

RESEARCH

Open Access



# Modification effect of fenofibrate therapy, a longitudinal epigenomic-wide methylation study of triglycerides levels in the GOLDN study

Runmin Wei<sup>1</sup> and Yanyan Wu<sup>2\*</sup>

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

## Abstract

**Background:** Identification of interactions between epigenetic factors and treatments might lead to personalized intervention of diseases. This paper aims to examine the modification effect of fenofibrate therapy on the association of methylation levels and fasting blood triglycerides (TG), and the related biological pathways among methylation sites.

**Results:** Mixed-effects models were employed to assess pre- and posttreatment associations and drug modification effects simultaneously. Five cytosine-phosphate-guanine (CpG) sites were found to be associated with TG levels before and after the fenofibrate therapy: cg00574958, cg17058475, and cg01082498 on *CPT1A* gene, chromosome 11; cg03725309 on *SARS*, chromosome 1; and cg06500161 on *ABCG1*, chromosome 21. In addition, fenofibrate therapy modified the methylation levels on the following 4 CpG sites: cg20015535 (gene *EGLN1*, chromosome 1); cg24870738 (gene *RNF220*, chromosome 1); cg06891775 (gene *LOC283050*, chromosome 10); and cg00607630 (gene *USP7*, chromosome 16). Further, gene set enrichment analysis (GSEA) identified cancer- and metabolism-related pathways that were associated with TG-related CpG sites.

**Conclusions:** We identified modification effects of fenofibrate on the associations between blood TG levels and several CpG sites. Pathway enrichment analysis indicated the alternations in some metabolism and cancer-related pathways. Our findings have important implications for future research in pharmacoepigenetics and personalized medicine.

**Keywords:** Methylation, TG, Repeated measures, Mixed models, Interaction, GSEA

## Background

Large-scale genome-wide association studies (GWAS) have identified numerous loci associated with fasting blood lipids and other cardiovascular diseases (CVDs) [1]. Epigenetic analysis has gained attention in the past few years as an alternative perspective on the etiology of complex diseases. Epigenetic adaptations alter gene expressions and are heritable through many cell divisions, even across generations, while they do not alter the primary DNA sequence. To advance blood lipids and CVD

research, it is important to apply epigenome-wide association study (EWAS) to detect the epigenetic risk factors. The study of molecular mechanisms underlying epigenetic inheritance, such as DNA methylation, will provide insights in advancing and shaping ideas of the role that epigenetic phenomena play in high blood lipids and CVD. In this paper, we used data from the Genetics of Lipid-lowering Drugs and Diet Network (GOLDN) study provided by the GAW20 to examine the methylation levels of lipid-lowering treatment on the fasting blood triglycerides [2].

\* Correspondence: [yywu@hawaii.edu](mailto:yywu@hawaii.edu)

<sup>2</sup>Office of Public Health Studies, University of Hawai'i at Mānoa, 1960 East-West Rd, Biomed D-104F, Honolulu, HI 96822, USA

Full list of author information is available at the end of the article



## Methods

### Data

The GAW20 data sets are drawn from the GOLDN study with a total number of 1105 participants [2]. The data sets include GWAS and EWAS data before and after the fenofibrate (blood lipid-lowering drug) intervention. The EWAS data set contains 2 triglyceride (TG) measurements and methylation levels of 463,995 cytosine-phosphate-guanine (CpG) sites for 995 pretreatment individuals and 530 post-treatment individuals, respectively. The log-transformed mean pre- and posttreatment TG levels were used as the outcome variable in our model. Control variables include age, gender, study center, and family pedigree.

### EWAS model

We applied mixed-effects models for two repeated measures of log TG levels with fixed effects of time (0 = pre, 1 = post), methylation level, and their interactions, adjusting for age (18 years of age to approximately 87 years of age), sex, study site, and top 4 methylation principal components. Pedigree and subject IDs are controlled as nested random effects. These fixed effects of time, methylation levels, and the interaction term, measure the associations for both pre- and posttreatment periods, and the treatment modification effects, respectively.

