Author's response to reviews

Title: Combined comparative genomic hybridization and transcriptomic analyses of ovarian granulosa cell tumors point to novel candidate driver genes.

Authors:

S. Caburet (caburet.sandrine@ijm.univ-paris-diderot.fr)
M. Anttonen (mikko.anttonen@helsinki.fi)
A. L. Todeschini (todeschini@ijm.univ-paris-diderot.fr)
L. Unkila-Kallio (leila.unkila-kallio@hus.fi)
D. Mestivier (denis.mestivier@gmail.com)
R. Butzow (Ralf.Butzow@hus.fi)
R. Veitia (veitia.reiner@ijm.univ-paris-diderot.fr)

Version: 4 Date: 12 February 2015

Author's response to reviews: see over
Dear Editor,
Please find enclosed the revised version of our manuscript "Combined comparative genomic hybridization and transcriptomic analyses of ovarian granulosa cell tumors point to novel candidate driver genes."
by S. Caburet et al.

We thank the two referees for their helpful comments and suggestions. We have addressed all their points as described below.

The revised portions are highlighted in red in the manuscript.

We hope that this revised version will be deemed suitable for acceptance in BMC Cancer.

Best regards,
S. Caburet and RA Veitia
on behalf of the authors.

Comments from Reviewer 1:

General comments: If the strategy used in this study is pertinent to define a comprehensive candidate driver genes list, the molecular heterogeneity associated with OGCTs would have required an analysis on a larger set of tumoral samples. However, we have to take account (i) the scarcity of OGCTs samples, (ii) the absence of large gene expression public data reported and (iii) authors did not report a molecular signature specific of OGCTs. This study is really the first tentative to identify potential driver genes and it deserve to be shared with the community.

We do agree with the referee that a larger set of tumor samples would have yielded results with better statistical support. Nevertheless, as pointed out by the referee himself, these tumors are rare, and our study deals with one of the largest cohorts analyzed so far by CGH and transcriptomics.
We have access to additional samples, but unfortunately they are formalin-fixed paraffin-embedded tumors, and we have been unable to extract RNA in proper quality allowing correct transcriptomic analysis.

Q1: p4 lines 112-113: Could authors better precise the gene expression normalization method.
a – “Gene expression was normalized by dividing the value in rearranged tumors by the mean expression in non-rearranged tumors.”
In regard to reference 9, gene expression authors should correct for : Gene expression was normalized by subtracting the value in all tumors by the mean expression in non-rearranged tumors.

The reference 9 (Benayoun et al, 2013) was related to the publication of our transcriptomic data on these tumors, and is not pertinent to explain to type of normalization we performed to combine the transcriptomic data with the CGH data.
We clarified our procedure in the main text as follows: "To better measure the impact of large-scale genomic alterations on gene expression, we divided the expression values for genes located within aneuploid regions by their mean expression in the remaining tumors. Therefore, expression ratios above/below 1 in natural scale (or above/below 0 in log2 scale) in aneuploid regions are suggestive of a correlation between genomic rearrangements and gene expression. Finally, these ratios were averaged over 30 windows (of equal size) per chromosome."

**Q2:** Which CNA thresholds are used to define gains, losses or amplification or homozygous deletion?

The threshold that we used correspond to a heterozygous gain or loss in a minimum of 50% of the tumor cells. In log2 scale these ratios are 0.322 and -0.415 respectively.

We added this information in the MM section.

**Q3:** p5 lines 121-122: “The Pearson correlation coefficient R is significant when it is >0.63 for ten samples” Could authors better explain how they obtain a threshold of 0.63.

This threshold (and the other ones mentioned in the text) were estimated using the standard test for the significance of an observed R value. Namely, the threshold was determined considering that, for a sample size N (equal to or greater than 6), with observed values of R, there is a statistic t such that:

\[ t = R \sqrt{\frac{n-2}{1-R^2}} \]

which follows approximately a student distribution with N-2 degrees of freedom. Application of this formula to any particular observed value of R will test the null hypothesis that the observed value comes from a population in which the correlation between the two variables is \( \rho = 0 \).

