Reviewer's report

Title: Molecular subtypes of urothelial carcinoma are defined by specific gene regulatory systems

Version: 1 Date: 27 March 2015

Reviewer: Sherri Millis

Reviewer's report:

The data analysis and new information are comprehensive and important to the basic research in urothelial cancers. Extensive analysis has been performed; however, an extensive re-write is recommended to more clearly define the goals and the conclusions. A more grammatically correct and concise version of the manuscript, especially the discussion, would significantly increase the clarity of the significance and conclusions. Specifically:

Minor essential revisions

1) In the methods section a) provide a rationale for the cohorts chosen, b) note that TCGA data are specific to primary tumors, c) clarify the other datasets' derivations and limitations (ie the Lund 308 – are these primary, metastatic, ?), d) explain your specific exclusions.

2) Results section:
   a. Section “Stat3 regulated gene signature…” states “we could establish…” and references a figure but provides no verbiage documenting how this can be established. Supporting explanation is needed.
   b. Conclusions are often placed in the results section. Move to discussion/conclusions.

3) Discussion section:
   a. Authors note that ‘RAR expression did not vary across the data’ but do not explain in context to subtype differences or comment regarding potential downstream differences. Additionally, ‘across the data’ is vague. Be specific to what is being compared, please.
   b. Statement ‘from a clinical perspective FOXM1 may be involved in, and create, the known aggressiveness’ is also vague and may be speculative, as the comment is not supported, nor is the statement further documented.
   c. Perhaps break the discussion into biomarker-focused and subtype focused sections for more clear focus of the key points. Currently, this section is very hard to follow.
   d. Last sentence of discussion is not a section concluding statement.

4) Conclusion section:
   a. Incomplete.
b. Provide context for what these different subtypes mean in terms of treatment/biomarker ‘targetability’.
c. No limitations noted.
d. No mention of the comparison to breast cancer data.
e. No clear wrap-up.

5) Comparison to breast cancer. Mentioned in methods, noted again in discussion, but not in conclusions. Please provide more context regarding why only breast cancer, basal subtypes, is compared to your analysis, including any supporting data that would indicate this is the only other type of cancer with similarities. Otherwise, analysis is incomplete.

6) Grammatical errors and vague sentences, unclearly defined sections, missing words, etc. decrease the ability to fully analyze the novelty, importance of the work. Example errors include incomplete sentences, non-parallel sentence structure, and incorrect word usage.

While the overall content may provide new information in the field, the manuscript needs to be rewritten to increase the clarity of the content, and the methods must be further defined.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Not suitable for publication unless extensively edited

Statistical review: No, the manuscript does not need to be seen by a statistician.