Let  $Y_{ijk}$  denote the log TG measurements at  $k$ th time (0 = pre, 1 = post) for the  $i^{\text{th}}$  individual in the  $j^{\text{th}}$  pedigree;  $X_{ijk}$  denote the methylation level; and  $t_k$  denote treatment while  $t_0 = 0$  and  $t_1 = 1$ . The model equation can be written as:

$$Y_{ijk} = \beta_0 + \beta_1 X_{ijk} + \gamma t_k + \delta(X_{ijk} t_k) + \beta_2 \text{Age}_i + \beta_3 \text{Site}_i + \beta_4 \text{PC1}_{ik} + \beta_5 \text{PC2}_{ik} + \beta_6 \text{PC3}_{ik} + \beta_7 \text{PC4}_{ik} + S_{ij} + \varepsilon_{ijk}$$

where the main effect  $\beta_1$  is the pretreatment methylation effect on log TG;  $\gamma$  is the main treatment effect;  $\delta$  is the interaction effect between methylation and treatment (i.e., the treatment modification effect); and  $S_{ij}$  is the random effect of the individual nested within the pedigree. The general linear hypothesis tests were applied to calculate postmethylation effect ( $\beta_1 + \delta$ ), the standard errors, and the  $p$  values. We examined each CpG site on the whole genome (463,995 sites). Mixed-effects models for repeated measures enable us to examine the individual patterns of change by excluding between-individual variability and provide more efficient estimators of treatment effects. The main effects and interactions work together to identify the epigenetic risk factors of TG levels for pretreatment, posttreatment, and potential gene–drug interactions simultaneously [3, 4]. Compared to cross-sectional study, the repeated measure analysis has the advantage of making reliable inferences by capturing the systemic changes within individuals, thereby achieving more sensitive tests and higher statistical power for a fixed number of individuals [5, 6]. Statistical software

R (version 3.2.3) was used for the entire analyses, with R package *nlme* for mixed-effects modeling [7], *car* for linear hypothesis tests [8], and *qqman* for Manhattan plots [9]. We applied a relatively loose significance threshold ( $p$  value  $< 1E-5$ ) for modification effects and posttreatment associations because of the exploratory nature of proposed method and the moderate sample size ( $N = 536$  posttreatment measures). A less-stringent threshold might imply potential drug modification effects, as empirical evaluation suggests a possible relaxation in the current GWAS threshold for replication studies [10].

### Pathway-enrichment analysis

After EWAS analyses of CpG sites for pretreatment, posttreatment, and interaction effects, we mapped them to corresponding genes. To provide a functional insight of the results, we applied a gene set enrichment analysis (GSEA) [11] preranked test to each of 3 gene lists with log-transformed  $p$  values. To compute the empirical  $p$  values and false discovery rates (FDRs) for pathways, we performed 1000 permutations. Pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [12] were used in our analysis.

GSEA is a robust technique that searches for pathways (gene sets) that contain abundant highly significant genes (CpG sites) based on a Kolmogorov-Smirnov test [11] to reveal biological insights of genome/epigenome data.

