

PROCEEDINGS

Open Access



# Family-based genome-wide association of inflammation biomarkers and fenofibrate treatment response in the GOLDN study

Sarmistha Das<sup>†</sup>, Pronoy Kanti Mondal<sup>†</sup>, Saurabh Ghosh and Indranil Mukhopadhyay<sup>\*</sup>

From Genetic Analysis Workshop 20  
San Diego, CA, USA. 4 - 8 March 2017

## Abstract

In this paper we analyzed whole-genome genetic information provided by GAW20 from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study for family data. Lipid levels such as triglycerides (TGs) and high-density lipoprotein (HDL) are measured at different time points before and after administration of an anti-inflammatory drug fenofibrate. Apart from that, the data contain some covariates and whole-genome genotype information. We propose 2 novel approaches based on Henderson's iterative mixed model to identify associated loci corresponding to (a) inflammatory biomarkers like TGs and HDLs together over time, and (b) the response to fenofibrate treatment. We developed a mixed-model approach using both TG and HDL phenotypes at all 4 time points for a genetic association study whereas we used TGs only to study genetic association with response to the drug. We expect that use of complete family data in a longitudinal framework under a single model involving the appropriate correlation structures would be able to exploit the maximum possible information contained in the sample. Our analysis of whole-genome single nucleotide polymorphisms (SNPs) and genomic regions corresponding to drug treatment finds no significant locus after multiple correction. Arguably, the moderately small sample size of the data set, as compared to the sample size usually used in genome-wide association studies (GWAS), could be a reason for such a result. Nevertheless, we report the top 20 SNPs associated with the phenotypes, and the top 20 SNPs and genomic regions associated with a response to fenofibrate treatment. Application of our methods to larger GWAS and further functional validation of the reported top SNPs and genomic regions might provide important biological insight into the genetic constitution of the trait.

## Background

Understanding the genetic architecture underlying complex phenotypes is crucial in decoding disease mechanisms as well as treatment and drug development. Genome-wide association studies (GWAS) have contributed significantly to the identification of associated variants with numerous traits. Although the sample size requirement of GWAS is high, the proportion of the disease risk explained by a single variant always remains low. However, availability of longitudinal data on multiple phenotypes might carry more information in identifying

associated variants. The Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study provides a data set with triglyceride (TG) and high-density lipoprotein (HDL) levels at 4 time points for a fixed set of families with few missing observations during follow-up. Consequently, given a moderate sample size, reduction of the multiple testing burden and/or use of longitudinal information is required. Moreover, with available data on multiple interacting phenotypes, it is informative to study both the inherent environmental correlation and the genetic correlation.

Increased levels of TG and decreased HDL levels are well-known causes of heart disease. So we developed one model that captures the genome-wide genetic association with interacting phenotypes, that is, TG and HDL, over time by introducing both a temporal

\* Correspondence: [indranil@isical.ac.in](mailto:indranil@isical.ac.in)

<sup>†</sup>Sarmistha Das and Pronoy Kanti Mondal contributed equally to this work.  
Human Genetics Unit, Indian Statistical Institute, 203 B T Road, Kolkata  
700108, India



covariance structure and a genetic covariance structure between phenotype measurements. Our method is expected to increase power as a result of the inclusion of more information through genetic and environmental correlation structures.

On the other hand, fenofibrate is an antiinflammatory drug, well known for TG-lowering effects. Some studies report modulation of a lipid response to fenofibrate as the result of genetic variants involved in lipid metabolism [1], but the response to treatment by fenofibrate varies across individuals in a population [2]. Some GWAS ventured to find associated variants behind such a response, but met no success in the sense of identifying variants significantly associated with fenofibrate response [3]. The reasons might be low sample size, noncomparable baseline lipid profiles, and environmental exposures of the individuals. Thus, along with simple GWAS, we studied multiloci association of response to fenofibrate treatment with the genomic regions, which reduces the chance of missing a moderately associated locus. We also examined the association of single variants using multiple TG and HDL phenotypes in the GOLDN study. We found that 1 single nucleotide polymorphism (SNP) is associated with the TG and HDL phenotypes, although we did not find any significant SNP or gene that is associated with the drug response. It is important to note that because the sample in this study is not very large, we report a few top significant loci that might be associated with phenotype and response to drug.

