EMBO
reports

Manuscript EMBO-2016-43581

CircHIPK3 sponges miR-558 to suppress heparanase expression in bladder cancer cells

Yawei Li, Fuxin Zheng, Xingyuan Xiao, Fei Xie, Dan Tao, Chao Huang, Dong Liu, Miao Wang, Liang Wang, Fuqing Zeng, and Guosong Jiang

Corresponding authors: Guosong Jiang and Fuqing Zeng; Union Hospital, Tongji Medical College, Huazhong University of Science and Technology

Review timeline:

Submission date:	28 October 2016
Editorial Decision:	25 November 2016
Revision received:	31 March 2017
Editorial Decision:	11 May 2017
Revision received:	01 June 2017
Editorial Decision:	23 June 2017
Revision received:	25 June 2017
Accepted:	30 June 2017

Editor: Achim Breiling

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 25 November 2016

Thank you for the submission of your research manuscript to EMBO reports. We have now received reports from the three referees that were asked to evaluate your study, which can be found at the end of this email.

As you will see, all referees acknowledge the potential interest of the findings. However, all three referees have raised a number of concerns and suggestions to improve the manuscript, or to strengthen the data and the conclusions drawn, which need to be addressed during a revision. As the reports are below, I will not detail them here. However, in particular points 1 and 2 by referee #1 need specific attention, as well as points 2 and 3 of referee #2. Regarding point 1 of referee #2 (limited novelty), we do not think that the previous report (PMID 27050392) severely affects the novelty of the present study, as you here focus on the functional characterization of this circRNA in a specific cancer setting by providing extensive in vitro and in vivo data. However, we ask you to clearly state in the revised version of the manuscript that BCRC2 has already been described previously and to discuss these previous findings in the context of your data.

Given these constructive comments, we would like to invite you to revise your manuscript with the understanding that all referee concerns must be fully addressed in the revised manuscript and in a complete point-by-point response. Acceptance of your manuscript will depend on a positive outcome of a second round of review. It is EMBO reports policy to allow a single round of revision only and acceptance or rejection of the manuscript will therefore depend on the completeness of

© European Molecular Biology Organization

your responses included in the next, final version of the manuscript.

REFEREE REPORTS

Referee #1:

General comments:

Li, Zheng, Xiao, and colleagues describe a bladder cancer related circular RNA (BCRC-2) and its function as a likely sponge for microRNA miR-558. By sequencing ribosomal RNA (rRNA)-depleted and RNase R digested total RNA from paired bladder cancer and normal bladder tissue samples, the authors identified 42,335 distinct circRNAs, of which 111 were differentially expressed between cancer and normal samples. The authors select one of these differentially expressed circRNAs (BCRC-2), and move on show that BCRC-2 is circular, that over expressing BCRC-2 inhibits migration and invasion of bladder cancer cell lines in vitro, that BCRC-2 can can pull down miR-558 and vice versa, that BCRC-2 and miR-558 have opposite expression patterns in bladder cancer and opposite effects on migration and invasion in vitro, and that constitutive expression of BCRC-2 inhibits growth and metastasis of bladder cancer cells in nude mice xenografts. The authors' pull down experiments provide compelling results for BCRC-2 acting as a miR-558 sponge, but the mechanism behind this interaction is currently not clear. Moreover, the parts of the work on the RNA-seq data analyses should be improved.

Major specific comments:

1) The authors predict six possible binding sites for miR-558 in BCRC-2, but only one of these (site 1, Fig. EV3) is a canonical miRNA site with perfect seed region complementarity to miR-558. This does not preclude miR-558 from binding the other sites, but makes site 1 the most likely true binding site. To test whether this site is functional, the authors should mutate the site and test whether BCRC-2 still can pull down miR-558 and vice versa. This experiment can show whether miR-558 binds directly to BCRC-2 through miRNA-like base pairing, which is unclear based on the authors' current data. Finally, the authors should also consider mutating the other five potential miR-558 binding sites, as this can reveal wether any of these are true binding sites as well.

2) The authors report 42,335 distinct circRNA candidates, but this number seems overly optimistic, as the candidates are only based on finding at least one single back-spliced read from a single sample. For example, the authors report four circRNA candidates that map to the RPPH1 non-coding RNA. This gene is annotated as a single exon, unspliced transcript at chr14:20811230-20811570, but the authors' candidates all map to partially overlapping parts within the transcript: chr14:20811288-20811559, chr14:20811417-20811566, chr14:20811283-20811436, and chr14:20811288-20811529. Moreover, the authors report that three of the four candidates are differentially expressed between normal and cancer, but two (chr14:20811417-20811566 and chr14:20811283-20811436) are up in normal and one (chr14:20811288-20811559) is up in cancer. These three variants are also the most abundant circRNAs with > 1.6 million reads; the last candidate is based on one single read from one cancer sample. Although the authors' bioinformatics pipeline identify these as distinct circRNAs, the candidates are partially overlapping and apparently lack the characteristics of well studied circRNAs, which originate from spliced transcripts. Similarly, the authors report multiple (9) circRNA candidates from the HIPK3 gene, including BCRC-2. Two of these candidates have the same genomic 5' end as BCRC-2, but one is shorter than BCRC-2 and the other includes the next exon from HIPK3. It is unclear which of these candidates have back-splice events that map to known or possible splice sites.

A more robust approach to find circRNA candidates would be to require support for the backsplice event from at least two distinct reads in at least two different samples, as this reduces false positives from RT template switching (see e.g. https://www.ncbi.nlm.nih.gov/pubmed/24811520). Requiring that the backsplice event is supported by annotated splice sites or canonical splice signals further reduces false positives. The authors should redo their analyses of the RNA-seq data and report both (1) which and how many candidates are supported by multiple distinct reads in multiple samples and (2) which and how many candidates are supported by known or possible splice events.

Minor comments:

3) The authors should describe the statistical analyses used to identify the significant circRNA candidates from the RNA-seq data.

4) The authors should describe exactly how the miRNA target predictions were done. Presumably, the three programs were run on the BCRC-2 sequence, but how did the authors allow for potential miRNA target sites within the backsplice region?

5) The manuscript needs some language correction, including the following:
* p.4: "Such as CDR1as or ciRS-7 [10,14], which has been reported harbors over 70 conventional binding sites for miR-7, can sequester miR-7 away from its target sites in the sense CDR1 mRNA."
* p. 7: "...BCRC-2 was insensitivity to digestion with RNase R..." (change "insensitivity" to "resistant").
* p. 8: "...we take this finding as a proper pointcut to investigate the interaction between..." (replace "pointcut")
* p. 12: "Due to the expressions of circRNAs are in a complex tissue or cell type- and development stage-specific manner [10,12], even the same circRNA could be diversely expressed in different cells or the same cell type but at different stage."