## Results

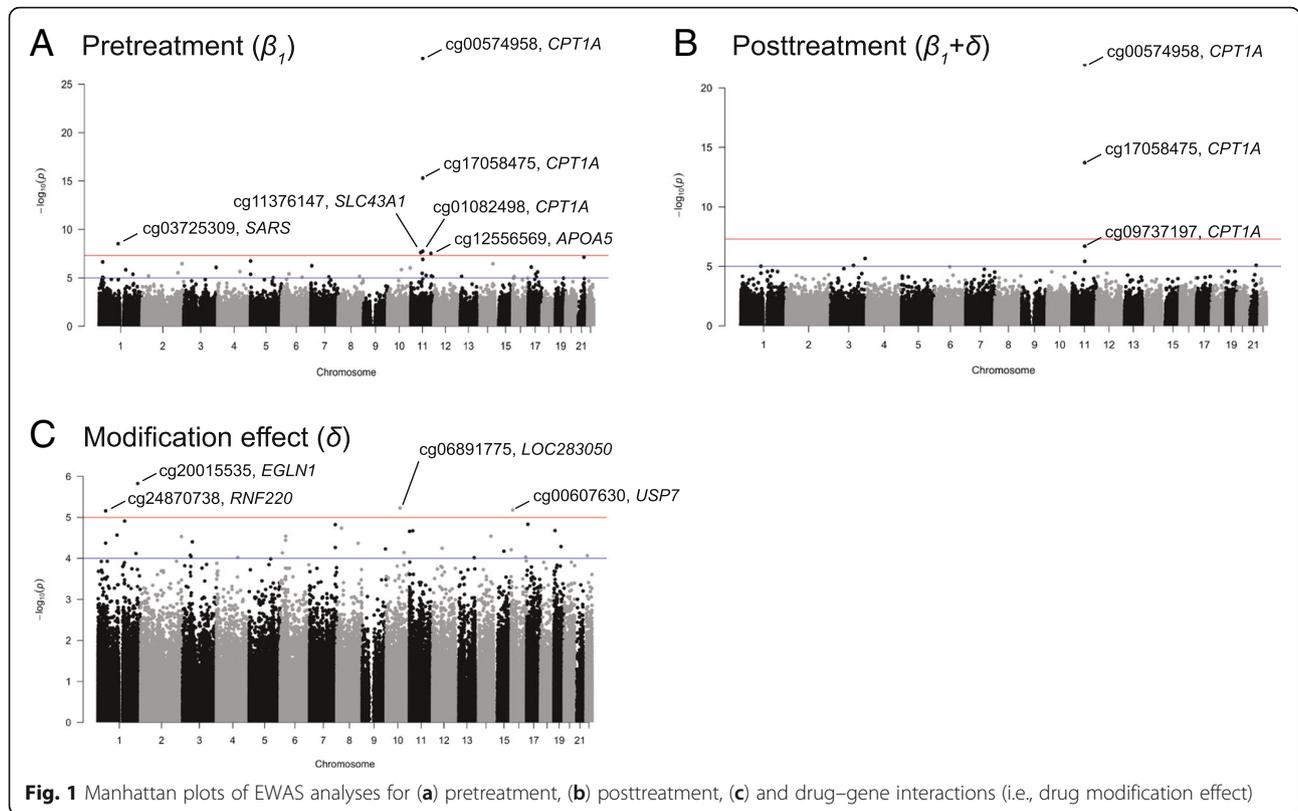
Table 1 lists selected CpG sites that are associated with pre- and post-log TGs, and modified by treatment, and Fig. 1 shows the corresponding Manhattan plots. The methylation level of 3 CpG sites (cg00574958, cg17058475, and cg01082498) on *CPT1A* gene, chromosome 11, and 2 other CpG sites (cg03725309 on gene *SARS*, chromosome 1, and cg06500161 on *ABCG1*, chromosome 21) were found to be associated with both pre- and posttreatment TG levels ( $p$  values  $< 1E-5$ ). Moreover, the methylation levels of 2 sites on chromosome 11 are associated with pretreatment log TG but not with posttreatment log TG levels (cg12556569 on gene *APOA5* and cg11376147 on gene *SLC43A1*).

In addition, the associations between methylation level and log TG levels were modified by the fenofibrate therapy on the following four CpG sites with  $p$  value  $< 1E-5$ : cg20015535 (gene *EGLN1*, chromosome 1), cg24870738 (gene *RNF220*, chromosome 1), cg06891775 (gene *LOC283050*, chromosome 10), and cg00607630 (gene *USP7*, chromosome 16).