This explanation was added to the MM section for clarity.

**Q4:** p5 lines 125-126: Why did authors decide to eliminate some samples from the analysis? If authors did not eliminate such samples and did not adjust the threshold, do not they think that they could better test robustness of their data? Comment.

We analyzed the effect of large-scale genomic imbalances on the expression of genes included in those genomic regions in the first part of our study. For the identification of candidate driver genes, we focused on local analysis of the correlation between the genomic copy number changes and expression. In order to exclude that the local correlation was driven by the large-scale imbalances, we removed the corresponding tumor(s) for the analysis of the relevant genomic regions. For example, the H24 tumor in our cohort presents a trisomy 8 and a trisomy 14, and the H18 tumor also presents a trisomy 14. So the correlation coefficient was calculated on only 9 tumors for genes on chr8, and on only 8 tumors for genes on chr14.

In the cases we eliminated samples for this local analysis, the threshold for the significance of R was adjusted, to R>=0.67 for 9 samples, R>=0.71 for 8 samples and R>=0.76 for 7 samples.

This point was clarified in the MM section.

Chapter “Expressional correlation, .... candidate drivers”
p4 lines 134-135 : “The correlation matrix... was built using the Pearson correlation coefficient R... 10 tumors”.

Q5: With 10 samples, why did authors use a parametric statistic test (Pearson test)?
Q6: Would not be more pertinent to use the Spearman correlation test because N<30 samples?

We provide an answer to the two questions at the same time.

We agree with the referee that Spearman's ranks correlation would have been a possible choice to explore our data.

The basis of our hypothesis is that the relationship between the CGH status and the level of expression of a given gene should be mostly linear. Since the Pearson is better suited to explore linear relationship, whereas Spearman is better for monotonic relationship, we feel that our choice of using Pearson correlation coefficient is valid.

Pearson correlation has been widely used for small samples. As mentioned above, we took into account the size of our cohort when determining the threshold for significance of R. For consistency, we used the Pearson correlation coefficient for all our study.

In the precise case of the hierarchical clustering of the candidate driver genes on the basis of their expression (for the MM section cited by the referee), choosing the Spearman Rank correlation as the measure for similarity during clustering leads to exactly the same gene grouping as when using Pearson correlation coefficient (see figure below).

Hierarchical clustering of gene expression in MeV, using Spearman Rank correlation (left) of Pearson correlation coefficient (right).

Chapter Results and discussion / “CGH of Ovarian GCT shows chromosome imbalances”

Q7: I would suggest to mention in results p6 line 167 “Our analysis combined with a review of the literature (refs) compiles then data of 94 adult-type GCTs. 64 of them presented large scale alterations. This compilation shows.... “

We thank to referee for this suggestion, and changed the sentences as suggested.
Chapter Results and discussion /Identification of putative drivers….

Lines 216-218. “We set the threshold for statistical significance of Pearson’s correlations R to 0.63, which is the standard cut-off for ten samples”

Q8: The word “standard” should use only if 0.63 has been defined as a standard value from literature? In this case authors should mention in which reference it was reported.

As mentioned above in response to Question 3, this threshold (and the other ones mentioned in the text) were estimated using the test for the significance of an observed R value. This test is standard, and equivalent values can be found in classic tables.

Line 218. “this value was adjusted to 0.72 and 0.81 in cases where 1 or 2 samples were removed”. These values are in discrepancy with adjusted values mentioned lines 122-126 (0.69 and 0.71 for 9 and 8 samples, respectively).

We thank the referee for pointing out this typo. The correct values are the ones mentioned in the MM section. We corrected the values in this paragraph accordingly.

In the discussion about candidate driver genes identified as amplified or deleted in OGCTs, with correlated expression, authors should comment:
a-the exclusively amplified genes (never found deleted) with correlated expression identified with the decreasing order: AKT1 (6/10), NRD1 (5/10), MMAB (4/10), TSPAN32 (3/10),…
b-the exclusively deleted genes (never found amplified) with correlated expression identified with the decreasing order: HSPA4 (3/10), RTF1 (3/10), TADA2, c19ORF18, FAM177A1, LIMA1 (2/10)

We thank the referee for this helpful suggestion. We revised the results section accordingly, and highlighted the known or putative roles of the candidate drivers that are found altered in most tumors first. Table 1 was also modified accordingly.