Referee #2:

In this manuscript, Li Y et al identified a circular RNA, BCRC-2, by comparing human bladder cancer and normal bladder tissue using RNA sequencing. The results demonstrated that overexpression of circular BCRC-2 significantly suppressed migration, invasion and angiogenesis of bladder cancer cells and restrained tumor growth and metastasis in nude mice. In addition, it was also found that BCRC-2 contained six binding sites for miR-558, thus exerted its anti-cancer function through the microRNA sponge mechanism. The present study hired a variety of advanced techniques to prove the hypotheses. It is a nice combination of in vitro and in vivo study. However, although this study, for the first time, elucidated the function of circular RNA BCRC-2 in bladder cancer and demonstrated its binding of miR-558, the microRNA sponge concept was not new which limited the novelty of this study. There are also some other key points that need to be addressed.

1. In this study, the BCRC-2 was derived from HIPK3 gene which circularized with single large exon2 (1099bp). Although the name was different, another study (Zheng Q et al. Nature Communication 2016) reported the same circular RNA named circHIPK3 in cancer recently. Therefore, BCRC-2 is not a new circular RNA that was identified in cancer.

2. No Northern blot results were shown to compare the mRNA and circular RNA expression upon RNaseR treatment. It is a critical experiment to demonstrate the circular status of BCRC-2. Only Northern blot in which the samples were treated with or without RNAseR, in which a molecular weight marker is used and in which only one band of the expected size is observed will not allow the authors to claim with confidence that the effects are due to overexpression of the circRNA.

3. In Figure 1H, the authors only normalize the RNase treated group to the mock treated group in RNAs extracted from cell lines. However, the RNase treatment was not used after BCRC-2 overexpression. Hence, I cannot be convinced that the effect of BCRC-2 in migration and invasion is only due to the circular RNA but not the linear RNA.

4. After siRNA treatment, the linear mRNA levels need to be shown in Figure 2E to demonstrate whether only circular RNA but not mRNA was involved in this anti-cancer function.

5. In Figure 3H, only one cell were shown in each group. A field picture containing more cells needs to be shown.

Referee #3:

In this study, the authors identify a circular RNA (called BCRC-2), which is downregulated in bladder cancer. They show that it sponges miR-558 and that its expression correlates with hespanase (HPSE) expression, migration and invasion. They finally show that BCRC-2 expression levels affect tumour growth and lung metastasis formation in mouse xenograft models.

Though BCRC-2 has been previously noted in the literature under a different name and also found to be perturbed in cancer, including bladder cancer, the functional characterization is novel and of interest. The main strength of the paper is a convincing array of functional studies. In general the molecular assays are replicated and robustly carried out. However, the variance of the measurements (standard errors on plots) were often surprisingly low, raising questions of whether biological replicates rather than technical replicates were used in all cases. The clinical samples were generally not sufficiently well characterized. Similarly, the statistical analysis for candidate discovery and selection was not described in sufficient detail to evaluate the analysis.

Specific comments:

1. Bladder cancer is a heterogeneous disease - clinically and molecularly. The authors only analyzed three pairs of tumor and normal tissue in the discovery cohort for identifying dysregulated circRNAs. This is a very small set and results may not be robust in many cases. As a results the 111 circRNAs reported in Suppl Table 1 are unlikely to all be relevant.

2. Bladder cancer arises from the urotheilum, which is the appropriate normal tissue type to compare against. The authors should clarify how their normal samples were prepared: was the urothelial biopsies enriched (or laser micro dissected)?

3. Along this line, the authors do not describe the stage and grade of the tumors analyzed with RNA-Seq. Was there a difference in carcinoma cell % in tumor section that also may explain differences? The comparison may simply reflect differences in cell composition between normal and cancer - and not true differences between cancer cells and normal urothelial cells.

4. In general, the analysis underlying the identification of the candidate (BCRC-2) is not sufficiently clear. No details are given of the statistical analysis. How did you deal with read counts of zero, which occurred often? Were pseudo-counts added? Extreme fold-changes are reported in the supplementary table and in Fig. 1C (up to 10^6). However, it appears these often arise by ration that involve zero in the denominator - this does not appear robust or trustworthy. Also, from Fig. 1C it appears the results are generally driven by small read counts (the trajectories of points). Explain what was done in more detail and use statistical analysis methods that are robust to small sample sizes.

5. Differential expression analysis: The paired structure of the samples appears not to be exploited. Could block variance and strengthen results.

6. We acknowledge that the discovery analysis is only used to pick the candidate and that an interesting candidate was chosen. However, the data should be analysed and presented in a form that conveys as much information as possible.

7. Fig 1B,C: indicate the chosen candidate.

8. Table 1: How is low and high expression defined? Relative to the normals?

9. BRCR-2 is already known in the literature as circHIPK3. The authors should use this name instead.

10. The authors must reference and discuss the existing literature on circHIPK3 and its role in cancer:
a. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. Nature comm. 2016 (doi:10.1038/ncomms11215).

11. Why were the used cell lines chosen?

12. Fig. 1I: positive control for cytoplasm lacking - why?

13. Fig 2B. The errors bars on the relative migration and invasion rates are small. We are surprised how reproducible the measurements have been, given the heterogeneity shown in the plots and our own experiments with these types of experiments.

14. Fig 2G,H: we recommend that the experiment are also done for the cell line with a high endogenous level of BCRC-2 (SV-HUC-1 from Fig. 1G).

15. Fig. 3G: Once again we surprised by the small error bars on the measurements. Are these based on biological or technical replicates? Given the conclusions drawn, these should be biological replicates.

16. Fig. 4A: Define the clinical characteristics of the 224 tumors on the TMA.

17. Fig. 4A: The TMA analysis results may simply reflect stage differences. The authors should only include tumors in this analysis comparable to the 44 samples obtained from cystectomy. HPSE may not add prognostic information beyond what is available from existing tumour grade and stage information. The authors should acknowledge / discuss this.

18. The Discussion includes a lengthy literature review. This should be trimmed and focus instead placed on discussing the results and their implications.

19. Throughout the paper. Which statistical tests were used where? Indicate together with p-values.

20. Many sentences do not read well and have syntactical flaws. Please have the manuscript seen by a native English speaker.

Minor comments:

1. Fig 2A: caption scaled, which makes it harder to match up vector and BCRC-2 columns.
2. Fig. 2: Define NC. (Negative control, we assume.)
3. Fig. 3G: note GAPDH on figure to help reader.

1st Revision - authors' response — 31 March 2017

Response to Referee #1:

Li, Zheng, Xiao, and colleagues describe a bladder cancer related circular RNA (BCRC-2) and its function as a likely sponge for microRNA miR-558. By sequencing ribosomal RNA (rRNA)-depleted and RNase R digested total RNA from paired bladder cancer and normal bladder tissue samples, the authors identified 42,335 distinct circRNAs, of which 111 were differentially expressed between cancer and normal samples. The authors select one of these differentially expressed circRNAs (BCRC-2), and move on show that BCRC-2 is circular, that over expressing BCRC-2 inhibits migration and invasion of bladder cancer cell lines in vitro, that BCRC-2 can pull down miR-558 and vice versa, that BCRC-2 and miR-558 have opposite expression patterns in bladder cancer and opposite effects on migration and invasion in vitro, and that constitutive expression of BCRC-2 inhibits growth and metastasis of bladder cancer cells in nude mice xenografts. The authors' pull down experiments provide compelling results for BCRC-2 acting as a miR-558 sponge, but the mechanism behind this interaction is currently not clear. Moreover, the parts of the work on the RNA-seq data analyses should be improved.

