

PROCEEDINGS

Open Access



Homozygosity disequilibrium associated with treatment response and its methylation regulation

Hsin-Chou Yang* and Chia-Wei Chen

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

Abstract

Homozygosity disequilibrium (HD), indicating a nonrandom pattern of sizable runs of homozygosity that deviates from a random allocation of homozygous and heterozygous genotypes in the genome, is an important phenomenon in population genomics and medical genomics. We performed the first genome-wide study investigating the roles of HD in pharmacogenomics and pharmacoepigenomics by analyzing GAW20 data. We inferred whole-genome profiles of homozygosity intensities and performed genome-wide homozygosity association analyses to identify regions of HD associated with triglyceride (TG) response to fenofibrate by using LOHAS (Loss-of-Heterozygosity Analysis Suite) software. The analysis identified a region of HD contained in *MACROD2* at 20p12 to be significantly associated with TG response to fenofibrate. We also examined the common genetic component in TG and methylation responses to fenofibrate. The methylation response to fenofibrate was regarded as a methylation quantitative trait, and our methylation quantitative trait locus analysis identified a *cis*-acting regulation association with marginal significance between the homozygosity intensity of *MACROD2* and the methylation response to fenofibrate. These findings may help delineate the genetic basis of pharmacogenomic and pharmacoepigenomic responses to fenofibrate intervention.

Background

Homozygosity disequilibrium (HD), coined by Yang et al. [1], indicates a nonrandom pattern of sizable runs of homozygosity that deviates from a random allocation of homozygous and heterozygous genotypes in the genome. The major genetic mechanisms of HD include, but are not limited to, autozygosity [2], natural selection [3], and chromosomal aberrations [4]. HD is natural variation among individuals and has interethnic differences [5, 6]. HD has familial aggregation, suggesting a genetic component of HD. Genomic distribution of HD in humans has been characterized [5, 7, 8]. Genetic contributions of HD to the susceptibility of Mendelian diseases, complex disorders, and cancers have been reviewed [9]; HD is especially crucial for neurodevelopment-related diseases [10,

11] and autoimmune diseases [1, 12]. Gene regulation of disease-associated HD also has been observed [8].

Previously, we developed statistical method and software (Loss-of-Heterozygosity Analysis Suite [LOHAS]) to detect HD based on genotypes of single nucleotide polymorphisms (SNPs) in SNP microarrays [5]. LOHAS was applied to investigate genetic association between HD and disease susceptibility [1, 5, 13] and the relationship between HD and continental populations [5]. We also developed another method and software (AF/LOH/LCSH/AI/CNV/CNA Enterprise [ALICE]) to detect HD through a whole-genome SNP hybridization intensity analysis [14].

Because no studies had investigated HD by using whole-genome sequencing data, LOHAS was extended to analyze the whole-genome sequencing data set in GAW18 [7]. The extension of LOHAS was based on the assumption that all rare variants (RVs) have an equal weight, even though common homozygotes of RVs with a lower minor allele frequency carry less homozygosity

* Correspondence: hsinchou@stat.sinica.edu.tw
Institute of Statistical Science, Academia Sinica, No 128, Sec 2, Academia Rd,
Nankang 115, Taipei, Taiwan



information [7]. In GAW19, we further enhanced LOHAS by considering a local property and genetic information of homozygosity in the homozygosity intensity estimation. In contrast to the previous estimation procedure, the new method did not assume equal importance of RVs when defining HD. In addition to the higher computational efficiency, simulation studies suggested the new method has well-controlled type 1 error and higher power than our previous homozygosity association test [7]. The new method not only identified the regions of HD associated with blood pressure, but also discovered unreported evidence of gene regulation by the regions of HD associated with blood pressure.

No studies had investigated the role of HD in pharmacogenomics and pharmacoepigenomics. GAW20 provides a real data set from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) project [15]. The data set consists of treatment response, and whole-genome genotypes of SNPs and whole-genome methylation levels of cytosine-phosphate-guanine (CpG) sites, providing an unmet opportunity to study HD in pharmacogenomics and pharmacoepigenomics. This study aims to evaluate the effects of HD on the treatment responses to fenofibrate (i.e., triglyceride [TG] and DNA methylation changes resulting from fenofibrate intervention) and methylation regulation.

