Author’s response to reviews

Title: Low expression levels of hepsin and TMPRSS3 are associated with poor breast cancer survival

Authors:

Mikko Pelkonen (mikkopel@student.uef.fi)
Kaisa Luostari (kaisa.luostari@uef.fi)
Maria Tengström (maria.tengstrom@kuh.fi)
Hermanni Ahonen (herma@student.uef.fi)
Bozena Berdel (bozena.berdel@uef.fi)
Vesa Kataja (vesa.kataja@ksshp.fi)
Ylermi Soini (ylermi.soini@uef.fi)
Veli-Matti Kosma (veli-matti.kosma@uef.fi)
Arto Mannermaa (arto.mannermaa@uef.fi)

Version: 2 Date: 29 March 2015

Author’s response to reviews: see over
Dear Editor and the Reviewers,

Thank you very much for your valuable notes and comments about our manuscript. We are very thankful for having the opportunity to resubmit our manuscript according to reviewers’ comments. In this letter, we have answered all comments point by point and revised manuscript accordingly. For each specific revision included in the manuscript, we have marked the page number and the line where the revision begins. In the resubmission we have also included a supplementary version of the manuscript in which we have highlighted all the revisions.

Reviewer #1, Tiziana Triulzi:

1. *The paper included several redundant analyses that make the paper hard to follow and not focused. The authors should choose to focus on TMPRSSs mRNA or protein expression levels and give corresponding results in the main text and put the others in the supplementary.*

First of all, we would like to thank you kindly for your comments. In our manuscript we wanted to approach the expression of hepsin and TMPRSS3 in human breast cancer regarding both mRNA and protein expression levels, especially as we found positive correlations between mRNA and protein expression levels. *TMPRSS1* expression levels correlated with hepsin protein expression levels \(r = 0.18; P = 0.05; n = 112\), and *TMPRSS3* expression levels correlated with *TMPRSS3* protein expression levels \(r = 0.24; P = 0.04; n = 147\).
We have revised our manuscript according to your valuable comments and we wish that in the current version of the manuscript our conclusions are presented more precisely.

2. The authors concluded that both TTSPs are potential novel therapeutic targets. It is not clear how two molecules that are associated with bad prognosis if low expressed could be considered as therapeutic targets. Instead, these molecules could be used as prognostic biomarkers.

Thank you for highlighting this. Our results support indeed the potential role of hepsin and TMPRSS3 as prognostic biomarkers. It is true that their potential as therapeutic targets cannot be shown clearly from our results and that is why we have decided to re-form the sentences regarding the potential role of these molecules as novel therapeutic targets from the conclusions chapter and also from the abstract as follows:

P4, line 67: The results suggest that the TTSPs hepsin and TMPRSS3 may have similar biological functions in the molecular pathology of breast cancer.

P23, line 522: The results showed that both TTSPs have potential as prognostic biomarkers.

3. It is not clear how the authors analyzed the immunoreactivity of TTSPs in tumor samples. What does the ‘median value of immunohistochemical scores’ means? Why did the author choose the median value? Usually, the immunoreactivity mean value of three areas of the tumor slides is used.

As presented in the comment above, we first analyzed the immunoreactivity in three areas of each tumor slide. Then we took the mean value of these three separate values and as a result we had a mean value for immunoreactivity in each tumor slide, for example 372 mean values for hepsin immunoreactivity. Then we divided these 372 samples into two groups, low and high expression, according to
the median value which was calculated based on the 372 single tumor immunoreactivity values. Hopefully this explanation answers your question.

4. **It is not clear how the authors analyzed the qPCR data. How did the authors use standard curves for sample quantification? The authors need to better explain these analyses.**

We would like to highlight that the exact PCR experiment type in these analyses is absolute quantification using the standard curve method. After we had created a standard curve we compared the unknown samples to the standard curve to achieve the exact expression values in the samples. We used Mx3000P PCR device in the analyses and the following description of the method is based on the MxPro QPCR software instruction manual by Agilent Technologies. The initial amount of the tumor sample targets was quantified using real-time PCR analysis based on threshold cycle (Ct) determinations.