Major specific comments:

1) The authors predict six possible binding sites for miR-558 in BCRC-2, but only one of these (site 1, Fig. EV3) is a canonical miRNA site with perfect seed region complementarity to miR-558. This does not preclude miR-558 from binding the other sites, but makes site 1 the most likely true

binding site. To test whether this site is functional, the authors should mutate the site and test whether BCRC-2 still can pull down miR-558 and vice versa. This experiment can show whether miR-558 binds directly to BCRC-2 through miRNA-like base pairing, which is unclear based on the authors' current data. Finally, the authors should also consider mutating the other five potential miR-558 binding sites, as this can reveal whether any of these are true binding sites as well.

Response: We appreciate the referee's positive comments on our work. We also thank the referee's good suggestion for better improving our manuscript. In this revised manuscript, we mutated all the six binding sites of miR-558 on circHIPK3 (we have replaced the name of BCRC-2 with circHIPK3 to avoid misunderstanding) respectively (Appendix Figure S3), and then performed pull-down assay. First, we confirmed that mutated circHIPK3 still could be significantly pulled down by biotin-circHIPK3 probe, which was consistent with the fact that all the mutated binding sites were not located in the bio-circHIPK3 probe complementary region (Fig.3E). Next, we found that after mutating site1 or site2, the relative binding of miR-558 was significantly decreased, while mutating of the other four sites had no significance difference compared with wild type circHIPK3, respectively (Fig.3F). Besides, it showed that binding site1 was more effective than site2 (Fig.3F). These results demonstrated that binding site1 and site2, but not the others four binding sites are critical for circHIPK3 to sponge miR-558.

2) The authors report 42,335 distinct circRNA candidates, but this number seems overly optimistic, as the candidates are only based on finding at least one single back-spliced read from a single sample. For example, the authors report four circRNA candidates that map to the RPPH1 non-coding RNA. This gene is annotated as a single exon, unspliced transcript at chr14:20811230-20811570, but the authors' candidates all map to partially overlapping parts within the transcript: chr14:20811288-20811559, chr14:20811417-20811566, chr14:20811283-20811436, and chr14:20811288-20811529. Moreover, the authors report that three of the four candidates are differentially expressed between normal and cancer, but two (chr14:20811417-20811566 and chr14:20811283-20811436) are up in normal and one (chr14:20811288-20811559) is up in cancer. These three variants are also the most abundant circRNAs with > 1.6 million reads; the last candidate is based on one single read from one cancer sample. Although the authors' bioinformatics pipeline identify these as distinct circRNAs, the candidates are partially overlapping and apparently lack the characteristics of well studied circRNAs, which originate from spliced transcripts. Similarly, the authors report multiple (9) circRNA candidates from the HIPK3 gene, including BCRC-2. Two of these candidates have the same genomic 5' end as BCRC-2, but one is shorter than BCRC-2 and the other includes the next exon from HIPK3. It is unclear which of these candidates have back-splice events that map to known or possible splice sites.

A more robust approach to find circRNA candidates would be to require support for the backsplice event from at least two distinct reads in at least two different samples, as this reduces false positives from RT template switching (see e.g. https://www.ncbi.nlm.nih.gov/pubmed/24811520). Requiring that the backsplice event is supported by annotated splice sites or canonical splice signals further reduces false positives. The authors should redo their analyses of the RNA-seq data and report both (1) which and how many candidates are supported by multiple distinct reads in multiple samples and (2) which and how many candidates are supported by known or possible splice events.

Response: We thank the referee's careful reviewing of our RNA-seq data and give us important suggestions for better improving our manuscript. The false positive results really affect the accuracy of RNA-seq data. In this revised manuscript, we re-analyze the RNA-seq data according to the referee's suggestions and following the instructions as reported ("Detecting and characterizing circular RNAs", doi:10.1038/nbt.2890). First, we mapped the RNA-seq data to the human reference genome (GRCH38/hg38) by TopHat2 and identified **16,353** circRNAs with at least two distinct reads. Next, we screened all these circRNAs with the restriction that the two reads were detected from at least two different samples, and then got **6,154** circRNAs (Dataset EV1). We confirmed that all the 6,154 circRNAs were supported by known or possible splice sites, among which **4,531** circRNAs were annotated in circBase (http://www.circbase.org/) and **1,623** were novel circRNAs. The referee mentioned two circRNAs in the comment. One is RPPH1 (ENSG00000277209), which was annotated as a single exon but detected four transcripts in the previous data. However, the backsplicing reads of RPPH1 were extremely high only in one sample and others were zero. In the revised data, RPPH1 was ruled out by the restriction that at least two distinct reads in at least two different samples. Another is HIPK3 (ENSG00000110422), which was also detected multiple

transcripts in the previous data. In present data, we identified two circular transcripts of HIPK3 gene: 1) circHIPK3 (hsa_circ_0000284) was formatted by Exon2 (1099bp); 2) circRNA.1188 (hsa_circ_0008887, chr11:33286413|33328633 (-)) was formatted by Exon2 and Exon3 (1223bp). According to the results (Dataset EV1), circHIPK3 was detected in all the six samples and the reads number were much higher than circRNA.1188, which indicated circHIPK3 was the predominant circular isoform of HIPK3 gene in human bladder tissues.

3) The authors should describe the statistical analyses used to identify the significant circRNA candidates from the RNA-seq data.

Response: We thank the referee to give us this useful suggestion for better improving our manuscript. In this revised manuscript, we added "Differentially expressed circRNA selection" in *Materials and Methods (Page 19)***.** We adopted edgeR package to analysis the different expression of circRNAs between bladder cancer tissues and normal bladder tissues, and with the restriction of FDR (false discovery rate) to calculate P-value. Fold change was calculated by SRPBM (spliced reads per billion mapping). SRPBM = number of circular reads/total mapped reads (units in billion). Differentially expressed circRNAs were filtered by |FC (fold change)| ≥2 and P<0.05. Detailed information is listed in Dataset EV2.

4) The authors should describe exactly how the miRNA target predictions were done. Presumably, the programs were run on the BCRC-2 sequence, but how did the authors allow for potential miRNA target sites within the backsplice region?

Response: We thank the referee for giving us useful suggestion for better improving our manuscript. In this revised manuscript, we added "miRNA target prediction of circHIPK3" in *Materials and Methods (Page 19)***.** We predicted the miRNA binding sites of circHIPK3 using the bioinformatic database miRanda (http://www.microrna.or g/microrna/getMirnaForm.do), PITA (http://genie.weizmann.ac.il/pubs/mir07/mir07 data.html) and RNAhybrid (http://bibi serv.techfak.uni-bielefeld.de/rnahybrid/). Filtering restrictions as follows: i) Total Sore ≥140, Total Energy <-17kcal/mol; ii) combined interaction energy (ΔΔG<-10; iii) minimum free energy (MFE) ≤-20kcal/mol. Besides, to avoid omitting the miRNA target sites within the backsplice region, we selected 60 nucleotides that span the junction area of circHIPK3, which composed of 30 nucleotides from the 3' end of and 30 nucleotides from the 5' end of HIPK3 Exon2, and then analyzed the sequence using three programs with the same restrictions. Detailed information is listed in Dataset EV3.