The GSEA results are recorded in Table 2, which shows FDR  $q$ -values for the top 15 pathways across pre- and posttreatment associations, and the treatment-modifying effects. Several cancer-related pathways (KEGG\_ENDOMETRIAL\_CANCER; KEGG\_PATHWAYS\_IN\_CANCER; KEGG\_CHRONIC\_MYELOID\_LEUKEMIA; KEGG\_BASAL\_C

**Table 1** Selected EWAS CpG sites for pretreatment, posttreatment, and gene-drug interactions ( $p$  value  $< 1E-5$  is considered statistical significance and italicized)

CpG site	Chr	Location	Genes	Pretreatment $\beta_1$ (95% CI)	Posttreatment $\beta_1 + \delta$ (95% CI)	Modification effect (95% CI)	$p$ value (Pre-)	$p$ value (Post-)	$p$ value (Modif.)
<i>Significant effect modification sites</i>									
cg20015535	1	231,556,123	<i>EGLN1</i>	1.30 (0.11, 2.49)	-3.1 (-4.77, -1.43)	-44 (-6.17, -2.63)	3.31E-02	2.80E-04	1.49E-06
cg24870738	1	45,097,499	<i>RNF220</i>	-3.8 (-5.9, -1.70)	2.23 (0.58, 3.88)	603 (3.44, 8.62)	4.40E-04	7.81E-03	6.90E-06
cg06891775	10	80,704,982	<i>LOC283050</i>	2.13 (1.28, 2.98)	0.52 (-0.29, 1.33)	-1.61 (-2.3, -0.92)	1.39E-06	2.10E-01	5.91E-06
cg00607630	16	9,030,093	<i>USP7</i>	2.82 (1.61, 4.03)	0.04 (-1.15, 1.23)	-2.78 (-3.97, -1.59)	7.18E-06	9.53E-01	6.60E-06
<i>Significant pre or post methylation sites</i>									
cg03725309	1	109,757,585	<i>SARS</i>	-1.65 (-2.18, -1.12)	-1.28 (-1.85, -0.71)	0.37 (0.04, 0.7)	3.00E-09	9.82E-06	2.50E-02
cg12556569	11	116,664,039	<i>APOA5</i>	0.35 (0.23, 0.47)	0.23 (0.08, 0.38)	-0.12 (-0.24, 0)	3.11E-08	2.34E-03	4.56E-02
cg11376147	11	57,261,198	<i>SLC43A1</i>	-2.24 (-3.01, -1.47)	-1.46 (-2.37, -0.55)	0.78 (0, 1.56)	2.46E-08	1.60E-03	5.20E-02
cg00574958	11	68,607,622	<i>CPT1A</i>	-3.81 (-4.44, -3.18)	-3.55 (-4.26, -2.84)	0.26 (-0.31, 0.83)	2.12E-28	1.07E-22	3.65E-01
cg01082498	11	68,608,225	<i>CPT1A</i>	-2.59 (-3.48, -1.7)	-2.38 (-3.39, -1.37)	0.21 (-0.66, 1.08)	1.86E-08	3.83E-06	6.30E-01
cg17058475	11	68,607,737	<i>CPT1A</i>	-2.31 (-2.84, -1.78)	-2.34 (-2.94, -1.74)	-0.04 (-0.45, 0.37)	5.00E-16	1.91E-14	8.63E-01
cg06500161	21	43,656,587	<i>ABCG1</i>	2.26 (1.45, 3.07)	2.16 (1.21, 3.11)	-0.09 (-0.86, 0.68)	7.55E-08	8.18E-06	8.16E-01



ELL\_CARCINOMA; and KEGG\_NON\_SMALL\_CELL\_LUNG\_CANCER) show consistent significant results. In addition, metabolism related pathways were also observed (KEGG\_TYPE\_II\_DIABETES\_MELLITUS and KEGG\_ADIPOCYTOKINE\_SIGNALING\_PATHWAY).