First we synthesized a solution of cDNA from 5 randomly-selected KBCP tumor samples in a final concentration of 50 ng/µl. Then we created a series of standards, containing a serial dilution of this known amount cDNA, and we amplified them to generate a curve that relates the initial quantity of the target in the standard sample to the Ct. Each PCR plate included a separate standard curve. In the qPCR reaction volume was 20 µl per well. The strongest standard curve point was 75 ng and we used a 3-fold method leading to the sixth and last standard curve point containing 0.309 ng. We used 96-well plates so with a separate standard curve in each plate we could run 24 samples in each plate as all samples were analyzed in triplicate. Then we used the standard curve to derive the initial template quantity in unknown wells based on their Ct values. As a measure of quality we included NTC (no template control) and No RT (no reverse transcriptase control) wells both in triplicate in each plate. As already presented in the methods section of the manuscript, the PCR thermal profile was 1 cycle at 95-°C for 3 minutes followed by 45-55 cycles at 95-°C for 20 seconds plus 30 seconds at 60-°C. The assays for the studied gene and the endogenous control (PPIA gene) were in the same reaction.
For each standard curve plotted, the graph legend displays the R squared (RSq) value and a value for slope of the curve. In addition, a PCR amplification efficiency is displayed for each plate. Our quality standard value for efficiency ≒ 100% (acceptable 90 – 100%), for RSq 1 (acceptable ≥ 0.985), and for slope -3.32 (acceptable -3.1 – -3.6). These quality standard values needed to be met in order to proceed in the analyses. After approving the standard curve the software automatically calculates and displays the initial template quantities in each well based on the standard curves for all the unknown samples in the same experiment. The quantities for each well are given separately for both the target gene (in this case TMPRSS1 or TMPRSS3) and the endogenous control gene (in this case PPIA). After this we calculated the relative expression values for each sample by dividing the raw expression value of the studied gene in the sample with the raw PPIA expression in the sample. The samples were then divided into two groups to be used in the SPSS analyses (low and high expression) based on the median expression value of all the samples. (Source used: Instruction Manual, MxPro QPCR Software for Mx3000P and Mx3005P QPCR Systems, Agilent Technologies).

We hope that this information better explains our PCR analyses. We think that the methods section in the manuscript is sufficient in its current form as it follows the guidelines of the previously published manuscripts of our group. However, we have slightly re-formed the first sentence in the qPCR section:

**P9, line 172:** Of the KBCP sample set, 125 invasive breast cancer samples and 16 benign breast tumor samples were available for TMPRSS1 mRNA absolute quantification by real-time PCR, and 167 invasive and 23 benign samples were available for TMPRSS3 mRNA quantification.

5. *In Table 1, the authors need to give data regarding patients’ treatment and insert analysis of TTSPs association with breast cancer survival according to treatment group if possible.*

Thank you very much for this valuable comment. We asked an experienced clinician, specialized oncologist Maria Tengström from our study group to help us
with the treatment analyses. We found major new results which especially reassert the importance of low \( \text{TMPRSS1} \) mRNA expression regarding poor breast cancer prognosis. These results improve the quality of our manuscript and are now presented in the results section of the revised manuscript under the heading: “\( \text{Low TMPRSS1 mRNA expression is associated with poor survival among patients treated with radiotherapy} \)”.

As a result we have also added a new figure (Fig. 5) and a legend to it and re-formed Table 1 to include the adjuvant therapy data as suggested. We have also included a section in discussion and a few sentences in abstract and in conclusions to emphasize the new results:

**P3, line 62:** Low TMPRSS1 mRNA expression was also an independent marker of poor breast cancer prognosis in patients treated with radiotherapy \( (P = 0.034; \text{HR}, 2.344; 95\% \text{ CI, 1.065-5.160}) \).

**P20, line 475:** When the treatment data was included in the multivariate survival analyses, low \( \text{TMPRSS1} \) mRNA expression remained an independent factor of poor prognosis in patients who were treated with radiotherapy (followed by the rest of the section).

**P22, line 519:** Furthermore, low \( \text{TMPRSS1} \) mRNA expression is an independent marker of poor clinical outcome in patients treated with radiotherapy.

6. In Table 2, results regarding all covariates used in the multivariate analyses need to be included. Moreover the cut-off used to classify low tumors as low expressing need to be better specified.

Thank you bringing up these points as re-forming those points improves legibility and intelligibility of the manuscript. We have re-formed table 2 so that the results regarding all covariates used in the analyses are now presented. Regarding Cox regression multivariate analyses presented in Table 2, in mRNA expression analyses we divided the patient cases in low and high expression groups according to the calculated median value of the expressions. Please read our answer for comment 4. above in which we have discussed this closely. We have the following sentence also in the legends for Figure 3 and 4.
Patients were divided into high and low expression groups relative to the median expression values.