Methods

Materials

GAW20 provides clinical variables [TG (mg/dL) measurements before and after fenofibrate intervention] and covariates [sex, age, field center, smoking status, and two metabolic syndrome indices—Adult Treatment Panel (ATP) and International Diabetes Federation (IDF)] for 1105 individuals, derived from 188 independent pedigrees in the GOLDN project. Among the 1105 individuals, 1105 and 818 individuals had TG measurements before and after fenofibrate intervention, respectively. GAW20 provides a whole-genome genotype data set of 718,542 autosomal SNPs for 822 individuals, derived from 173 independent pedigrees. The SNP data were obtained using the Affymetrix 6.0 array. SNPs were removed if they violated Mendelian segregation, had a minor allele frequency of < 0.01 or a call rate of $< 96\%$, or failed the Hardy-Weinberg equilibrium test in $p < 1 \times 10^{-6}$. Details of quality control can refer to the GOLDN project [15]. GAW20 also provides whole-genome methylation pre- and post-fenofibrate treatment constituting of 463,995 CpG sites for 446 individuals from 140 independent pedigrees. The methylation data were obtained using the Illumina Infinium Human Methylation450 BeadChip. CpG sites were removed if they had insignificant detection p value, mismatch with annotation file, or were close to SNPs. Samples were removed if they had missing data of $> 1.5\%$ of CpG

sites and were outliers in principal component of T-cell purity. Combat normalization was used to correct for batch effect. For details of quality control, refer to the GOLDN project [15].

Statistical methods

We applied the double-weight local polynomial model [8] in LOHAS to estimate homozygosity intensities of 822 individuals. Sliding windows on each chromosome were constructed by using the nearest-neighbor method. The number of SNPs in each sliding window was 5% of SNPs on a chromosome. A cubic kernel weight and a locus weight with a threshold of minor allele frequency of 0.05 were considered. The estimates of homozygosity intensity range from 0 to 1. A higher value indicates a higher homozygosity.

In each sliding window, generalized estimation equation (GEE) analysis was performed to examine the relationship between homozygosity intensities and TG responses to fenofibrate, with concomitant adjustment for covariates (sex, age, field center, smoking status, and two metabolic syndrome indices—ATP and IDF) of 778 individuals. To consider potential population stratification, we performed principal component analysis based on a linkage disequilibrium pruned set of 80,930 SNPs at $r^2 < 0.2$. The top 10 principal components were included as covariates in the GEE analysis. In this study, TG response to fenofibrate was measured by a reduction of TG after fenofibrate intervention (i.e., the average of two TG measurements after fenofibrate intervention [visit 3 and visit 4] minus the average of two TG measurements before fenofibrate intervention [visit 1 and visit 2]). In general, fenofibrate intervention has a TG-lowering effect.

Bonferroni correction was performed for a multiple-testing problem in this GEE-based genome-wide homozygosity association study. Because of a high dependency of statistical tests across overlapping sliding windows, we estimated the effective number of independent tests (n_e) by adapting the method in Li et al. [16] and then adjusted for multiple tests as follows: Let n_j denote the number of the homozygosity association tests (i.e., the number of sliding windows) on autosome j . Let $I[A]$ denote an indicator taking a value of 1 if event A holds and 0 otherwise. By each autosome, we estimated the effective number of independent tests $n_{e,j}$ using $n_j - \sum_{i=1, \dots, n_j} \{I[\lambda_{i,j} > 1](\lambda_{i,j} - 1)\}$, where $\{\lambda_{i,j}, i = 1, \dots, n_j\}$ indicates the eigenvalues of the correlation matrix of homozygosity intensities of sliding windows on autosome j . Finally, we used $\alpha / \sum_j n_{e,j}$ as the critical significance level with $\alpha = 0.05$.