**Discussion**

The proposed mixed-effects model examines the methylation sites that are associated with blood TG levels before and after the fenofibrate therapy, and the potential gene–drug interactions. Using the linear hypothesis test,

**Table 2** FDR *q*-values for the top 15 pathways across different results from GSEA

Pathway name	Pretreatment FDR <i>q</i> value	Posttreatment FDR <i>q</i> value	Modification FDR <i>q</i> value
KEGG_ENDOMETRIAL_CANCER	0.146	0.023	0.040
KEGG_PATHWAYS_IN_CANCER	0.166	0.059	0.137
KEGG_DORSO_VENTRAL_AXIS_FORMATION	0.182	0.067	0.122
KEGG_ADHERENS_JUNCTION	0.141	0.079	0.139
KEGG_TYPE_II_DIABETES_MELLITUS	0.198	0.079	0.092
KEGG_COLORECTAL_CANCER	0.205	0.078	0.107
KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY	0.160	0.081	0.138
KEGG_FOCAL_ADHESION	0.165	0.061	0.153
KEGG_MELANOGENESIS	0.178	0.077	0.139
KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	0.193	0.017	0.156
KEGG_CHRONIC_MYELOID_LEUKEMIA	0.160	0.134	0.133
KEGG_BASAL_CELL_CARCINOMA	0.153	0.055	0.199
KEGG_CALCIIUM_SIGNALING_PATHWAY	0.236	0.067	0.133
KEGG_WNT_SIGNALING_PATHWAY	0.230	0.075	0.133
KEGG_NON_SMALL_CELL_LUNG_CANCER	0.289	0.070	0.109

we identified 7 CpG sites that are associated with pretreatment TG levels ( $p$  value  $<1E-7$ ) and 5 sites for post-treatment ( $p$  value  $<1E-5$ ). All 5 posttreatment CpG sites are overlapped with pretreatment CpG sites. Among these CpG sites, 3 are located in gene *CPT1A*, which encodes a key enzyme in the carnitine-dependent transport of long-chain fatty acids across the mitochondria membrane whose deficiency will result in downregulation of fatty acid  $\beta$ -oxidation [13]. The consistent findings suggest a strong association between blood TGs and DNA methylation of *CPT1A* regardless of the interference of lipid-lowering drug. In addition, we also observed 4 potential drug-interacted CpG sites from our results that belong to genes *EGLN1*, *LOC283050*, *USP7*, and *RNF220*. Previous study shows that the inhibition of *EGLN1* improves the glucose and lipid metabolism, and protects against obesity and metabolic dysfunction [14]. Less-significant results were observed for drug modification effects, which were in part a result of the moderate sample size (536 posttreatment measures). But our results provide initial evidence of gene–drug interaction and warrant replication studies [10].

To provide further biological insight to 4 EWAS results, GSEA was applied to examine the associated biological pathways using KEGG database [12]. It is worth noting that 5 cancer-related pathways were enriched by TG-associated CpG sites. We observed potential associations between blood TG levels and cancer risk from an epigenetic point of view. Although obesity was recognized as a risk factor for several different cancers, for example, endometrial cancer [15], further mechanism research is necessary to determine whether there is any association between methylation level and cancer risk. In addition to these cancer-related pathways, 2 metabolism-related pathways were also observed. For Type II diabetes, elevated TG levels are common dyslipidemic features [16] and could be identified as an independent risk factor [17]. Another significant pathway is the KEGG\_ADIPOCYTOKINE\_SIGNALING\_PATHWAY. In addition to the fatty acid metabolism and  $\beta$ -oxidation, this pathway is also associated with glucose uptake and insulin resistance.

## Conclusions

In summary, we used linear mixed models with interaction terms to study pre- and posttreatment associations between blood TGs and CpG methylation levels and drug–gene interactions simultaneously across the whole-genome. We found several CpG sites that were consistently associated with blood TG levels in both pre- and posttreatment. In addition, by testing on the interaction term, we found potential treatment modification effects on certain CpG sites. Our pathway-enrichment analysis revealed a number of cancer-related biological pathways that were significantly enriched by TG-associated CpG sites. The results suggest

connections between TG levels and cancer risk through an epigenetic point of view. However, because only 1 cohort with a limited sample size was studied in our analyses, further research on independent cohorts and experimental biology validations are needed for convincing conclusions.