5) The manuscript needs some language correction.

Response: We thank the referee for pointing out this. In this revised manuscript, we have corrected the language mistakes and hope these changes are good enough for better improving our manuscript.

Response to Referee #2

In this manuscript, Li Y et al identified a circular RNA, BCRC-2, by comparing human bladder cancer and normal bladder tissue using RNA sequencing. The results demonstrated that over-expression of circular BCRC-2 significantly suppressed migration, invasion and angiogenesis of bladder cancer cells and restrained tumor growth and metastasis in nude mice. In addition, it was also found that BCRC-2 contained six binding sites for miR-558, thus exerted its anti-cancer function through the microRNA sponge mechanism. The present study hired a variety of advanced techniques to prove the hypotheses. It is a nice combination of in vitro and in vivo study. However, although this study, for the first time, elucidated the function of circular RNA BCRC-2 in bladder cancer and demonstrated its binding of miR-558, the microRNA sponge concept was not new which limited the novelty of this study.

There are also some other key points that need to be addressed:

1. In this study, the BCRC-2 was derived from HIPK3 gene which circularized with single large exon2 (1099bp). Although the name was different, another study (Zheng Q et al. Nature Communication 2016) reported the same circular RNA named circHIPK3 in cancer recently. Therefore, BCRC-2 is not a new circular RNA that was identified in cancer.

© European Molecular Biology Organization

Response: We thank the referee's useful comment for better improving our manuscript. We fully agree with the referee's opinion that BCRC-2 (circHIPK3) is not a new circRNA and has been reported by another article, thus we have replaced BCRC-2 with circHIPK3 to avoid misunderstanding in the revised manuscript. In the present study, we mainly focus on the functional characterization of circHIPK3, and have revealed the novel regulatory mechanism of circHIPK3 that act as a tumor suppressor through targeting miR-558/heparanase axis in bladder cancer.

2. No Northern blot results were shown to compare the mRNA and circular RNA expression upon RNase R treatment. It is a critical experiment to demonstrate the circular status of BCRC-2. Only Northern blot in which the samples were treated with or without RNase R, in which a molecular weight marker is used and in which only one band of the expected size is observed will not allow the authors to claim with confidence that the effects are due to over expression of the circRNA.

Response: We appreciate the referee for this constructive comment. We agree that Northern blot is a critical experiment to demonstrate the circular status of circHIPK3. In this revised manuscript, we performed Northern blot using the probe targeting the backsplicing junction area to detect circHIPK3 with or without RNase R treatment on T24T and UMUC3 cells. The digestion efficiency of RNase R was confirmed by RNA electrophoresis. As the result displayed, circHIPK3 was detected by Northern blot with only one band of the expected size (1099bp), and was resistant to RNase R digestion (Figure 1E). On the other hand, detection of mRNA status of HIPK3 by Northern blot is currently ongoing project in our group and we hope that we can include these results in our near future studies to provide additional information. We thank the referee's important suggestion for better improving our manuscript.

3. In Figure 1H, the authors only normalize the RNase treated group to the mock treated group in RNAs extracted from cell lines. However, the RNase treatment was not used after BCRC-2 over expression. Hence, I cannot be convinced that the effect of BCRC-2 in migration and invasion is only due to the circular RNA but not the linear RNA.

Response: We appreciate the referee for pointing out this. In this revised manuscript, we tested circHIPK3 and HIPK3 mRNA expression levels after over-expression of circHIPK3 with or without RNase treatment (Figure 2A). We found that transfection with circHIPK3 plasmids can significantly increase the expression of circHIPK3 but have no obvious effect on the expression of HIPK3 mRNA. Meanwhile, circHIPK3 still could resistant to RNase R digestion after over-expression. These results demonstrate that the effect of BCRC-2 in migration and invasion is due to the circular RNA but not the linear RNA.

4. After siRNA treatment, the linear mRNA levels need to be shown in Figure 2E to demonstrate whether only circular RNA but not mRNA was involved in this anti-cancer function.

Response: We thank the referee for pointing out this. In this revised manuscript, we tested circHIPK3 and HIPK3 mRNA expression levels in bladder cancer cells that transfected with siRNAs respectively (Figure 2E). We found that all the three siRNAs had no significant effect on the expression of HIPK3 mRNA. These results further confirm that only circular RNA, but not mRNA, is involved in this anti-cancer function.

5. In Figure 3H, only one cell were shown in each group. A field picture containing more cells needs to be shown.

Response: We appreciate the referee's important suggestion for better improving our manuscript. In this revised manuscript, we have replaced the previous picture with a filed picture containing more cells (Figure 3H).

Response to Referee #3:

In this study, the authors identify a circular RNA (called BCRC-2), which is down regulated in bladder cancer. They show that it sponges miR-558 and that its expression correlates with heparanase (HPSE) expression, migration and invasion. They finally show that BCRC-2 expression levels affect tumor growth and lung metastasis formation in mouse xenograft models.

© European Molecular Biology Organization

Though BCRC-2 has been previously noted in the literature under a different name and also found to be perturbed in cancer, including bladder cancer, the functional characterization is novel and of interest. The main strength of the paper is a convincing array of functional studies. In general the molecular assays are replicated and robustly carried out. However, the variance of the measurements (standard errors on plots) were often surprisingly low, raising questions of whether biological replicates rather than technical replicates were used in all cases. The clinical samples were generally not sufficiently well characterized. Similarly, the statistical analysis for candidate discovery and selection was not described in sufficient detail to evaluate the analysis.

Specific comments:

1. Bladder cancer is a heterogeneous disease - clinically and molecularly. The authors only analyzed three pairs of tumor and normal tissue in the discovery cohort for identifying dysregulated circRNAs. This is a very small set and results may not be robust in many cases. As results the 111 circRNAs reported in Suppl Table 1 are unlikely to all be relevant.

Response: We thank the referee's useful comment for better improving our manuscript. We agree with the referee's opinion that three pairs of bladder cancer and normal bladder tissues may not enough to discover cohort for identifying dysregulated circRNAs. In this revised manuscript, we have re-analyzed the RNA-seq data with more stringent restrictions: at least two distinct backsplicing reads in at least two different samples. Thousands of false positive results are ruled out which was detected in previous data. We identified 6,154 circRNAs that were supported by known or possible splice sites, among which 4,531circRNAs were annotated in circBase (http://www.circbase.org/) and 1,623 novel circRNAs (Dataset EV1). We also identified hundreds of differentially expressed circRNAs (Dataset EV2). RNA-seq results provide useful information for revealing the general tendency of circRNAs expression, and help us to select candidate circRNAs for further research. We agree with the referee's suggestion that each candidate circRNA listed in Dataset EV2 (the previous Suppl Table 1) still need to be verified in clinical samples and cell lines. Take circHIPK3 for example, through RNA-seq, we found it down-regulated in bladder cancer tissues, then we tested the expression of circHIPK3 in 44 pairs of human bladder cancer and normal bladder tissues, also tested its expression in T24T, UMUC3 and SV-HUC-1 cells. After that, we finally confirmed the results were consistent with RNA-seq results.