**Methods**

We analyzed the real data provided by GAW20 (ie, GOLDN study) data set [4]. The data set contains data on age, smoking status, etc. as covariates, pedigree information, and genome-wide genetic variation, as well as TG and HDL levels measured before and after the drug at 4 time points. Information on genetic variation was available for 822 individuals, while other information, except kinship structure, had a sample size of 1105 individuals. Pedigree data was available for 4151 individuals. The kinship structure, covariates, and TG and HDL phenotypes were available for all 822 individuals, but genotype information was missing for 1 individual. Consequently, in the subsequent analysis we used the remaining 821 subjects. Next, the missing genotypes and monomorphic SNPs were removed from the analysis. For variants with only 2 observed genotypes, we eliminated the SNP if 1 genotype frequency was < 5%. We imputed missing phenotype data using a mixed-model approach under the null model, and used log–log transformation of the phenotype variables for the entire analysis. This transformation made the data normally distributed and, hence, the resultant test statistic followed a standard distribution under a null hypothesis (H<sub>0</sub>) of no association. During imputation, we assumed constant heritability, and this value was from an existing

study [5]. We calculated *p* values using the asymptotic distribution of test statistics under H<sub>0</sub> after Benjamini-Hochberg (BH) correction.

To meet the objectives as stated in the previous section, we first do a GWAS based on longitudinal data with TG and HDL together as phenotypes. We use a mixed model that includes environmental as well as genetic correlation structure.

With TG and HDL at all time points as response vector, our model is:

$$\begin{pmatrix} Y^{TG} \\ Y^{HDL} \end{pmatrix} = \begin{pmatrix} X^{TG}\beta^{TG} + Z^{TG}u^{TG} \\ X^{HDL}\beta^{HDL} + Z^{HDL}u^{HDL} \end{pmatrix} + \begin{pmatrix} \epsilon^{TG} \\ \epsilon^{HDL} \end{pmatrix} \tag{1}$$

where  $Y^{TG} = (Y_1^{TG}, Y_2^{TG}, Y_3^{TG}, Y_4^{TG})'$  and  $Y^{HDL} = (Y_1^{HDL}, Y_2^{HDL}, Y_3^{HDL}, Y_4^{HDL})'$  denote TG and HDL respectively, at 4 time points, for *n* individuals. In this model,  $X^{TG}$  and  $X^{HDL}$  are fixed effects design matrices, where  $X^{TG} = X^{HDL} = [I_4 \otimes 1_n \ 1_4 \otimes g_n]$ ,  $g_n$  is the genotype vector for *n* individuals at a single marker locus,  $u^{TG}$  and  $u^{HDL}$  are random effect vectors for *n* individuals, and  $Z^{TG} = Z^{HDL} = 1_4 \otimes I_n$  is the corresponding matrix. Here  $\beta^{TG} = (\beta_1^{TG}, \beta_2^{TG}, \beta_3^{TG}, \beta_4^{TG}, \beta_5^{TG})'$ , where the first 4 components are temporal effects for 4 different time points that includes the drug effect and  $\beta_5^{TG}$  is the effect of the SNP. We assume that,  $Var(u^{TG}) = \sigma_{u,TG}^2 K$ ,  $Var(u^{HDL}) = \sigma_{u,HDL}^2 K$  where *K* is the kinship matrix,  $Var(\epsilon^{TG}) = \Sigma_{TG} \otimes I_m$ ,  $Var(\epsilon^{HDL}) = \Sigma_{HDL} \otimes I_m$

$$\Sigma_{TG} = \sigma_{TG}^2 \begin{pmatrix} 1 & \rho_{1,TG} & \rho_{1,TG}\rho_{2,TG} & \rho_{1,TG}\rho_{2,TG}\rho_{3,TG} \\ \rho_{1,TG} & 1 & \rho_{2,TG} & \rho_{2,TG}\rho_{3,TG} \\ \rho_{1,TG}\rho_{2,TG} & \rho_{2,TG} & 1 & \rho_{3,TG} \\ \rho_{1,TG}\rho_{2,TG}\rho_{3,TG} & \rho_{2,TG}\rho_{3,TG} & \rho_{3,TG} & 1 \end{pmatrix}$$

where  $\sigma_{TG}^2$  is the common variance of TG at all 4 time points and the correlation coefficient matrix is parameterized by three parameters:  $\rho_{1, TG}$ ,  $\rho_{2, TG}$  and  $\rho_{3, TG}$ . Similarly we define  $\beta^{HDL}$  and  $\Sigma_{HDL}$ . Now, denoting  $\rho_g$  and  $\rho_e$  as genetic and environmental correlations respectively, we assume the correlation structure as:

$$Var \begin{pmatrix} Y^{TG} \\ Y^{HDL} \end{pmatrix} = Z(\Sigma_{u,TG,HDL} \otimes K)Z' + (\Sigma_{\epsilon,TG,HDL} \otimes I_n)$$

where  $Z = \begin{pmatrix} Z^{TG} & 0 \\ 0 & Z^{HDL} \end{pmatrix}$ ,

$$\Sigma_{\epsilon,TG,HDL} = \begin{pmatrix} \Sigma_{TG} & \rho_e \Sigma_{TG}^{\frac{1}{2}} \Sigma_{HDL}^{\frac{1}{2}} \\ \rho_e \Sigma_{HDL}^{\frac{1}{2}} \Sigma_{TG}^{\frac{1}{2}} & \Sigma_{HDL} \end{pmatrix}$$

$$\Sigma_{u,TG,HDL} = \begin{pmatrix} \sigma_{u,TG}^2 & \rho_g \sigma_{u,TG} \sigma_{u,HDL} \\ \rho_g \sigma_{u,HDL} \sigma_{u,TG} & \sigma_{u,HDL}^2 \end{pmatrix}$$

We test the null hypothesis of no genetic association of an SNP with TG and HDL using a likelihood ratio test. The asymptotic distribution of the log-likelihood

ratio statistic can be shown to follow a  $\chi^2_2$  distribution. We apply this test at the GWAS level after appropriate multiple-testing correction.

To address our second objective (ie, to test association of response to fenofibrate treatment), we model our data using Henderson’s mixed-model approach with adequate modification. Incorporating correlation structure among family members, we propose our model as

$$Y = X\beta + Zu + \varepsilon \tag{2}$$

where  $Y$  is the vector of changes in phenotype (measured as log-log of TG) before and after the drug treatment;  $X$  is the design matrix of covariates, namely, age, smoking status, and SNP/genotypes in a genomic region (not in linkage disequilibrium);  $\beta$  is the associated fixed effect parameter;  $Z$  is the design matrix of random components;  $u$  is the random effect of the family; and  $\varepsilon$  is the error component. We assume  $u \sim N(0, \sigma_g^2 K)$ , where  $K$  is the kinship matrix and  $\varepsilon \sim N(0, \sigma_e^2 R)$  independently of  $u$ . Hence,  $V(Y) = \sigma_g^2 ZKZ' + \sigma_e^2 R$ . Note that because we are dealing with family data, a non-diagonal positive definite matrix  $R$  appears in the variance-covariance matrix of  $\varepsilon$ .

During analysis, we use 778 individuals after removing those with no response either before or after drug treatment. But in case of missing response at one of the time points before (after) drug treatment, we impute it with the other response value. To calculate kinship matrix we use R package “kinship2” with the entire family data. To find association of genomic regions, we first identify the genomic regions and then remove the SNPs that are in linkage disequilibrium ( $r^2 > 0.5$ ). The genomic regions are basically (a) the genic regions, and (b) the intergenic regions lying between 2 consecutive genes, that overlap the genotyped SNPs in our data. We use Henderson’s iterative procedure for mixed model approach [6] after substantial modification and after adjusting for age, smoking status as fixed effects, and random genetic effect within a family. We use the restricted maximum likelihood (REML) approach to test our  $H_0$  of no association, adopting the expectation-maximization (EM) algorithm for parameter estimation.

Maximization of joint likelihood of  $Y$  and  $u$  and eq. (3) [6] provide the best linear unbiased predictors (BLUPs) for the random component under normal assumption of the response variable.

$$\begin{pmatrix} X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{u} \\ ZR^{-1}X\hat{\beta} + (Z'R^{-1}Z + D^{-1})\hat{u} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix} \tag{3}$$

So to test the association of (a) whole-genome SNPs and (b) genomic regions, with response to the drug treatment, our null hypothesis will be,  $H_0 : M\beta = 0$  for an

arbitrary  $p \times q$  matrix  $M$  with  $rank(M) = p$ . Thus, if  $n$  be the number of observations and  $rank(M) = p$ , the test statistic [7],

$$F = \frac{(M\hat{\beta})' (M(X'\hat{V}^{-1}X)^{-1}M')^{-1} (M\hat{\beta})}{rank(M)} \sim F(p, n-q), \text{ under } H_0 \tag{4}$$

where  $\hat{V} = Var(\hat{\beta}) = X'\hat{V}_y^{-1}X$  and  $\hat{V}_y = \widehat{Var}(y)$ .