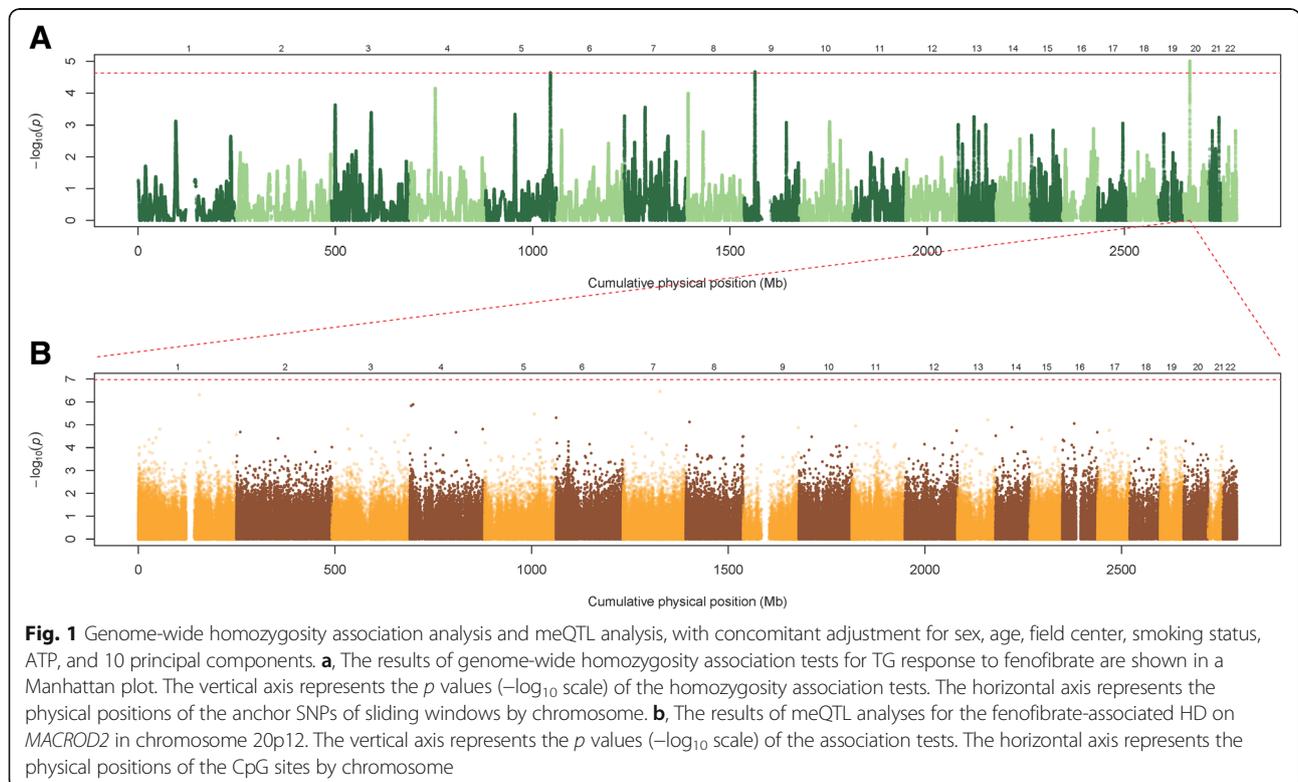
The changes in TG and DNA methylation patterns resulting from fenofibrate intervention were found, but no association was observed between the TG and methylation responses in previous studies [17]. We examined

whether the treatment responses in TG and DNA methylation were associated with the common genetic component. Once the regions of HD associated with the TG response to fenofibrate were identified, we further examined whether the identified regions of HD were also associated with the methylation response to fenofibrate. Similar to the homozygosity association study for the TG response to fenofibrate, GEE analysis was performed to examine the relationship between homozygosity intensities and methylation response to fenofibrate with concomitant adjustment for covariates (sex, age, field center, smoking status, the top 10 principal components, and two metabolic syndrome indices—ATP and IDF) of 429 individuals. The *cis*-acting and *trans*-acting methylation quantitative trait locus (meQTL) analyses were performed, where the methylation quantitative trait was the methylation response to fenofibrate. Before the meQTL analyses, a further methylation normalization was performed. Because a probe-type bias of Infinium I versus Infinium II was observed in the normalized data of GAW20, we performed the beta-mixture quantile normalization method [18] based on the methylation beta values (i.e., the ratio of methylated to combined intensity values) and then transformed the beta values to M values (i.e., \log_2 ratio of beta value to 1 – beta value). The methylation response to fenofibrate was calculated by a change of M values after fenofibrate intervention (i.e., M values at visit 4 minus M values at visit 2).