7. **Data presented in Figure 2 are redundant. Instead, Supplementary Table 1, in which association with all covariates is reported, could be moved as main table.**

   This is a valuable comment indeed and in the revised manuscript we moved Supplementary Table S1 as a main table (Table 2 in the current version). We moved Figure 2 as a supplementary file (Supplementary Figure S2). We wish not to remove it entirely from the manuscript as we think that the figure has important graphical value. In addition, we have improved the figure by including an expression bar of benign tumors next to the ones grouped by tumor grade. The figure presents clearly the differences for example between benign and grade I tumors as well as between different grades. In Table 2 we have also included \( P \) values for comparing mRNA expression in benign tumors versus grade I tumors.

8. **In Figure 3, it is not clear why the authors reported significances calculated by several tests. It looks like they do not know which one is the best for their data, since they are similar but not identical. Moreover, it is not clear why they reported association of Hepsin with survival in 10 years follow up. The authors need to show data, they have, in a 20 years follow-up and then discussed the results.**

   We started the Kaplan Meier survival analyses with a 10-year follow-up time and all of the analyses in Figure 3 were significant. After that we produced the same analyses with a longer 20-year follow-up time and all the other expression analyses, except hepsin analysis, remained significant. That is why we wanted to use the survival results with a 20-year follow-up and decided to present the hepsin results with a restricted 10-year follow-up time in the initial version of the manuscript.

   We are very proud that we are able to use a follow-up time as long as 20 years in our study. It is not exactly common in cancer-survival studies to have a follow-up time as long as 20 years. In general, a 5-year follow-up time is considered sufficient.
However, as presented in the comment above it appears to be too confusing to use the figure with merely 10-year follow-up time. That is why, to improve the consistency of the manuscript, we have replaced the figure with a 20-year follow-up time and included $P$ values with both 10-year and 20-year follow-up time. We have also re-formed the sentence related to this as follows:

**P12, line 277:** Similarly, low protein expression of hepsin (log rank, $P = 0.035$, Fig. 3C) predicted poorer breast cancer-specific survival during the 10-year follow-up period, yet not anymore during the 20-year follow-up period ($P = 0.315$, Fig. 3C).

In Kaplan-Meier survival analyses in Figure 3 we decided to present all the results obtained from the three different analysis methods to show the power of our survival results. However, to improve the manuscript we have decided to present merely the results from log rank test. Thank you very much for bringing up these points.

9. *Data presented in Figure 4 are redundant and do not add any informative data compared to what presented in Table 2. This figure should be removed.*

Thank you for this comment yet we wish to have Figure 4 as a part of the manuscript as we feel that graphical results are more illustrative especially regarding the legibility of the manuscript. In this context visual information has more power than presenting merely $P$ values or hazard ratios.

10. *The discussion is too long and unfocused. The authors demonstrated an association between TTSPs expression and breast cancer prognosis, without any data regarding the biological role of these TTSPs in human breast cancer, thus the discussion (from line 419) regarding their pro-metastatic activity is confusing and useless. The part regarding matriepase-2 is useless too. The authors need to focus the discussion on results presented in this study and discuss instead why the low levels of these two TTSPs are associated with poor prognosis even if their biological role suggested their pro-metastatic activity.*
Thank you very much this valuable comment. We think that one of the most important functions of discussion is to reflect the meaning of the new results to the previously published results. The discussion section of our manuscript begins with a short summary of our main findings. After that we have compared our results to previous studies. We think that the structure of the discussion is well-formed.

Even though in this study we don’t have any data to present regarding the biological role of the studied proteases, we wanted to make a synthesis based on the previously published studies although many of them are other than breast cancer-related. We have now shortened the beginning of the section as follows:

P19, line 440: Based on their proteolytic activity at the cell surface, TTSPs could contribute to tumor progression by affecting initiation of the metastatic process in primary breast cancer tumors.

We also removed the sentence regarding HAI-1 and HAI-2 inhibiting hepsin activity as well as a sentence introducing a plasmin-mediated proteolytic pathway. We re-formed a sentence regarding pro-uPA and uPA as follows:

P19, line 447: In addition, hepsin converts potently pro-uPA into active uPA, which initiates the degradation of ECM by cleaving plasminogen into plasmin.