We developed the above procedure to test association of multiple SNPs (genomic regions), which reduces the multiple-testing burden. However, this test can be seen as a single-marker association test that we have used to perform our GWAS study with appropriate multiple-testing correction.

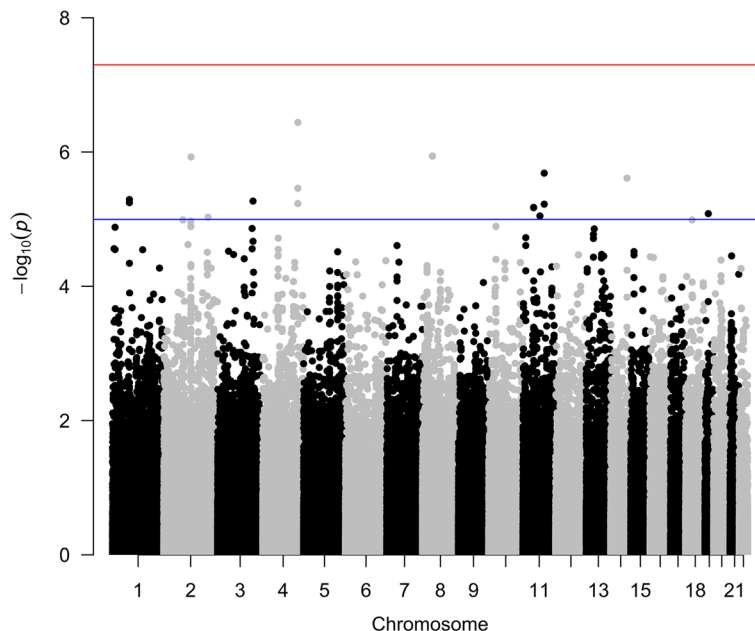
### Results

With longitudinal information using multiple phenotypes we identified only 1 significant SNP (Fig. 1) after BH correction. The SNP is rs2880301, located at *TPTE2* in an intron. rs2880301 is reported to be associated with HDL particle diameter and low-density lipoprotein (LDL) particle diameter [8] and is also known to confer protection against hepatocellular carcinoma [9]. However, we think that there might be other SNPs that remain unidentified as a consequence of small sample size. Hence we report the top 20 SNPs based on  $p$  value in Table 1. rs752273 is reported to be associated with cardiovascular diseases [10] while rs2896368 is known to be associated with  $\alpha_1$ -antitrypsin level [11].

To test the null hypothesis of no association with drug response, we examined 243,593 whole-genome SNPs and 18,266 genomic regions. The genomic regions in our study are (a) genic, that overlap the genotyped SNPs, and (b) intergenic, that lie between 2 consecutive genes and overlap the genotyped SNPs in the data set. After BH correction, none of the SNPs nor genomic regions showed significant association with the drug response (Fig. 2). However, we report the top 20 SNPs (Table 2) and top 20 genomic regions (Table 3). The moderately small sample size of the data compared to most of the GWAS might be a reason behind this result. Application of our methods to larger GWAS and further functional validation of the reported top loci might provide some directive for studying inflammatory biomarkers and response to fenofibrate treatment.

### Discussion and conclusions

In this paper, we developed novel methods for (a) GWAS using longitudinal data and (b) GWAS/genomic region association with response to fenofibrate treatment based on a family-based design. These methods are