2. Bladder cancer arises from the urotheilum, which is the appropriate normal tissue type to compare against. The authors should clarify how their normal samples were prepared: was the urothelial biopsies enriched (or laser micro dissected)?

Response: We thank the referee's useful comment for better improving our manuscript. We collected 44 pairs of bladder cancer tissues and paired adjacent normal bladder tissues from patients who underwent radical cystectomy with the standard clinical sample collection procedure. With the instruction of a skillful pathologist, we collected the normal bladder urothelium samples (≥200mg/sample) with a distant of ≥3cm from the edge of cancer tissues in the resected bladder. We have provided this information in our revised manuscript.

3. Along this line, the authors do not describe the stage and grade of the tumors analyzed with RNA-Seq. Was there a difference in carcinoma cell % in tumor section that also may explain differences? The comparison may simply reflect differences in cell composition between normal and cancer - and not true differences between cancer cells and normal urothelial cells.

Response: We thank the referee's useful comment for better improving our manuscript. We selected two bladder cancer samples of T3G3 and one sample of T1G3 for RNA-seq. We collected all the 44 pairs of bladder cancer tissues and paired adjacent normal bladder tissues from patients who underwent radical cystectomy with the standard clinical sample collection procedure. We ensured that all the cancer samples were collected from the protuberance of the lesion's location with the similar size and cut into 0.5cm tissue pieces (≥200mg) for research.

4. In general, the analysis underlying the identification of the candidate (BCRC-2) is not sufficiently clear. No details are given of the statistical analysis. How did you deal with read counts of zero, which occurred often? Were pseudo-counts added? Extreme fold-changes are reported in the

supplementary table and in Fig. 1C (up to 10^6). However, it appears these often arise by ration that involve zero in the denominator - this does not appear robust or trustworthy. Also, from Fig. 1C it appears the results are generally driven by small read counts (the trajectories of points). Explain what was done in more detail and use statistical analysis methods that are robust to small sample sizes.

Response: We thank the referee's important comment for better improving our manuscript. In this revised manuscript, we added "Differentially expressed circRNA selection" in *Materials and Methods (Page 19)*. We adopted edgeR package to analysis the different expression of circRNAs between bladder cancer tissues and normal bladder tissues, and with the restriction of FDR (false discovery rate) to calculate P-value. Fold change was calculated by SRPBM (spliced reads per billion mapping). SRPBM =number of circular reads/total mapped reads (units in billion). Differentially expressed circRNAs were filtered by |FC (fold change)| ≥2 and P<0.05 (Dataset EV2). As explained in comment 1, we have re-analyzed the RNA-seq data with more stringent restrictions and ruled out thousands of false positive results or pseudo-counts to make the RNA-seq more robust and reliable. We also choose the circRNA candidates that can be detected in multiple samples, such as circHIPK3 which was detected in all six samples.

5. Differential expression analysis: The paired structure of the samples appears not to be exploited. Could block variance and strengthen results.

Response: We thank the referee's important comment for better improving our manuscript. In this revised manuscript, we added "Differentially expressed circRNA selection" in Materials and Methods. We have replaced the previous picture with re-analyzed hot map and volcano map which displayed in Figure 1B and Expanded view figure 2B. We also provide the detailed differential expression information in Dataset EV2.

6. We acknowledge that the discovery analysis is only used to pick the candidate and that an interesting candidate was chosen. However, the data should be analyzed and presented in a form that conveys as much information as possible.

Response: We thank the referee's important suggestion for better improving our manuscript. In this revised manuscript, we provided detailed information of seven candidate circRNAs which were amplified by divergent primers and tested by RT-PCR (Appendix Figure S1 and Appendix Figure S2).

7. Fig 1B,C: indicate the chosen candidate.

Response: We appreciate the referee for pointing out this. In this revised manuscript, we pointed out circHIPK3 in Figure 1B and Expanded view figure 2B.

8. Table 1: How is low and high expression defined? Relative to the normal?

Response: We thank the referee for pointing out this. High expression is defined as the relative quantification ≥ 2-fold compared with normal. Low expression is defined as the relative quantification ≤ 0.5-fole compared with normal. We have provided this information in our revised manuscript.

9. BRCR-2 is already known in the literature as circHIPK3. The authors should use this name instead.

Response: We thank the referee for pointing out this. In this revised manuscript, we have replaced BCRC-2 with circHIPK3 to avoid misunderstanding.

10. The authors must reference and discuss the existing literature on circHIPK3 and its role in cancer: a. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. Nature comm. 2016 (doi:10.1038/ncomms11215).

Response: We thank the referee's important comment for pointing out this. In this revised manuscript, we have referenced this report and also discussed it in discussion.

© European Molecular Biology Organization

11. Why were the used cell lines chosen?

Response: T24T and UMC3 are two characteristic invasive bladder cancer cell lines. In present study, we are focus on the function of circHIPK3 that regulates the migration and invasion of bladder cancer, thus we choose these two cell lines.

12. Fig. 1I: positive control for cytoplasm lacking - why?

Response: We thank the referee's important comment. In this revised manuscript, we added 18S, which was abundantly expressed in cytoplasm, as positive control (Figure 1I).

13. Fig 2B. The errors bars on the relative migration and invasion rates are small. We are surprised how reproducible the measurements have been, given the heterogeneity shown in the plots and our own experiments with these types of experiments.

Response: We thank the referee for pointing out this. We strictly performed cell wound healing assay with standard procedure and repeated the experiment three times independently.

14. Fig 2G, H: we recommend that the experiment is also done for the cell line with a high endogenous level of BCRC-2 (SV-HUC-1 from Fig. 1G).

Response: We thank the referee's important comment for better improving our manuscript. SV-HUC-1 was a human immortalized uroepithelium cell line with no invasive ability and used as normal bladder cells. We have preformed the transwell migration and invasion experiments on SV-HUC-1 cell after knocking down of circHIPK3. However, knock-down the expression of circHIPK3 did not give the abilities of SV-HUC-1 cells to migration and invasion. We supposed that only knock-down the expression of circHIPK3 was not enough to change the non-invasive feature of SV-HUC-1 cells.

15. Fig. 3G: Once again we surprised by the small error bars on the measurements. Are these based on biological or technical replicates? Given the conclusions drawn, these should be biological replicates.

Response: We thank the referee's important comment for better improving our manuscript. We strictly performed biotin-miRNA pull down experiment with standard procedure and repeated the experiment three times independently.

