Author's response to reviews

Title: Decreased GATA5 mRNA expression associates with CpG island methylation and shortened recurrence-free survival in clear cell renal cell carcinoma

Authors:

Inga Peters (peters.inga@mh-hannover.de)
Natalia Dubrowskaja (dubrowskaja.natalie@mh-hannover.de)
Michael Kogosov (michael.kogosov@stud.mh-hannover.de)
Mahmoud Abbas (abbas.mahmoud@mh-hannover.de)
Jörg Hennenlotter (joerg.hennenlotter@med.uni-tuebingen.de)
Christoph von Klot (Klot.Christoph@mh-hannover.de)
Ralph Scherer (scherer.ralph@mh-hannover.de)
Arnulf Stenzl (stenzl.arnulf@med.uni-tuebingen.de)
Axel S Merseburger (merseburger.axel@mh-hannover.de)
Markus A Kuczyk (kuczyk.markus@mh-hannover.de)
Jürgen Serth (serth.juergen@mh-hannover.de)

Version: 2
Date: 20 November 2013

Author's response to reviews: see over
Dear Dr Lars Dyrskjot and Ms Roselyn Remoto

Thank you for giving us the opportunity to revise our manuscript. We also thank the reviewers for the time and effort spent on the review and the helpful comments. Please find enclosed our revised manuscript and the rebuttal letter with a point-by-point response to the reviewer’s comments. We have considered most of remarks made by the reviewers.

We believe that our revision substantially improved the manuscript and it is now appropriate for the publication in BMC Cancer.

Sincerely,
Inga Peters

Reviewer(s) Comments to the Author:

Reviewer#1

Major Compulsory Revisions
Materials and Methods
1. The authors described that “Sampling, collection, processing, and histopathological examination of tissue samples were performed as described previously [12].“. However, in ref [12], it was not accurate how they obtained the tissues. The identity and origin of the normal control tissue is a critical factor. Therefore, it is vital to accurately describe the generation of normal tissue samples. What was the distance to the tumor lesion? If the normal control tissue was extracted in close proximity to the primary tumor lesion, did the authors account for epigenetic cancer field effects? Please include this information.
2. Why is the number of tissue samples between tumor and non-tumor tissues not equal? For example; some of kidney might be totally replaced by the tumor? Please include this information.

Results
3. The authors should describe the mean ±#S.D. of natural logarithms of relative expression (lnRQ) values of GATA5 mRNAs in tumor and non-tumor.
4. The authors did not describe the patient numbers above and below groups of
the cut off value.

5. A prognostic factor identified by univariate Cox regression can, by definition, not be independent. The authors analyzed the data by bivariate statistical models. Why did the authors not examine the multivariate Cox regression analysis?

**Reviewer#2**

*Major Compulsory Revisions*

1. Data for non-dichotomized GATA5 mRNA expression in the cox regression analysis should be shown.

2. The number of patients in the low and high GATA5 mRNA expression groups should be shown in figure 3.

3. In the discussion, the strength of the survival data should be discussed in relation to the number of samples included in the survival analysis.

*Minor Essential Revisions*

1. Figure 3 and Table 3 are labeled as Figure 2 and Table 2

2. In Table 1, there appears to be a mistake in the calculations of percentages.

3. For the logistic regression analysis please state all clinicopathological parameters tested.

4. In table 3 there is a typo in lymph.

*Discretionary Revisions*

1. Inclusion of more samples in the RFS analysis would greatly improve the manuscript.

2. In the few patient samples where methylation and RNA expression are not correlated, it would be interesting to see whether transcription from alternative transcription start sites plays a role.
#1 Materials and Methods

1. The authors described that “Sampling, collection, processing, and histopathological examination of tissue samples were performed as described previously [12].“. However, in ref [12], it was not accurate how they obtained the tissues. The identity and origin of the normal control tissue is a critical factor. Therefore, it is vital to accurately describe the generation of normal tissue samples. What was the distance to the tumor lesion? If the normal control tissue was extracted in close proximity to the primary tumor lesion, did the authors account for epigenetic cancer field effects? Please include this information.

We revised the “Material and Methods” section now describing the method how tissues were obtained and sample processing has been carried out. We also pointed out that morphologically normal kidney tissue (designated as adN in the manuscript) were isolated with a minimum 0.5 – 2 cm distance to the primary tumor lesion.

Our previous study, cited in the manuscript (Peters et al., 2012), described tumor-specific hypermethylation of \textit{GATA5} (p<0.001). However, we also noted detectable methylation of \textit{GATA5} in adN indicating the possible presence of an epigenetic field effect.

On the other hand we observed a substantial difference in \textit{GATA5} mRNA expression between adN and tumor tissues indicating that a reduction of \textit{GATA5} mRNA expression within the adN tissues due to a possible epigenetic field effect is of limited influence.

As figure 1A shows this effect might be present at the most for a small subset of tumor and adN tissue pairs showing absent or minor expression differences.