Results

We estimated the whole-genome profiles of homozygosity intensities of 822 individuals. We obtained all regions of HD satisfying homozygosity intensity of ≥ 0.9 and region length of ≥ 5 Mb in the genomes of all individuals. The minimum, first quartile, second quartile, third quartile, and maximum of lengths of regions of HD were 5.01 Mb, 8.82 Mb, 11.06 Mb, 12.90 Mb, and 40.91 Mb, respectively. We also calculated the total length of regions of HD in each genome. The minimum, first quartile, second quartile, third quartile, and maximum of total length of regions of HD carried by an individual were 5.15 Mb, 9.44 Mb, 12.46 Mb, 23.38 Mb, and 168.05 Mb, respectively.

We performed two genome-wide homozygosity association studies to identify fenofibrate response-associated regions of HD. In the first study, a GEE analysis with concomitant adjustment for covariates sex, age, field center, smoking status, ATP, and 10 principal components were performed based on 778 individuals. Figure 1a shows the Manhattan plot. For a multiple-testing correction, we estimated the effective number of independent tests was 2145. After a Bonferroni correction, this study identified three genomic regions strongly associated with a TG-lowering effect to fenofibrate intervention. The most significant sliding window at each of the three regions was anchored at rs254239 (chr5:164916163, $p = 2.308 \times 10^{-5}$), rs7037978 (chr9:27330668, $p = 2.15 \times 10^{-5}$), and rs17704829



(chr20:15691912, $p = 9.86 \times 10^{-6}$). In addition to a number of significant principal components, field center (Utah vs. Minnesota) was the only significant covariate in one of the three significant sliding window; $p = 0.0309$ in the significant sliding window anchor at rs17704829.

