Author’s response to reviews

Title: Circadian pathway genetic variation and cancer risk: evidence from genome wide association studies

Authors:

Simone Mocellin (simone.mocellin@unipd.it)
Saveria Tropea (saveria.tropea@iov.veneto.it)
Clara Benna (clara.benna@unipd.it)
Carlo Rossi (carlor.rossi@unipd.it)

Version: 1 Date: 21 Dec 2017

Author’s response to reviews:

Dear Editors,

We are submitting you the revised version of our manuscript entitled “Circadian pathway genetic variation and cancer risk: evidence from genome wide association studies”, which we hope you will find suitable for publication in BMC Medicine.

We thank the reviewers for their comments/criticisms, which we believe have helped us improve the quality of our paper.

We have addressed all the issues raised by the reviewers (see below point-by-point reply) and have changed the text accordingly.

All changes are highlighted in green for prompt identification.

We thank you for taking into consideration our work and look forward to hearing from you.

Best regards

Simone Mocellin

University of Padova, Italy
Point-by-point reply:

Reviewer #2:

This paper reports circadian pathway genetic variation in breast, prostate and lung cancer predisposition based on in silico analysis of public available three GWAS meta-analyses and using an adaptive rank truncated product (ARTP) method. It should be emphasized that this is the first study investigating a possible role of clock genes variations in cancer using SNPs data repository from GWAS studies (the NCBI database of Genotypes and Phenotypes (GaP)) and ARTP approach.

Novel set-based methods, including ARTP which analyze sets of SNPs jointly, can detect variants with a potentially smaller effects acting within a gene or a pathway. Thus, a challenge for the proposed circadian gene loci-based pathway study was to analyze association of a set of SNPs in 1) a gene or 2) a pathway using ARTP.

The study is of great interest, because of its novelty and very interesting findings regarding the importance circadian variants in cancer risk, but it definitely needs several corrections.

Major remarks:

* It should be more clearly stated that association of set of circadian SNPs was analyzed in a pathway-based and a gene-based ARTP methods.

REPLY: Done as requested (Abstract, page 4; Materials and Methods, page 7 and page 8; Results section, page 12; Discussion section, page 15; Table 1 and Table 2).

* Additionally, circadian pathway definition with particular clock gene and number of SNPs within should be included (in the text and/or additional table).

REPLY: Done as requested (besides the already existing definition on pages 7 and 8 of Materials and Methods section, we have now added some text in the Results section on page 12 and have added Supplementary Table 1).

* I also suggest to include pathway P-values for each type of cancer and cancer subgrups in the table 1 and 2.

REPLY: The pathway P-values are already in the text and we believe it would be redundant and confusing for the reader to find the same information in the tables. We believe it is preferable to
use tables for providing information not already present in the text (and vice versa), as it is usually advised by editorial guidelines.

* Results (tables). I suggest to present set of SNPs or single SNPs (rs number) as a supplementary file.

REPLY: Done as requested (Supplementary Table 1, Supplementary Table 2, Supplementary Table 3).

* Study design requires more precisely description.

REPLY: Done as requested (Materials and Methods section, page 7).

* Clock gene selection. In my opinion the selection of clock genes seems to be incomplete. What about two repressors of E-box driven transcription BHLHE40 (DEC1), BHLHE 41 (DEC2); four D-box binding transcription factors; FBXL3 and TIPIN?

REPLY: following the reviewer’s suggestion we searched for information on these genes: unfortunately, the data sources we used did not include SNP information on none of these genes. Please consider we did not find information either on two genes we had selected in the first place (CSNK1D and TIMELESS).

* SNPs selection. How many tagSNPs/representative clock SNPs/a priori selected candidate SNPs have been chosen, globally (pathway ARTP) and in specific circadian gene (gene loci-based ARTP)?

REPLY: We thank the reviewer for this comment, because it shows the text was unclear. Briefly, there was no SNP selection based on tag or candidate strategy: we used all the SNPs representing variations of clock genes that were reported in already published GWAS (or their meta-analysis). In other words, we first selected clock genes, then we searched GWAS data for germline variants of these genes and finally we performed ARTP-based gene and pathway analysis. We have now tried to make this concept clearer in the text (Materials and Methods section, page 7).
* NCBI GaPdb data repository. dbGaP study accession number in the repository should be presented.