**Fig. 1** Manhattan plot of genome-wide  $p$  values of SNPs on interacting phenotypes, namely, TG and HDL

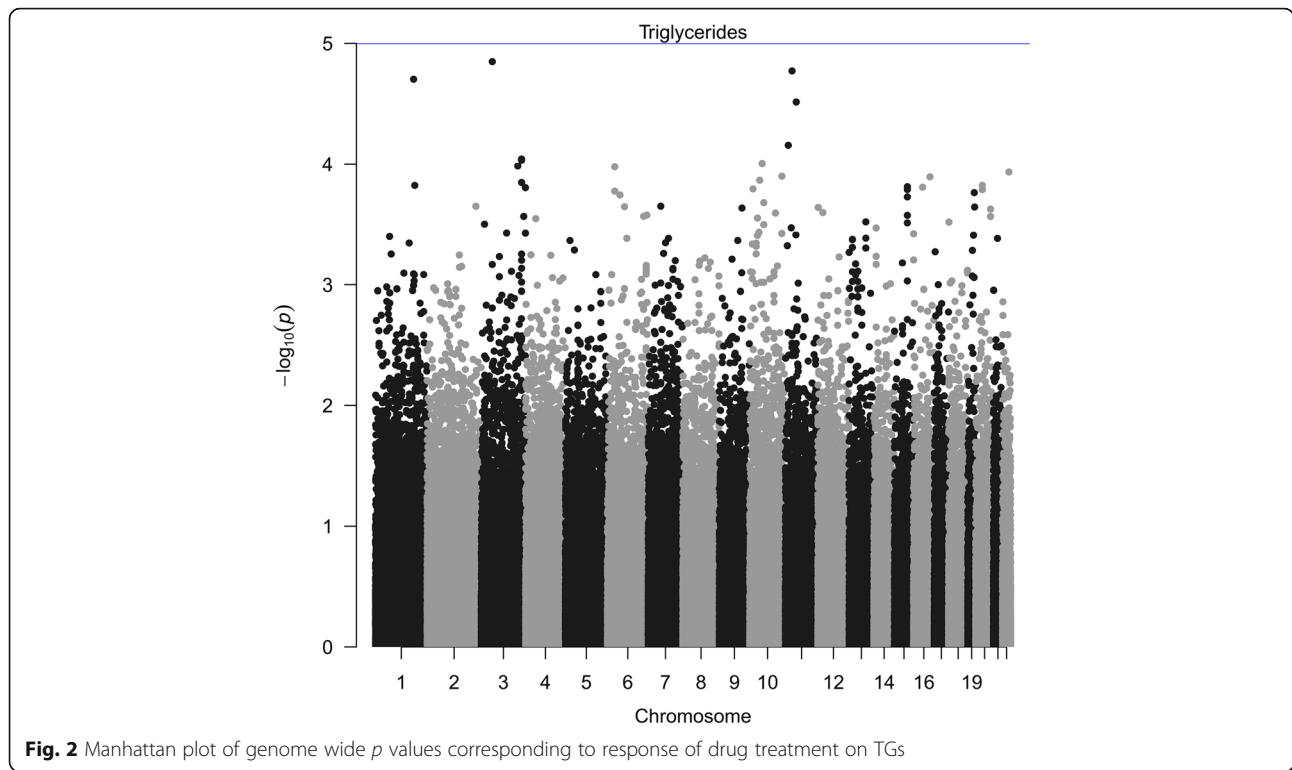
**Table 1** Top 20 SNPs associated with TG and HDL

Chr	SNPs	$p$ Value	Base position
13	rs2880301	$1.095660 \times 10^{-9}$	18,998,534
4	rs1909882	$3.610927 \times 10^{-7}$	170,922,834
4	rs12510928	$3.610927 \times 10^{-7}$	170,954,378
8	rs6558172	$1.148044 \times 10^{-6}$	22,547,997
2	rs752273	$1.183976 \times 10^{-6}$	108,297,537
11	rs1793368	$2.066321 \times 10^{-6}$	96,521,084
14	rs2896268	$2.448163 \times 10^{-6}$	93,935,461
13	rs2770297	$2.707521 \times 10^{-6}$	46,344,666
4	rs6835612	$3.468092 \times 10^{-6}$	170,955,790
1	rs4844913	$5.117515 \times 10^{-6}$	208,134,740
3	rs6440833	$5.392711 \times 10^{-6}$	154,128,934
1	rs924297	$5.693772 \times 10^{-6}$	76,996,366
4	rs13113929	$5.873701 \times 10^{-6}$	169,320,943
11	rs1255523	$5.999600 \times 10^{-6}$	95,019,155
11	rs395297	$6.692577 \times 10^{-6}$	37,093,845
19	rs4805303	$8.219183 \times 10^{-6}$	34,106,471
11	rs7929919	$8.890781 \times 10^{-6}$	78,529,128
2	rs2353319	$9.293606 \times 10^{-6}$	204,594,629
2	rs4676175	$1.016761 \times 10^{-5}$	108,054,810
18	rs2419041	$1.017765 \times 10^{-5}$	26,293,205