16. Fig. 4A: Define the clinical characteristics of the 224 tumors on the TMA.

Response: We thank the referee's important comment for better improving our manuscript. We obtained the clinical characteristics of the 224 tumors from *R2: Genomics Analysis and Visualization Platform (*https://hgserver1.amc.nl/cgi-bin/r2/ main.cgi). We concluded the results as follows:

Parameters	Group	Cases
Gender	Male	163
	Female	61
Age at surgery	<55	14
	≥55	210
Pathological stage	pTa-T1	173
	pT2-T4	51
Grade	G1	45
	G2	85
	G3	94
Total	224	

© European Molecular Biology Organization

17. Fig. 4A: The TMA analysis results may simply reflect stage differences. The authors should only include tumors in this analysis comparable to the 44 samples obtained from cystectomy. HPSE may not add prognostic information beyond what is available from existing tumor grade and stage information. The authors should acknowledge / discuss this.

Response: We thank the referee for pointing out this. We obtained Kaplan–Meier survival result from *R2: Genomics Analysis and Visualization Platform (*https://hgserver1.amc.nl/cgi-bin/r2/main.cgi). The result demonstrates that bladder cancer patients with high HPSE expression have worse prognosis than that of HPSE low expression patients. We also discussed it in discussion.

18. The Discussion includes a lengthy literature review. This should be trimmed and focus instead placed on discussing the results and their implications.

Response: We thank the referee's important suggestion for better improving our manuscript. In this revise manuscript, we have re-written the discussion according to the referee's suggestion.

19. Throughout the paper. Which statistical tests were used where? Indicate together with p-values.

Response: We thank the referee for pointing out this. In this revise manuscript, we have added statistical tests that were used in Figure legends and Expanded view figure legends.

20. Many sentences do not read well and have syntactical flaws. Please have the manuscript seen by a native English speaker.

Response: We thank the referee for pointing out this. In this revise manuscript, we invited a native English speaker to correct the language flaws to make the article read well.

<u>Minor comments:</u>

1. Fig 2A: caption scaled, which makes it harder to match up vector and BCRC-2 columns.

Response: We thank the referee's useful comments for better improving our manuscript. In this revise manuscript, we tested circHIPK3 and HIPK3 mRNA expression levels after over-expression circHIPK3 with or without RNase treatment (Figure 2A), according to the referee's suggestion.

2. Fig. 2: Define NC. (Negative control, we assume.)

Response: We thank the referee kindly reminds us this problem. In this revised manuscript, we have defined NC as Negative control in Figure legends and Expanded view figure legends.

3. Fig. 3G: note GAPDH on figure to help reader.

Response: We thank the referee for pointing out this. In this revise manuscript, we have noted GAPDH in Figure 3G.

2nd Editorial Decision — 11 May 2017

Thank you for the submission of your revised manuscript to our editorial offices. We have now received the reports from the three referees that were asked to re-evaluate your study that you will find enclosed below. As you will see, all three referees now support the publication of your manuscript in EMBO reports. However, both referee #2 and referee #3 have some further comments that we ask you to address in a final revised version. Please add the data requested by referee #2 to the current manuscript, and address the remaining points mentioned by referee #3. Please also have the final manuscript corrected by a native speaker. Further, I have these editorial requests:

The title of your manuscript is currently too long. Please provide a simpler and shortened title.

Please include dataset legends as tab in the dataset .xls files.

The Appendix needs a more detailed ToC including page numbers and figure/table legends. Fig. S1 seems to be a table, thus name this Appendix Table S1. Figure S2 seems to be a mix of table and figure. Please separate these and rename accordingly. The Appendix items are also not called-out chronologically in the text (S2 before S1). Please change the order of the Appendix items accordingly. Then please update the call-outs in the article file.

Could you please check for all figures if they are conforming to our guidelines? (E.g. figures should not have landscape format.)

Could you provide higher resolution images for the Western and Northern blots?

Please separate more clearly the panels in Figures 4F and 5E.

The scale bars are rather small in some microscopic panels and also vary in type. Please provide uniform and thicker scale bars. Please also use scale bars without writing on them (this cannot be read in the online version of the figures). Please state the size of the scale bars only in the figure legend.

Finally, please provide an ORCID digital identifier for co-corresponding author Fuqing Zeng.

I look forward to seeing a revised version of your manuscript when it is ready. Please let me know if you have questions or comments regarding the revision.

REFEREE REPORTS

Referee #1:

The authors have appropriately addressed my previous concerns.

Referee #2:

The authors have reasonably addressed the issues raised in the comments, except the detection of HIP3 mRNA by Northern blot. They indicated that this is in their currently ongoing project. I don't understand why detection of HIP3 expression belongs to an ongoing project but not this one.

Referee #3:

Yawei et al. have done a good job of addressing many of my comments and concerns and have also added additional experimental to address concerns of the other referees. However, some of my comments were not considered and not addressed, as detailed below. There are still numerous language errors and typos that needs to be fixed as it affects the reading and in some place makes the meaning unclear.

The miR-558 family is special in the sense that it is derived from MADE1 repetitive elements (e.g., http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000203). Some mention and discussion of this would be appropriate. Furthermore, the mirBase read evidence supporting miR-558 is atypical in that the start form is lacking: http://www.mirbase.org/cgi-bin/mirna_entry.pl?acc=MI0003564.

1) The statistical analysis of the RNAseq data is still not clear. In the methods the authors say that they "adopted the edgeR package to analysis [analyse] the different [differences in] expression of circRNAs between bladder cancer tissues and normal bladder tissues, and with the restriction of FDR (false discovery rate) to calculate p-value". In addition to the language issues (my suggestions given in the square brackets), the meaning is not clear. What does "restriction of FDR" mean? Are the results thresholded on q-values and not p-values? P-values are reported in the text and their meaning is thus not clear.

2) Figure EV2 is not especially informative and hard to read. Which samples are shown? (Some faint color caption may be given centrally in the plot, but it cannot be read of easily.) What is the

scale of the axis? If the intention is that the different samples should be comparable, then a zoomed-in on some section should be provided. It is not possible to evaluate if the signal in the samples correlate in the given plot.

3) In my comment #17, I had requested that the authors should acknowledge or discuss that the survival analysis in Fig. 4A of low and high levels of HPSE correlation with survival might not contribute independent prognostic value. In particular, that HPSE expression level may contribute additional prognostic information beyond what is present in stage and grade. A multivariate analysis incl. grade and stage should be made to evaluate if this is the case. If not performed, the authors should acknowledge the above. In the response the author say they discuss the results in the Discussion, which I could not confirm.

Examples of language issues [my suggestions given in square brackets]:

P. 6: "can quantitatively compared [with] backsplices"
P. 6: "for each sample and [the] computational"
P. 7: "one that [was] stably"
P.7: "from [the] HIPK3 gene and was [skip was] consisted by [of] the head-to-tail...]
P.10: "one that [was] abundantly"
P.12: "In bladder cancer cell which [with] stable over-expression"
P.15: second last sentence on page not clear, starting with: :"Exonic circRNAs also contain..."
P.26: ",," -> ","
P.37: legend of Fig. EV2.A: "DAN [DNA]"

2nd Revision - authors' response 01 June 2017

Thank you for your letter and for the referee's recognition of our research work. We have revised our manuscript according to the editor's and referee's requests. The major changes of the manuscript are listed as below:

1. We have shortened our manuscript title as "CircHIPK3 sponges miR-558 to suppress heparanase expression in bladder cancer cells."