Authors showed that the +14/-22 and +7/-16q associations were not random. Which candidate driver genes might affect in such combinations?

For the +14/-22 association, the candidate driver genes that could be involved are respectively AKT1 and RUNX1. For the +7/-16q combination, only PIEZO1 is present on chr16. Nevertheless, these genes are not located in the smallest regions of overlap that we could define by compiling large-scale rearrangements from several studies (see figure 1). Therefore, we do not think that it is relevant to highlight this point.

In the table supp2, a- Data on broken Amp Del.
In the footnote at the top “…that contains the genomic alterations involving segments….”. Replace “involving” by “involving”

Line 276 Figure 2b has to be replaced by Figure 3b. Figure 3b should contain p values for both concerned pathways.
We thank the referee for pointing out these errors and typos. Each one has been corrected.

For figure 3b, Bonferroni-corrected p-values for enrichment of the two functional pathways for depicted genes were added, as provided by Enrichr. The legend was clarified accordingly.

Comments from Reviewer 2:

Reviewer 2 mentioned three concerns about our study.

- "The first one is on the statistical power of the study, considering the small sample size"

As said above in response to referee 1, we do agree that a larger cohort of tumor samples would have yielded results with better statistical support. Nevertheless, as pointed out by the referee 1, these tumors are rare, and our study deals with one of the largest cohorts analysed so far conjointly by CGH and transcriptomics.

As mentioned above, we have access to additional samples, but these are formalin-fixed paraffin-embedded tumors, and we have been unable to extract RNA in proper quality allowing correct and reliable transcriptomic analysis.

As noted by the referee, it would have been preferable to apply a correction for multiple testing. But unfortunately, the current corrections such as Bonferroni or the softer Benjamini-Hochberg are way too stringent because of the small sample size of our cohort. We are conscious that this study provides leads to be further explored, and this is acknowledged at the end of the discussion: "However, we are aware that, given the small sample size for which CGH and transcriptomic data were available, this genomic exploration only provides leads for functional analyses to formally demonstrate the implication of the candidate drivers in GC tumorigenesis."

- "Secondly, the authors define the driver genes “if one or several CGH probe(s) mapped within 25 kb of at least one of its transcripts, and its transcriptomic value correlated significantly with the CGH data over all the tumors”.

These two criteria were only the first ones that we used to define potential candidate driver genes.

To be considered as a candidate driver gene in our study:

- the gene had to be within 25 kb of one or several amplified/deleted CGH probe(s) above the threshold that indicates that more that 50% of the tumor cells harbor the alteration (namely 0,322 for amplification and -0,415 for deletion, in log2 scale)
- this alteration had to be present in at least 2 different tumors
- the transcriptomic values for the gene had to be significantly correlated with its CGH data
- the CGH data was manually inspected to verify that the gene is being completely included in the genomic alteration (if not, it would be classified as broken)
- the genomic alterations are not included in known CNVs in healthy individuals.
Furthermore we refined the list of candidate driver genes, by assessing a clear functional implication in cancer or cancer-related pathways from bibliography. If a direct implication in cancer or cancer-related pathways was not evident (or unknown because the gene is not studied yet), we performed a gene set enrichment analysis on transcriptomic neighbors (genes whose expression is tightly correlated with the expression of the candidate driver gene in our tumors), in order to detect a strongly suggestive implication in cancer-related pathways.

- "Thirdly, it is well known CGH has low resolution and SNP array or NGS is a better alternative with much higher resolution. The significance of the results drawn on the CGH data would be much reduced due to the low resolution."

We understand the point raised by the referee, especially if we had used CGH array. However, in our case, we used a DNA chip containing probes distributed every 13 kb in average. This provides essentially the same resolution as a SNP microarray, where the 250k probes are distributed every 12 kb in average. Obviously, NGS would have provided a better resolution, but this is beyond the scope of this study.