We have removed the section regarding matriptase-2 according to the comment. In the initial version of the manuscript we wanted to mention the matriptase-2 study because the results were similar to ours. However, based on your comment we agree that it is better to remove it from the manuscript to clarify our discussion. We have also re-formed quite vastly the fourth section of discussion. We moved the last sentence of the section from the initial manuscript to the beginning of the section:

P18, line 412: In this study we have shown that altered TMPRSS1 and TMPRSS3 expressions are associated with the occurrence of relapses and that low TMPRSS3 mRNA and protein expression are independent factors affecting distant metastasis occurrence.

11. Limitations of the work need to be included in the discussion session.
We have now included a new section to discuss the limitations of our study:

**P20, line 460:** The limitations of our study include that TMPRSS3 expression in cancer has not nearly been studied as extensively as the expression of TMPRSS1. More work needs to be done to study the biological role of TMPRSS3 in cancer. Nonetheless, our study presents in a coherent clinical breast cancer sample set that TMPRSS3 is a credible prognostic biomarker. In contrast to our results, a previous study presented that hepsin overexpression was associated with positive nodal status and tumor stage in breast cancer. However, no survival analyses were done in that study and the analysis methods were different.

To highlight our results the following section continues with discussion regarding validation of TMPRSS1 and TMPRSS3 as prognostic biomarkers. Please see our answer for your comment number 13 for more details.

12. Based on data the authors previously published in PLoS One, and on the evidence that mRNA expression levels of these two molecules resulted more prognostic than protein levels, it would be interesting to add in this paper information regarding the association between TTSPs expression and SNPs in TMPRSS1 and TMPRSS3 and association with breast cancer prognosis.

We thank the reviewer for exploring also the previous studies of our group. The next step in our TTSP study in breast cancer is precisely to analyze the associations between our data from the published and ongoing SNP studies and expression studies. However, the amount of data is extremely vast so we had to exclude those analyses from the current study.

13. Validation of TMPRSS1 and TMPRSS3 mRNA expression levels as prognostic marker in public gene expression datasets should be included.

This is an excellent comment indeed. We used an online Kaplan-Meier survival analysis tool to validate the value of TMPRSS1 and TMPRSS3 as prognostic biomarkers in breast cancer. The Kaplan-Meier plotter uses gene expression data
and relapse-free and overall survival information which are downloaded from GEO (Affymetrix microarrays only), EGA and TCGA. The patient samples are divided into two groups according to the median gene expression value similar to our analysis method. Then the two groups are compared by Kaplan-Meier plot and the hazard ratio with 95% confidence intervals and log rank $P$ values are calculated. The databases are updated biannually. Both low $TMPRSS1$ and $TMPRSS3$ expression were significantly associated with poorer relapse-free survival (RFS, $n = 3554$), overall survival (OS, $n = 1117$), and distant metastasis-free survival (DMFS, $n = 1609$). The results were exactly similar to ours and the figures obtained from the Kaplan-Meier analysis tool are presented below. (Source used: http://kmplot.com/analysis/index.php?p=background; Reference: Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients, Breast Cancer Res Treatment, 2010 Oct;123(3):725-31).

$TMPRSS1$ RFS ($n = 3554$)  

$TMPRSS3$ RFS ($n = 3554$)
We have also discussed the validation results in the revised manuscript:

P20, line 466: In addition, to validate our results and the prognostic value of hepsin and TMPRSS3 in a large clinical breast cancer microarray database, we used an online Kaplan-Meier survival analysis tool. Similar to our study,
in these analyses the cohorts were divided into two groups according to the median expression of TMPRSS1 and TMPRSS3. Based on the survival curves displayed and the logrank P values both low TMPRSS1 and TMPRSS3 expression was significantly associated with poorer relapse-free survival, overall survival, and distant metastasis-free survival. To sum up, the survival trend was exactly alike compared to our results.

Reviewer #2, Brian F. Clem:

Major compulsory revisions:

1. In paragraph 1, line 5, within the results section, the authors state that both TMPRSS1 and TMPRSS3 mRNA expression is higher in malignant breast tumors compared to the benign tumors. However, while the data for TMPRSS1 is confirmed within the statistical analysis in supplementary table S1, TMPRSS3 levels show no statistical difference and even a lower median expression within malignant tumors compared to benign. Within paragraph 1 of the discussion section, the authors narrow this description to well-differentiated tumors. In order to make this conclusion for TMPRSS3, statistical analysis should be performed between benign and the samples that the authors are describing as “well-differentiated” malignant tumors.