2. We have added the dataset legend for each dataset file, and we also have rearranged and renamed the tables and figures in Appendix.

3. We have provided higher resolution images for Western and Northern blots as well as modified the figures according to editorial requests. Thanks a lot for your suggestions for better improving our manuscript!

4. The ORCID of co-corresponding author Fuqing Zeng is 0000-0003-2059-0952.

5. According to referee #2 requesting, we used the probe targeting Exon2 of HIPK3 mRNA to detect both the circular and linear form of HIPK3 transcripts, and it showed that the fragment of linear form of HIPK3 was digested by RNase R and circHIPK3 was retained after RNase R treatment (Fig.1E).

6. According to the points mentioned by referee #3, we replied as follows:

1) As referee #3 pointed out, some miRNAs were derived from miniature inverted-repeat transposable elements, such as miR-548 family, which was reported in the reference that the referee provided. It has been reported that miR-558 is co-expressed with its host gene baculoviral IAP repeat containing 6 (BIRC6), indicating that they are likely to be expressed from the same promoter. We have discussed these in the fourth paragraph of the discussion.

2) The samples that used for RNA-seq were obtained from human bladder tissues, and the results were thresholded on p-values, rather than q-value. As results listed in Dataset EV1 and Dataset EV2, differentially expressed circRNAs were filtered by |FC (fold change)| 2 and P<0.05. We have corrected the confusing statement and hope the meaning is thus clear.

3) We have modified the Figure EV2A to make it easier to be read and understood. As showed in Fig.EV2A, the outside circle represents the genomic DNA, and the red color represents circRNAs reads junction of each sample. Different color represents different sample. Because the locus of circHIPK3 is chr11:33286413|33287511 (-), so we zoomed-in chromosome 11. The scale of the axis is 106 bp. Fig.EV2A provides the general distribution of circRNA detected by RNA-seq, and detailed information can be found in Dataset EV1 and Dataset EV2.

4) We fully agreed with referee #3 and also discussed it in the discussion that only HPSE expression might not provide independent prognostic value, and a multivariate analysis that including tumor grade and stage might provide more solid prognostic information.

5) We appreciated that referee #3 pointed out and kindly corrected the language errors in our manuscript. This final manuscript has been corrected by a native speaker.

We hope that this revised manuscript meet the requirement of EMBO Reports.

POINT BY POINT RESPONSE

Response to Referee #2:

The authors have reasonably addressed the issues raised in the comments, except the detection of HIP3 mRNA by Northern blot. They indicated that this is in their currently ongoing project. I don't understand why detection of HIP3 expression belongs to an ongoing project but not this one.

Response: We appreciate the refereeís positive comments on our work. We used the probe targeting Exon2 of HIPK3 mRNA to detect both the circular and linear form of HIPK3 transcripts, and it showed that the fragment of linear form of HIPK3 was digested by RNase R and circHIPK3 was retained after RNase R treatment (Fig.1E).

Response to Referee #3:

1 The miR-558 family is special in the sense that it is derived from MADE1 repetitive elements (e.g.,http://journals.plos.org/plosone/article?id=10.1371/ journal. pone.0000203). Some mention and discussion of this would be appropriate.

Response: We thank the referee to give us this useful suggestion for better improving our manuscript. Some miRNAs were derived from miniature inverted-repeat transposable elements, such as miR-548 family, which was reported in the reference that the referee provided. It has been reported that miR-558 is co-expressed with its host gene baculoviral IAP repeat containing 6 (BIRC6), indicating that they are likely to be expressed from the same promoter. We have cited this reference and also discussed these in the discussion. (page 16, paragraph 2, line 1~4).

2) The statistical analysis of the RNAseq data is still not clear. In the methods the authors say that they "adopted the edgeR package to analysis [analyse] the different [differences in] expression of circRNAs between bladder cancer tissues and normal bladder tissues, and with the restriction of FDR (false discovery rate) to calculate p-value". In addition to the language issues (my suggestions given in the square brackets), the meaning is not clear. What does "restriction of FDR" mean? Are the results thresholded on q-values and not p-values? P-values are reported in the text and their meaning is thus not clear.

Response: We thank the refereeís careful reviewing and give us important suggestions for better improving our manuscript. The samples that used for RNA-seq were obtained from human bladder tissues, and the results were thresholded on p-values, rather than q-value. As results listed in Dataset EV1 and Dataset EV2, differentially expressed circRNAs were filtered by |FC (fold change)| 2 and P<0.05. We have corrected the confusing statement in ìMaterials and Methods, Differentially expressed circRNA selectionî (page 19~20), and hope the meaning is thus clear.

3) Figure EV2 is not especially informative and hard to read. Which samples are shown? (Some faint color caption may be given centrally in the plot, but it cannot be read of easily.) What is the scale of the axis? If the intention is that the different samples should be comparable, then a zoomed-

in on some section should be provided. It is not possible to evaluate if the signal in the samples correlate in the given plot.

Response: We appreciate the referee for pointing out this. We have modified the Figure EV2A to make it easier to be read and understood. As showed in Fig.EV2A, the outside circle represents the genomic DNA, and the red color represents circRNAs reads junction of each sample. Different color represents different sample. Because the locus of circHIPK3 is chr11:33286413|33287511(-), so we zoomed-in chromosome 11. The scale of the axis is 106 bp. Fig.EV2A provides the general distribution of circRNA detected by RNA-seq, and detailed information can be found in Dataset EV1 and Dataset EV2.

4) In my comment #17, I had requested that the authors should acknowledge or discuss that the survival analysis in Fig. 4A of low and high levels of HPSE correlation with survival might not contribute independent prognostic value. In particular, that HPSE expression level may contribute additional prognostic information beyond what is present in stage and grade. A multivariate analysis incl. grade and stage should be made to evaluate if this is the case. If not performed, the authors should acknowledge the above. In the response the author say they discuss the results in the Discussion, which I could not confirm.

Response: We thank the refereeís important comment for better improving our manuscript. We fully agreed with referee3 and also discussed it in the discussion (page 17, line 3~6) that only HPSE expression might not provide independent prognostic value, and a multivariate analysis that including tumor grade and stage might provide more solid prognostic information. Finally, we appreciated that referee#3 pointed out and kindly corrected the language errors in our manuscript. This final manuscript has been corrected by a native speaker.

3rd Editorial Decision 23 June 2017

Thank you for the submission of your revised manuscript to our editorial offices. We have now received the reports from the two referees that were asked to re-evaluate your study that you will find enclosed below. As you will see, both referees now support the publication of your manuscript in EMBO reports. However, referee #3 has a final comment that we ask you to address in a further revised version. Further, I have a few more editorial requests:

The abstract should be written in present tense. Please change this.