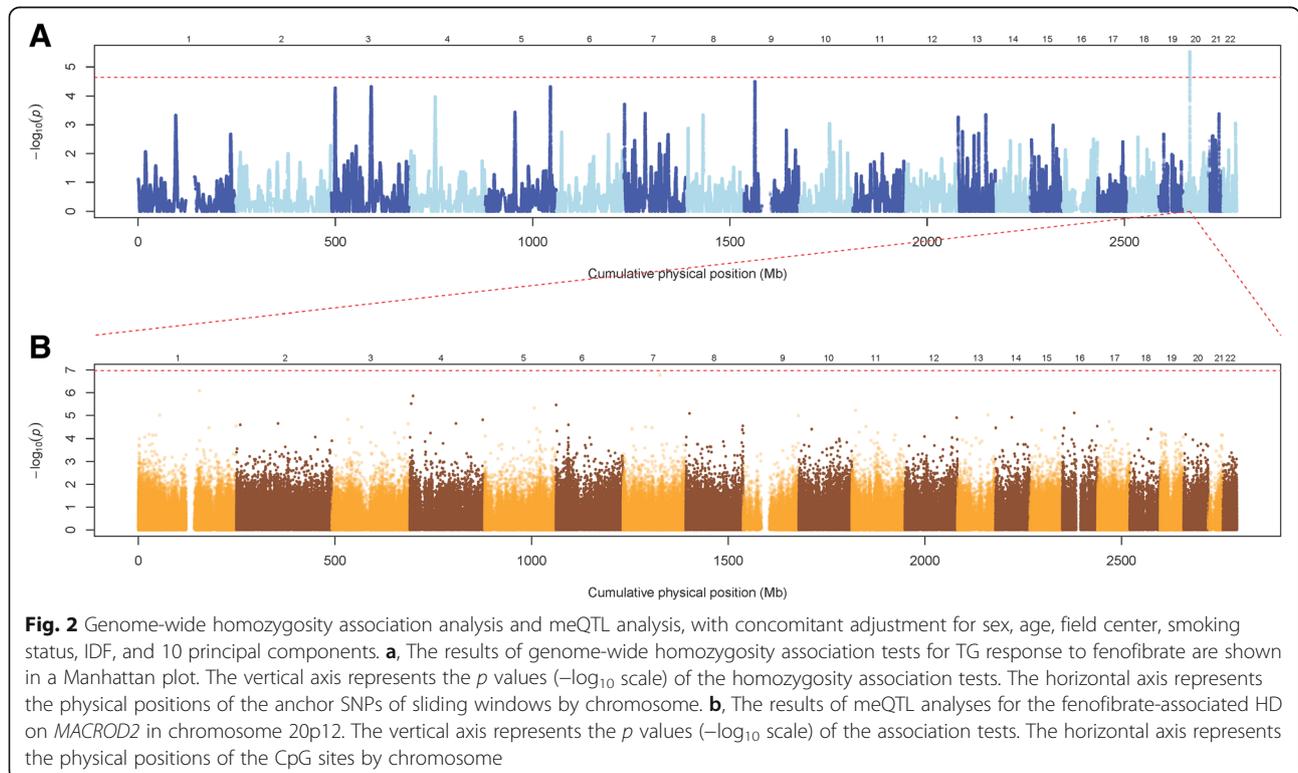
In the second study, the GEE model was similar to the previous model except for replacing ATP with IDF. The second study only identified one window as associated with a TG-lowering effect to fenofibrate intervention; the anchor SNP was rs764140 (chr20:15713167, $p = 2.95 \times 10^{-6}$) (Fig. 2a). Both of the studies identified *MACROD2* (MACRO domain containing 2) in chromosome 20p12 as an important gene region associated with treatment response to fenofibrate. In addition to a number of significant principal components, field center (Utah vs Minnesota) was the only significant covariate; $p = 0.0306$ in the significant sliding window anchor at rs764140.

To investigate the common genetic component in the TG and methylation responses to fenofibrate, we performed a *cis*-meQTL analysis for the two fenofibrate response-associated windows of HD anchor at rs17704829 and rs764140 on *MACROD2*. In total, 24 CpG sites on *MACROD2* were provided in the Illumina Infinium Human Methylation450 BeadChip. Homozygosity intensities of the two windows were marginally significantly associated with the methylation response to fenofibrate at two CpG sites: cg07953890 (chr20:13976782; $p = 2.318 \times 10^{-2}$

for the window anchor at rs17704829; $p = 2.555 \times 10^{-2}$ for the window anchor at rs764140) and cg09937190 (chr20:15177509; $p = 3.614 \times 10^{-2}$ for the window anchor at rs17704829; $p = 3.911 \times 10^{-2}$ for the window anchor at rs764140). To identify more meQTL associations, we also performed *trans*-meQTL analyses. The results are summarized in Fig. 1b and Fig. 2b. However, no *trans*-acting regulation association was identified after a multiple-testing correction.

Discussion and conclusions

We performed the first genome-wide study investigating the role of HD in pharmacogenomics and pharmacoepigenomics. The homozygosity intensity estimation method and homozygosity association tests in LOHAS software [5] (<http://www.stat.sinica.edu.tw/hsinchou/genetics/loh/LOHAS.htm>) we used are proven useful tools for studying HD in simulation studies and real data analyses [1, 5, 7, 8, 13]. In this study, we inferred whole-genome profiles of homozygosity intensities and examined the genomic distributions of HD. Our two genome-wide homozygosity association analyses pinpointed the same region of HD, contained in *MACROD2* at 20p12, strongly associated with TG response to fenofibrate. *MACROD2* was reported to be associated with metabolic diseases such as hypertension [19], as well as drug resistance such as tamoxifen for breast cancers



[20]. Our meQTL analysis identified a *cis*-acting regulation association with marginal significance between the homozygosity intensity of *MACROD2* and the methylation response to fenofibrate. These findings may help delineate the genetic basis of pharmacogenomic and pharmacoepigenomic responses to fenofibrate.