agnostic to the choice of phenotype and can be generalized to any such study. Although we could not detect any novel biologically relevant locus that is significantly associated with response to fenofibrate treatment, we identified a few loci that are associated with TG and HDL levels. Our belief is that the primary reason for obtaining only a small number of significant association findings is the much smaller sample size in our analyses as compared to conventional GWAS. Validation in a larger sample might throw more light on the roles of the top few significantly associated SNPs and/or genomic regions in controlling TG and HDL levels. Nevertheless, this study emphasizes the effect of administering fenofibrate to individuals with specific genetic profiles.

We pruned the available set of SNPs to an independent set of SNPs in our GWAS, primarily to reduce the multiple-testing burden. Because many studies impute SNPs, and hence augment the number of available SNPs, to explore association findings for previously reported SNPs that have not been genotyped, our strategy has a caveat in the sense of reduction in the overall power of the GWAS. Similarly, while our proposed method involves simultaneous testing of multiple SNPs within a gene in order to evaluate association at the gene level, it may yield lower powers compared to the usual single SNP analyses in GWAS.

We imputed the missing phenotype data using a known heritability value [5] and have applied the EM algorithm. Although studies show that such imputation



**Table 2** Top 20 SNPs associated with fenofibrate treatment response

Chr	SNPs	$p$ Value
1	rs7555566	$1.982194 \times 10^{-5}$
1	rs6664332	$1.499459 \times 10^{-4}$
3	rs1401072	$1.416564 \times 10^{-5}$
3	rs13072632	$9.056298 \times 10^{-5}$
3	rs11564450	$9.281710 \times 10^{-5}$
3	rs1631395	$1.034264 \times 10^{-4}$
3	rs9311268	$1.418409 \times 10^{-4}$
3	rs6599150	$1.418409 \times 10^{-4}$
6	rs13217251	$1.048927 \times 10^{-4}$
10	rs2482583	$9.882478 \times 10^{-5}$
10	rs3122227	$1.257541 \times 10^{-4}$
10	rs7917347	$1.358690 \times 10^{-4}$
11	rs2406928	$1.692430 \times 10^{-5}$
11	rs10790087	$3.064359 \times 10^{-5}$
11	rs1940088	$6.968692 \times 10^{-5}$
15	rs1777266	$1.537405 \times 10^{-4}$
16	rs7197943	$1.272486 \times 10^{-4}$
20	rs6075087	$1.497509 \times 10^{-4}$
20	rs11087178	$1.497509 \times 10^{-4}$
22	rs10427772	$1.160345 \times 10^{-4}$

**Table 3** Top 20 genomic regions associated with fenofibrate treatment

Chr	Genomic regions	$p$ Value
1	<i>intergenic.snps</i> CAPZB LOC105378614	$4.705439 \times 10^{-4}$
1	<i>intergenic.snps</i> RCC2 ARHGEF10L	$6.471417 \times 10^{-4}$
2	<i>snps</i> SCN1A	$6.896508 \times 10^{-4}$
3	<i>snps</i> ABCC5	$1.877981 \times 10^{-4}$
3	<i>snps</i> PQLC2L	$8.087247 \times 10^{-4}$
3	<i>snps</i> IQCJ-SCHIP1	$8.150236 \times 10^{-4}$
4	<i>intergenic.snps</i> C4orf33 LOC101927282	$2.817887 \times 10^{-4}$
5	<i>snps</i> FAM81B	$4.593152 \times 10^{-4}$
5	<i>snps</i> PKD2L2	$5.974084 \times 10^{-4}$
6	<i>intergenic.snps</i> LOC105377967 CEP85L	$1.270536 \times 10^{-4}$
7	<i>intergenic.snps</i> NUPL2 GPNMB	$7.784435 \times 10^{-4}$
8	<i>snps</i> TRPA1	$4.326193 \times 10^{-5}$
9	<i>intergenic.snps</i> TNFSF8 TNC	$1.822392 \times 10^{-4}$
9	<i>intergenic.snps</i> HACD4 IFNB1	$2.491489 \times 10^{-4}$
9	<i>intergenic.snps</i> DEC1 LOC101928775	$6.834450 \times 10^{-4}$
10	<i>intergenic.snps</i> MAP3K8 LYZL2	$2.765323 \times 10^{-4}$
14	<i>snps</i> NEMF	$3.979899 \times 10^{-5}$
18	<i>snps</i> LO $\times$ HD1	$5.430775 \times 10^{-4}$
20	<i>snps</i> LINC01524	$8.431111 \times 10^{-5}$
22	<i>intergenic.snps</i> MIR4762 WNT7B	$4.120036 \times 10^{-4}$