REPLY: Done as requested (Results section, page 11 and page 12)

Minor remark:
* Appropriate references for STREGA and STROBE should be included, as well as website link of each mentioned database.

REPLY: Done as requested (Materials and Methods section, page 7)

Reviewer #3:

In this article by Mocellin et al., the authors use publically available data from GWA studies to investigate the association between circadian rhythm related genes and risk of three cancers, breast, prostate and lung. The authors utilize a pathway analysis approach to evaluate the association and found variation in 15 genes were associated with risk of cancer, and four were associated with all three cancers. The paper is well-written and described, but I do have some comments.

Introduction:

1. The first paragraph is repetitive with the information included in the methods describing the pathway approach. Would suggest focusing and starting the introduction more on why the circadian rhythm genes in particular are interesting for risk of these three cancers. The justification for choosing these cancers was in my opinion lacking.

REPLY: We thank the reviewer for this comment: we have now changed the text as suggested (Introduction, page 5 and 6)

2. Figure 1 was confusing in the sentence it was ascribed to. Would suggest explaining the figure a bit more in the text. Perhaps mention the negative and positive feedback loops. You describe it in more detail in the methods - either move to introduction, or wait to introduce the figure until methods.

REPLY: Done as requested (we followed the last suggestion, see Materials and Methods section, page 8).
Methods:

1. Was there a reason data had to be from case-control studies? Could you have utilized cohort study data?

REPLY: Actually we chose to collect data from genome-wide association studies (GWAS): we did not choose case-control studies, although all GWAS we found were actually case-control studies. We changed the text to make it clearer (Materials and Methods section, page 8).

2. On page 8, it says the method adjusts the p-value for multiple testing. This was in parentheses and I missed it the first time. Would suggest taking out of parentheses and describing how it adjusts.

REPLY: Done as suggested (Methods section, page 9)

3. The first paragraph on page 9 is confusing and I don't know what it adds to describing the method.

REPLY: We believe this paragraph is important for people who are not expert in this field: reviewer #2 has asked for explanations just on this paragraph, so we believe it would be important to keep it.

4. It is stated that there are 3 values reported: the pathway p-value, the gene p-value and the top gene and SNP. Is the pathway p-value all of the genes you evaluated, or restricted to the top genes?

REPLY: Also according to one of the subsequent reviewer’s criticism, we have now rearranged the results section (see below reply).

5. The first part of the results describing the study populations belongs in the methods section under study population.

REPLY: As we looked for GWAS reporting on germline variants investigated for association with cancer risk and the entire analysis hinges upon these GWAS findings, we believe that the description of the GWAS we identified are the first results of our work.
Results:

1. The results are presented somewhat confusingly. I read through the results multiple times before figuring out where the pathway p-value was and what the top SNP was. I would suggest re-arranging the tables to include this information. Also, would be helpful to put all the results in supplementary information so that one can compare the results for each gene across all the cancer types.

REPLY: As suggested by the reviewer, we have now rearranged the results section. In particular:

A) ARTP-based pathway P-values are reported as the first result for each cancer, and it is clearly stated what they are (Results section, pages 12-14);

B) Please consider that top gene and top SNP are reported in the text of the Results section (not in tables); we believe it is preferable to use tables for providing information not already present in the text (and vice versa), as it is usually advised by editorial guidelines.

C) ARTP-based gene analysis findings are now available in Supplementary Table 2 (primary analysis) and Supplementary Table 3 (subgroup analysis), which parallels the description of the results in the main text (Results section, pages 12-14).

Discussion:

In general, I thought the discussion re-hashed the results rather than including a full discussion of the ‘why’. For example, a more detailed discussion of the four genes that came up in all three cancers would add to the paper. Breast and prostate are hormonally driven cancers, lung different, why would they have similar genes and/or why not.

REPLY: Done as suggested (Discussion section, page 16)

General minor comments:

1. The manuscript is well-written, but there are some instances where the meaning is a bit confusing. For example, on page 9: "this gene-based pathway analysis implies to obtain …." Do you mean, the gene-based pathway analysis obtains?

REPLY: following the reviewer’s suggestion, the English language/style has been revised throughout the text (Introduction section, page 6; Materials and Methods section, page 8 and page 9; Discussion section page 15)