## Abbreviations

CpG: Cytosine-phosphate-guanine; CVDs: Cardiovascular diseases; EWAS: Epigenome-wide association study; FDRs: False discovery rates; GAW: Genetic Analysis Workshop; GOLDN: Genetics of Lipid-lowering Drugs and Diet Network; GSEA: Gene set enrichment analysis; GWAS: Genome-wide association studies; KEGG: Kyoto Encyclopedia of Genes and Genomes; TG: Triglycerides

## 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 Genetics* Volume 19 Supplement 1, 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://bmccgenet.biomedcentral.com/articles/supplements/volume-19-supplement-1>.

## Authors' contributions

RW, YYW conceived the study, conducted the data analysis and wrote the paper. Both 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.

## Author details

<sup>1</sup>Molecular Biosciences and Bioengineering, University of Hawai'i at Mānoa, 1955 East-West Rd, Honolulu, HI 96822, USA. <sup>2</sup>Office of Public Health Studies, University of Hawai'i at Mānoa, 1960 East-West Rd, Biomed D-104F, Honolulu, HI 96822, USA.

Published: 17 September 2018

## References

1. Global Lipids Genetics C. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45:1274–83.
2. Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, et al. Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study. *Circulation.* 2014;130:565–72.
3. Fox J. Generalized linear models. *Appl Regres Anal Gen Linear Model.* 2008; 135:379–424.
4. Wu YY, Briollais L. Mixed-effects models for joint modeling of sequence data in longitudinal studies. *BMC Proc.* 2014;8:S92.
5. Hedeker DR, Gibbons RD. Longitudinal Data Analysis. Wiley Ser Probab Stat 2006;1–8. <https://www.wiley.com/en-us/Longitudinal+Data+Analysis-p-9780471420279>.

6. Fitzmaurice GM, Nan M, Laird JHW. Applied longitudinal analysis. 2nd ed: Wiley; 2011.
7. Pinheiro J, Bates D, DebRoy S, Sarkar D. Nlme: linear and nonlinear mixed effects models. R Packag version 3. 2007.
8. Fox J, Weisberg S. An {R} companion to applied regression, second edition. Sage. 2011;2016. <https://www.barnesandnoble.com/w/an-r-companion-to-applied-regression-john-fox/1119455638>.
9. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014. Available from: <https://cran.r-project.org/web/packages/qqman/README.html>
10. Panagiotou OA, Ioannidis JPA, Hirschhorn JN, Abecasis GR, Frayling TM, McCarthy MI, et al. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol*. 2012;41:273–86.
11. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, M a G, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102: 15545–50.
12. Kanehisa M, Goto S. Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
13. Longo N, Amat Di San Filippo C, Pasquali M. Disorders of carnitine transport and the carnitine cycle. *Am J Med Genet Semin Med Genet*. 2006;142C:77–85.
14. Rahtu-Korpela L, Maatta J, Dimova EY, Horkko S, Gylling H, Walkinshaw G, et al. Hypoxia-inducible factor prolyl 4-Hydroxylase-2 inhibition protects against development of atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2016; 36:608–17.
15. Sorosky JL. Endometrial cancer. *Obstet Gynecol*. 2012;120:383–97.
16. Ginsberg HN, Zhang Y-LL, Hernandez-Ono A. Regulation of plasma triglycerides in insulin resistance and diabetes. *Arch Med Res*. 2005;36:232–40.
17. Tirosh A, Rudich A, Shochat T, Tekes-Manova D, Israeli E, Henkin Y, et al. Changes in triglyceride levels and risk for coronary heart disease in young men. *Ann Intern Med*. 2007;147:377–85.

**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)