Please remove the bullet points and the summary from the main manuscript text. I have saved this in a separate file and will export this together with the manuscript to our publisher after acceptance.

The scale bars in Fig. 6C/D still have writing on them and are white. Please use also here scale bars without writing on them (this cannot be read in the online version of the figures). Please state the size of the scale bars only in the figure legend.

Please explain in the legend to figure 1E what the upper and lower panels show (probed blot and gel below, I guess).

Finally, we need the ORCID for Fuqing Zeng to be linked to the author profile on our website. This can only be done by the author.

REFEREE REPORTS

Referee #2:

Is suitable for publication in EMBO reports without revision.

Referee #3:

The authors have addressed all my comments. However, the statement they added to the discussion on multivariate analysis should be revised:

"In this study, we found that patients with higher HPSE expression had worse survival probability by using R2 Genomics Analysis, indicating that a multivariate analysis that including tumor grade and stage may provide more prognostic information."

The point of my comment was that the authors had not shown that HPSE added any additional prognostic value over grade and stage and might just correlate strongly with these. A step-wise multivariate analysis would be needed to demonstrate that it held independent prognostic value.

3rd Revision - authors' response 25 June 2017

Thank you for your letter and for the refereesí support of the publication of our manuscript. We have revised our manuscript according to the editor's and referee's requests. The major changes of the manuscript are listed as below:

1. We have written the abstracts with present tense.
2. We have removed the bullet points and the summary from the main manuscript text.
3. We have changed the scale bars in Fig 6C/D according to editorial requests.
4. We have added the explanation of Fig 1E to make it easier to understand. (Figure legends, page 30)
5. We also have added the ORCID of Fuqing Zeng in the EMBO Reports website.
6. According to referee #3's request, we have revised the statement in the discussion as "In this study, we found that patients with higher HPSE expression had worse survival probability by using R2 Genomics Analysis. However, we had not shown that HPSE added any additional prognostic value over grade and stage and might just correlate strongly with these, and a multivariate analysis would be needed to determine whether the high expression of HPSE held independent prognostic value." (Page 16, lines 3-8)

4th Editorial Decision 30 June 2017

I am very pleased to accept your manuscript for publication in the next available issue of EMBO reports. Thank you for your contribution to our journal.

© European Molecular Biology Organization

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Guosong Jiang & Fuqing Zeng
Journal Submitted to: EMBO Reports
Manuscript Number: EMBOR-2016-43581V1

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

USEFUL LINKS FOR COMPLETING THIS FORM

Link	Resource
http://www.antibodypedia.com	Antibodypedia
http://1degreebio.org	1DegreeBio
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo	ARRIVE Guidelines
http://grants.nih.gov/grants/olaw/olaw.htm	NIH Guidelines in animal use
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm	MRC Guidelines on animal use
http://ClinicalTrials.gov	Clinical Trial registration
http://www.consort-statement.org	CONSORT Flow Diagram
http://www.consort-statement.org/checklists/view/32-consort/66-title	CONSORT Check List
http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun	REMARK Reporting Guidelines (marker prognostic studies)
http://datadryad.org	Dryad
http://figshare.com	Figshare
http://www.ncbi.nlm.nih.gov/gap	dbGAP
http://www.ebi.ac.uk/ega	EGA
http://biomodels.net/	Biomodels Database
http://biomodels.net/miriam/	MIRIAM Guidelines
http://jjj.biochem.sun.ac.za	JWS Online
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html	Biosecurity Documents from NIH
http://www.selectagents.gov/	List of Select Agents

B- Statistics and general methods

Please fill out these boxes (Do not worry if you cannot see all your text once you press return)

Question	Answer
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	We collected 44 pairs of bladder cancer tissues and paired adjacent normal bladder tissues from patients who underwent radical cystectomy at our department of Urology between 2014 and 2016, which ensure adequate power to detect a pre-spcified effect size.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	For animal expriment, we divided the 4-week nude mice into 4 groups,each group contains 6 nude mice.
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Inclusion criteria:1)the patient without other cancer history;2)the pothological result confirmed to be bladder cancer; 3)hot ischemia<30min;4) tumor size>3cm.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	For human clinical samples, we chose the tumor size that >3cm and cut off the similar size of the samples with entier bladder wall to ensure adequate power to perform RNA-seq and other researches.
For animal studies, include a statement about randomization even if no randomization was used.	We randomly divided the nude mice into four groups, each group contains 6 mice.
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	For animal expriment, we chose 4-week nude mice and ensured the average weight of nude mice in each group without stastatic difference.
4.b. For animal studies, include a statement about blinding even if no blinding was done	We did animal study with no blinding.
5. For every figure, are statistical tests justified as appropriate?	Yes.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Yes. The data of present study meet the assumption of the tests, and we use SPSS 19.0 to assess it
Is there an estimate of variation within each group of data?	Yes.
Is the variance similar between the groups that are being statistically compared?	Yes.

C- Reagents

Question	Answer
6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	The antibodies we used for western blot as follows:HPSE(Cat No:24529-1-AP,Proteintech, USA),VEGF(Cat No:19003-1-AP,Proteintech, USA), MMP-9(Cat.No:10375-2-AP, Proteintech, USA),β-actin(Cat.No:60008-1-Ig, Proteintech, USA),HRP-conjugated secondary goat anti-mouse(
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	Human invasive bladder cancer cell line T24 and the paired metastatic cell line T24T was kindly provided by Dr. Dan Theodorescu (Departments of Urology, University of Virginia,

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

Question	Answer
8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	We used 4-week-old female BALB/c nude mice.
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	All animal experiments were approved by the Animal Care Committee of Tongji Medical College.
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	We have consulted the ARRIVE guidelines to ensure that other relevant aspects of animal were reported adequately.

E- Human Subjects

Question	Answer
11. Identify the committee(s) approving the study protocol.	All human specimens were obtained with appropriate informed consent from the patients and approved by the Institutional Review Board of Tongji Medical College of Huazhong University of Science and Technology.
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	We colleced the clinical samples with informed consent obtained from every patient and our experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Departmen of Health and Human Services Belmont Report.
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	We do not need to publicate the patient's photos.

14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	The RNA-seq data have been deposited in the Gene Expression Omnibus database under accession code GSE97239.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	We agree with the journal's data policy. We provide supplementary materials in Expanded View figure 1-5, Dataset EV1-EV3,and Appendix figure S1-S3.
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	We have provided the RNA-seq to GEO which has no restriction for ervery researcher to get the data freely. We also get the patients' permission to use their samples for scientific research. However, we haven't got the agreement from the patient to open their information to the public repositories.
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	In present study, we referenced the Kaplan–Meier survival data of 224 well-defined bladder cancer cases derived from R2 Genomics Analysis and Visualization Platform (http://r2.amc.nl). We cited it and also mentioned this in the article.
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	We agree with this policy. We have up-load the RNA-seq data to the public open access data GEO, the accession code is GSE97239.

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	Our study are complied with the dual use research restrictions.