may lead to some loss of power and hence seems to be a limitation of our current method, the intuition behind the imputation strategy was to use the maximum phenotype data in our analyses. A more general model that includes the genotype data can be developed in a likelihood framework for testing association, but this would increase substantial computational complexity while calculating the test statistic.

Association findings based on any real data set are susceptible to being false positives. If these findings validate previous reports of association, they are more likely to be true positives. In case of novel significant findings, it is necessary to either validate them in an independent data set or, alternatively, to perform extensive simulations under similar genotype and phenotype structures to evaluate the false-positive rates of the underlying test procedures.

#### Funding

Publication of this article was supported by NIH R01 GM031575.

#### 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://bmcpoc.biomedcentral.com/articles/supplements/volume-12-supplement-9>.

#### Authors' contributions

SD, PKM, and IM conceived the study and developed the statistical methods; SD and PKM conducted the data analysis; IM and SG formulated the problem; SD, PKM, and IM wrote the paper; and SG reviewed the paper. All authors have 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.

Published: 17 September 2018

#### References

- Brisson D, Ledoux K, Bosse Y, St-Pierre J, Julien P, Perron P, Hudson TJ, Vohl MC, Gaudet D. Effect of apolipoprotein E, peroxisome proliferator-activated receptor alpha and lipoprotein lipase gene mutations on the ability of fenofibrate to improve lipid profiles and reach clinical guideline targets among hypertriglyceridemic patients. *Pharmacogenetics*. 2002;12(4):313–20.
- Shen J, Arnett DK, Parnell LD, Peacock JM, Lai CQ, Hixson JE, Tsai MY, Province MA, Straka RJ, Ordovas JM. Association of common C-reactive protein (CRP) gene polymorphisms with baseline plasma CRP levels and fenofibrate response. *Diabetes Care*. 2008;31(5):910–5.
- Aslibekyan S, Kabagambe EK, Irvin MR, Straka RJ, Borecki IB, Tiwari HK, Tsai MY, Hopkins PN, Shen J, Lai CQ, et al. A genome-wide association study of inflammatory biomarker changes in response to fenofibrate treatment in the genetics of lipid lowering drug and diet network. *Pharmacogenet Genomics*. 2012;22(3):191–7.
- Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, Thibeault KS, Patel N, Day K, Jones LW, et al. Epigenome wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study. *Circulation*. 2014;130(7):565–72.
- Heckerman D, Gurdasani D, Kadie C, Pomilla C, Carstensen T, Martin H, Ekoru K, Nsubuga RN, Ssenyomo G, Kamali A, et al. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc Natl Acad Sci U S A*. 2016;113(27):7377–82.
- Searle SR, Casella G, McCulloch CE. *Variance components*. New York: Wiley; 2006.
- Kennedy BW, Quinton M, van Arendonk JA. Estimation of effects of single genes on quantitative traits. *J Anim Sci*. 1992;70(7):2000–12.
- Frazier-Wood AC, Manichaikul A, Aslibekyan S, Borecki IB, Goff DC, Hopkins PN, Lai CQ, Ordovas JM, Post WS, Rich SS, et al. Genetic variants associated with VLDL, LDL and HDL particle size differ with race/ethnicity. *Hum Genet*. 2013;132(4):405–13.
- Ginsburg GS, Willard HF, David SP. *Genomic and precision medicine: primary care, third edition*. London: Academic Press; 2017.
- Shendre A, Irvin MR, Wiener H, Zhi D, Limdi NA, Overton ET, Shrestha S. Local ancestry and clinical cardiovascular events among African Americans from atherosclerosis risk in communities study. *J Am Heart Assoc*. 2017;6(4)
- Setoh K, Terao C, Muro S, Kawaguchi T, Tabara Y, Takahashi M, Nakayama T, Kosugi S, Sekine A, Yamada R, et al. Three missense variants of metabolic syndrome-related genes are associated with alpha-1 antitrypsin levels. *Nat Commun*. 2015;6:7754.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