Therefore we believe that a possible \textit{GATA5} methylation within adN tissues does not affect the findings of our study.

2. Why is the number of tissue samples between tumor and non-tumor tissues not equal? For example; some of kidney might be totally replaced by the tumor? Please include this information.

We now describe in detail in the materials and methods section that of 77 ccRCCs a subset of 58 paired adjacent normal (adN) tissues were available. The different numbers are due to the fact that in 19 cases corresponding adN tissue could not be sampled.
#2 Results

3. The authors should describe the mean ± S.D. of natural logarithms of relative expression (lnRQ) values of GATA5 mRNAs in tumor and non-tumor.
We have now added this information in the result section.

4. The authors did not describe the patient numbers above and below groups of the cut off value.
We have added this information in the result section and also in figure 3.

5. A prognostic factor identified by univariate Cox regression can, by definition, not be independent. The authors analyzed the data by bivariate statistical models. Why did the authors not examine the multivariate Cox regression analysis?

The present study identified reduced GATA5 expression as a candidate biomarker of univariate prognostic value in ccRCC. Moreover our analyses demonstrated that expression of GATA5 is a non-independent parameter in bivariate models considering status of metastasis, advanced disease and tumor grade as described on pg. 6 and in table 3, while it remained as a significant parameter when compared to lymph node metastasis, age and gender. We nowhere described GATA5 expression as an independent prognosticator.

Bivariate analyses were used as a surrogate for multivariate analyses because they represent an appropriate statistical tool to estimate the effect of confounders and show only little or no constraints compared to multivariate analysis when the size of patient subgroups is limited. In respect to the explorative study design we do not see a significant drawback in using bivariate analyses instead of a single multivariate measurement particular when considering that limitations in study size often may lead to less confidential results of multivariate analyses. Therefore, in such case the application of bivariate analyses appears to be more informative.

Author comments to Reviewer 2

#1 Major Compulsory Revisions

1. Data for non-dichotomized GATA5 mRNA expression in the cox regression
analysis should be shown.

The statistical concept of our survival data evaluation in an explorative study setup *a priori* did not intend to carry out parametric cox regression analyses for some fundamental concerns. The most important counter-argument against parametric cox regression is that this method implies a linear relationship between *GATA5* expression and survival of patients. However, as the reviewer noted on the basis of figure 2, a subset of tumors showed independence of gene expression and methylation, suggesting additional factors, which alter gene expression. Moreover, we point out in our response to the comment #3.2 that expression of *GATA5* has been reported to be affected by a microRNA in kidney cells (miR-335-5p, Marques et al., 2011). Thus, our and published data indicate a multifactorial regulation of expression diminishing the probability of a linear relationship.

Our dichotomized Cox regression analysis revealed that only a subgroup of tumors, exhibiting very low expression, is associated with impaired progression-free survival. This potentially relevant effect, most likely, would have been overseen when applying parametric cox regression and a linear relationship. Conclusively, it seems debatable whether the current picture describing a subset of patients with impaired prognosis would be changed by the results of a parametric cox regression.

Apart from these fundamental concerns, we prefer the comprehensibility of dichotomized models as they allow a clear estimation of hazards of low and high expression groups as well as visualization of corresponding data in Kaplan-Meier plots.

2. The number of patients in the low and high *GATA5* mRNA expression groups should be shown in figure 3.

We have now added this information to figure 3 and the results section.

3. In the discussion, the strength of the survival data should be discussed in relation to the number of samples included in the survival analysis.

We revised the discussion section now giving more attention to the limits of our survival analysis due to the cohort size.

#2 Minor Essential Revisions

1. Figure 3 and Table 3 are labeled as Figure 2 and Table 2

We corrected this error and revised the manuscript for other typos.
2. In Table 1, there appears to be a mistake in the calculations of percentages. We checked the calculations of percentages and corrected the errors in Table 1.

3. For the logistic regression analysis please state all clinicopathological parameters tested. We have now added all clinicopathological parameters tested in logistic regression analysis and stated this information in the corresponding result section.

#3 Discretionary Revisions
1. Inclusion of more samples in the RFS analysis would greatly improve the manuscript. We agree with the reviewers comment. However, the results of this explorative study only now allow planning of a future evaluation study with an appropriate cohort size.

2. In the few patient samples where methylation and RNA expression are not correlated, it would be interesting to see whether transcription from alternative transcription start sites plays a role.

We agree that detailed functional analyses of transcription, regulation of transcription and epigenetic modulators of GATA5 are mandatory for a better understanding protein function and changes in ccRCC. Recently, we carried out an *in silico* analysis and found out that GATA5 has been identified as a candidate target for transcriptional control by the miR-335-5p microRNA both by computer as well as laboratory analysis (Marques et al., 2011). Hence, analysis of GATA5 expression is certainly a highly relevant issue of future studies. In view of the fact that the current manuscript solely describes statistical associations corresponding functional considerations were not included in our discussion.