**Reviewer’s report**

**Title:** Systematic characterization of germline variants from the DiscovEHR study Endometrial Carcinoma population

**Version:** 0 **Date:** 10 Dec 2018

**Reviewer:** Mari Kylleso Halle

**Reviewer's report:**

General comment:

The authors present a study identifying potential pathogenic germline variants by comparing whole exome sequencing data from a cohort of patients with known endometrial cancer diagnosis (EMCA, N=300), an elderly female non-cancer cohort (NCC, N=2120) and a separate hormone related malignancy cancer cohort (OHRM, N=1463). The purpose of this study is to detect rare genetic variants within the EMCA cohort with the assumption that these rare variants were potential pathogens for endometrial cancer development. The authors found four germline variants that were unique in the EMCA cohort when looking at overlap of loci compared to nine detected genes when utilizing a burden-based approach. The plausibility in which these variants could be pathogenic for endometrial cancer development and whether these methods are suitable for such investigations are further discussed. The study design is OK, however the manuscript should be thoroughly revised and spell checked. Several major and minor comments are outlined below.

**Major comments:**

Comment #1: A major limitation with this study is that a clear clinical application of these findings do not exist. Abnormal vaginal bleeding is an early symptom of endometrial cancer present in more than 90% of the patients. Postmenopausal bleeding is a disconcerting symptom, urging the majority of patient to seek medical care. The diagnosis is therefore often set at an early stage. On this background, I do not see a potential clinical application of distinguishing non-cancer from endometrial cancer. The challenge with endometrial cancer is, however, to distinguish the diagnosed patients into risk/treatment groups, and in my opinion, the suggested pathogenic drivers in this study is not well suited to classify patients into such risk groups.
However, the suggested genes and variants could point to causal links to the pathogenesis of cancer, and on this conclusion, I agree with the authors. Please elaborate more on the clinical utility of these findings.

Abstract:

Comment #2: Why is there a need to identify novel germline variants among EMCA patients? As mentioned above, patients with endometrial cancer is in general diagnosed at an early stage of disease. Please described more detailed what you see as the potential clinical application of these findings!

Comment #3: When I read the abstract alone, I get the impression that the purpose of this study was to detect endometrial cancer patients from non-cancer controls. However, there is really not a striking clinical need to use germline variants to do this. Already existing methods such as Pipelle/curettage sampling combined with histopathological evaluation and imaging analysis will more accurately identify patients with endometrial cancer. As I read through the manuscript I realize that early detecting probably is not what the authors sees as the main application of these finding. However, the abstract should be re-phrased to give a more precise description of the purpose of this study.

Methods:

Comment # 4: Patient cohort: The EMCA patient cohort should be further described. This population is clearly not population based, but consists as I can assume of American patients within the GHS cohort. As far as I can see, there is a prevalence of low stage and grade tumors and obese patients. In order to characterize the patient cohort and to assess the transferability to other cancer cohorts, a crosstab should be made with the same clinicopathological parameters from the whole American (or worldwide) population.

Comment # 5: Patient cohort: Leiomyosarcomas are usually not included in the same analysis as the rest of the endometrial carcinomas. These should be removed from the analysis.
Comment #6: In Table S1, 267 patients are registered with grade. Is this only the endometrioid patients? If so, please inform. And if so, why is there only 265 patients with endometrioid histology? This does not add up.

Comment #7: Table S1: Please emphasize/highlight what are headings and subheadings. Additionally, there is no use of mentioning that twice that there are 265 patients with endometrioid histology.

Results:

Comment #8: Please insert a crosstab with statistics where at least age, BMI and family history of EMCA or colon cancer and is compared among the cohorts. Both age and BMI are important factors for development of endometrial cancer.

Comment #9: DiscovEHR germline variants are reproduced in TCGA study: The TCGA study cohort consists of patients with more aggressive disease. Additionally, only endometrioid and serous carcinomas are included in the TCGA study. Please insert a crosstab that quantifies these differences and elaborate a bit on the impact this plays on the analyses.

Discussion:

Comment #10: Page 9, Line 55: Please put in more references on important biomarker discovery in the field of endometrial cancer. The TCGA study is far from the only study that has investigated somatic alterations to identify clinically relevant biomarkers.

Comment #11: Page 9, Line 58: ‘(..) it is equally important to study germline variants as they serve as potential prognostic indicator for the disease.’ I disagree: It is not equally important to detect markers for early detection of endometrial cancer. The current clinical challenge is not early detection, but rather risk stratification. Risk and further treatment stratification depend on
characterization of somatic alterations within the tumor that could serve as prognostic or predictive biomarkers. Please rephrase.

Comment #12: Page 12, Line 23: Please also comment on differences in BMI and age between these cohorts.

Approvals/declarations:

Comment #13: Is there a local IRB approval for this particular project? WES data are considered sensitive. How is the anonymity of patients ensured?

Minor comments:

Background:

Comment #1: Page 3, Line 18: Please insert hormone after steroid: 'Steroid hormone receptor' or just 'hormone receptors' are more commonly used to name these receptors.

Comment #2: Page 3, Line 30-34: Please state which nation's treatment guidelines you are referring to? Are these the US treatment guidelines?

Comment #3: Page 3, Line 43: Please insert 'also' in the sentence (e.g. but an increased risk 'also' of EMCA for women.)

Comment #4: Page 4, Line 6: Please put in references. Germline variant analyses have been performed for endometrial cancer!
Methods:

Comment #5: Page 5, Line 12: 'WES was performed' can be misunderstood. As I understand, WES data was collected from the DiscovEHR database[10]. Please rephrase!

Comment #6: Please define 'genomic boundaries'.

Results:

Comment #7: Page 6, Line 35: Please write the abbreviation 'MAF' in full the first time it is mentioned!

Comment #8: Page 7, Line 7: Please insert a comma after study (split up sentences to clarify the message).

Comment #9: Page 9, Line 38: Please remove 'rare variants'. It has been written twice!

Discussion:

Comment #10: Page 11, Line 32: Please remove 'which' in the beginning of the sentence and write 'are parts of' instead of are a part of. This appears to be typos. The whole paragraph should be revised for typos!

Conclusions:

Comment #11: Page 13, Line 4: Please remove 'And' in the beginning of the sentence and please insert 'that' after illustrates.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.