(6):e99407.
4. Belonogova NM, Svishcheva GR, Axenovich TI. FREGAT: an R package for region-based association analysis. Bioinformatics. 2016;32(15):2392–3.
5. Warodomwichit D, Shen J, Arnett DK, Tsai MY, Kabagambe EK, Peacock JM, Hixson JE, Straka RJ, Province MA, An P, et al. ADIPOQ polymorphisms, monounsaturated fatty acids, and obesity risk: the GOLDN study. Obesity (Silver Spring). 2009;17(3):511–7.
6. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ, Second-generation PLINK. Rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7.
7. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2(12):e190.
8. Sinnwell JP, Therneau TM, Schaid DJ. The kinship2 R package for pedigree data. Hum Hered. 2014;78(2):91–3.
9. Kang J, Huang KC, Xu Z, Wang Y, Abecasis GR, Li Y. AbCD: arbitrary coverage design for sequencing-based genetic studies. Bioinformatics. 2013;29(6):799–801.
10. Chen M-H, Yang Q. GWAF: an R package for genome-wide association analyses with family data. Bioinformatics. 2010;26(4):580–1.
11. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B. 1995;57(1):289–300.
12. Strimmer K. Fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics. 2008;24(12):1461–2.
13. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34(8):816–34.
14. Xu Z, Duan Q, Yan S, Chen W, Li M, Lange E, Li Y. DISSCO: direct imputation of summary statistics allowing covariates. Bioinformatics. 2015;31(15):2434–42.
15. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, et al. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet. 2010;42(8):684–7.
16. Barber MJ, Mangravite LM, Hyde CL, Chasman DI, Smith JD, McCarty CA, Li X, Wilke RA, Rieder MJ, Williams PT, et al. Genome-wide association of lipid-lowering response to statins in combined study populations. PLoS One. 2010;5(3):e9763.
17. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, Li Y, Lin S, Lin Y, Barr CL, et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. Cell Rep. 2016;17(8):2042–59.
18. Kamei Y, Suganami T, Ehara T, Kanai S, Hayashi K, Yamamoto Y, Miura S, Ezaki O, Okano M, Ogawa Y. Increased expression of DNA methyltransferase 3a in obese adipose tissue: studies with transgenic mice. Obesity (Silver Spring). 2010;18(2):314–21.
19. Benton MC, Johnstone A, Eccles D, Harmon B, Hayes MT, Lea RA, Griffiths L, Hoffman EP, Stubbs RS, Macartney-Coxson D. An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. Genome Biol. 2015;16:8.
20. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, et al. The UCSC genome browser database: 2016 update. Nucleic Acids Res. 2016;44(D1):D717–25.
21. Xu Z, Zhang G, Duan Q, Chai S, Zhang B, Wu C, Jin F, Yue F, Li Y, Hu M. HiView: an integrative genome browser to leverage Hi-C results for the interpretation of GWAS variants. BMC Res Notes. 2016;9(1):159.