Acknowledgements

We are grateful to GAW20 for providing the whole-genome pharmacogenomics and pharmacoepigenomics data. We sincerely thank two anonymous reviewers for their constructive and insightful comments that helped in preparing our manuscript.

Funding

Publication of this article was supported by NIH R01 GM031575. This work was partially supported by grants from the Ministry of Science and Technology of Taiwan (MOST 103-2314-B-001-008-MY3).

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

HCY conceived of the study and prepared the manuscript. CWC analyzed the data and discussed the results with HCY. Both 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.

Published: 17 September 2018

References

1. Yang HC, Chang LC, Liang YJ, Lin CH, Wang PL. A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. *PLoS One*. 2012;7(4):e34840.
2. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*. 2006;15(5):789–95.
3. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie XH, Byrne EH, McCarroll SA, Gaudet R, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007;449(7164):913–8.
4. Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong LC, White RL. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature*. 1983;305(5937):779–84.
5. Yang HC, Chang LC, Huggins RM, Chen CH, Mullighan CG. LOHAS: loss-of-heterozygosity analysis suite. *Genet Epidemiol*. 2011;35(4):247–60.
6. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet*. 2012;91(2):275–92.
7. Yang HC, Li HW. Analysis of homozygosity disequilibrium using whole-genome sequencing data. *BMC Proc*. 2014;8:S15.
8. Yang HC, Lin YT. Homozygosity disequilibrium and its gene regulation. *BMC Proc*. 2016;10:S27.
9. Ku CS, Naidoo N, Teo SM, Pawitan Y. Regions of homozygosity and their impact on complex diseases and traits. *Hum Genet*. 2011;129(1):1–15.
10. Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, Launer L, Hardy J, Singleton AB. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics*. 2009;10(3):183–90.
11. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A*. 2007;104(50):19942–7.
12. Baschal EE, Aly TA, Jasinski JM, Steck AK, Noble JA, Erlich HA, Eisenbarth GS. Defining multiple common "completely" conserved major histocompatibility complex SNP haplotypes. *Clin Immunol*. 2009;132(2):203–14.
13. Huggins R, Li LH, Lin YC, Yu AL, Yang HC. Nonparametric estimation of LOH using Affymetrix SNP genotyping arrays for unpaired samples. *J Hum Genet*. 2008;53(11–12):983–90.
14. Huang MC, Chuang TP, Chen CH, Wu JY, Chen YT, Li LH, Yang HC. An integrated analysis tool for analyzing hybridization intensities and genotypes using new-generation population-optimized human arrays. *BMC Genomics*. 2016;17:266.
15. Irvin MR, Zhi DG, Aslibekyan S, Claas SA, Absher DM, Ordovas JM, Tiwari HK, Watkins S, Arnett DK. Genomics of post-prandial lipidomic phenotypes in the genetics of lipid lowering drugs and diet network (GOLDN) study. *PLoS One*. 2014;9(6):e99509.
16. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet*. 2012;131(5):747–56.
17. Das M, Irvin MR, Sha J, Aslibekyan S, Hidalgo B, Perry RT, Zhi DG, Tiwari HK, Absher D, Ordovas JM, et al. Lipid changes due to fenofibrate treatment are not associated with changes in DNA methylation patterns in the GOLDN study. *Front Genet*. 2015;6:304.
18. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29(2):189–96.
19. Slavin TP, Feng T, Schnell A, Zhu XF, Elston RC. Two-marker association tests yield new disease associations for coronary artery disease and hypertension. *Hum Genet*. 2011;130(6):725–33.
20. Mohseni M, Cidado J, Croessmann S, Cravero K, Cimino-Mathews A, Wong HY, Scharpf R, Zabransky DJ, Abukhdeir AM, Garay JP, et al. *MACROD2* overexpression mediates estrogen independent growth and tamoxifen resistance in breast cancers. *Proc Natl Acad Sci U S A*. 2014;111(49):17606–11.

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