First of all, we would like to thank you very much for your valuable comments as by making the following revisions the legibility and consistency of our manuscript are clearly improved. We have re-formed the sentence mentioned above in the results section as follows:

P11, line 247: TMPRSS1 and TMPRSS3 mRNA expression was high in well-differentiated malignant breast tumors compared to benign breast tumors (Table 2; Supplementary Fig. S1A-B).

Based on your excellent comment we have also removed the initial Supplementary Figure S1 and in the revised version we have merged it with the initial Figure 2. As
a result we have a new figure (Supplementary Figure S1) where we included an expression bar of benign tumors next to the ones grouped by tumor grade. The figure presents clearly the differences for example between benign and grade I tumors as well as between different grades. In Table 2 we have also included the significant $P$ values for comparing mRNA expression in benign tumors versus grade I tumors. With these revisions the results section and the discussion are now in accordance regarding the expression level comparisons between benign and malignant tumors.

2. From the published literature, there appears to be a discrepancy between levels of these TTSPs and their association with cancer risk among various tumor types with some reports associating high expression with advanced disease while others demonstrate lower expression correlating with lower survival. The authors cite a published report by Xing et.al. (2011) as a reference for increased expression of Hepsin in breast cancer. However, they did not discuss the clinical portion of that study, which demonstrated a significant association between high Hepsin expression and higher tumor stage as well as lymph node metastasis. This is in direct contrast with the data presented within this manuscript. At the least, the authors need to acknowledge this difference and discuss them in light of the present findings.

Our results are not similar to the ones presented in the previous study but it, to our opinion, does not flatten the value of our findings. We were also able to validate our survival results in a large clinical microarray database and our mRNA expression results correlated with the protein expression results. This is a relevant comment indeed and we have discussed the matter in the revised manuscript:

P20, line 464: In contrast to our results, a previous study presented that hepsin overexpression was associated with positive nodal status and tumor stage in breast cancer. However, no survival analyses were done in that study and the analysis methods were different. In addition, to validate our results and the prognostic value of hepsin and TMPRSS3 in a large clinical breast cancer microarray database, we used an online Kaplan-Meier survival analysis tool. Similar to our study, in these analyses the cohorts
were divided into two groups according to the median expression of TMPRSS1 and TMPRSS3. Based on the survival curves displayed and the logrank P values both low TMPRSS1 and TMPRSS3 expression were significantly associated with poorer relapse-free survival, overall survival, and distant metastasis-free survival. To sum up, the survival trend was exactly alike compared to our results.

As mentioned above, to validate our results we used an online Kaplan-Meier survival analysis tool and please see our answer also for the other reviewers’ comment number 13 for more information.

In the prior study (reference 14) the authors have analyzed the total immunoreactivity of the cells. In our study, the immunoreactivity of hepsin and TMPRSS3 was analyzed in the cytoplasm of epithelial tumor cells, and the intensity and the extent of staining were scored (0, negative; 1, weak; 2, moderate; 3, high). Then we divided the tumors into two groups according to the median value of the scores. In the previous study the tumors were divided into two groups differently (low: less than 30% of the cells were immunoreactive; high: 30% to greater than 70% of the cells were immunoreactive). That is why it is challenging to compare their results to ours.

In the same previous study in Table 1 one might think at first that Hepsin-L stands for low hepsin expression and Hepsin-H for high expression. However, in the immunohistochemistry section of materials and methods, it is said that that Hepsin-H is for scores 0 to 1 and Hepsin-L is for scores 2 to 3. All in all, we felt somewhat confused about interpreting the results of that study. In the previous study a rabbit polyclonal antibody against hepsin was also used, however in that study it was from a different manufacturer (Santa Cruz). That is one reason why one shouldn’t compare our results to theirs too explicitly. This is an example of a typical challenge in an IHC study: How detailed comparisons between two different studies are possible to produce if the methods and reagents differ from each other?
3. Notwithstanding their data demonstrating the association of low expression with decreased survival, the authors introduce TTSPs and their activity as a cellular mechanism to promote tumor progression and invasion and then conclude their findings by suggesting that Hepsin and TMPRSS3 have potential as therapeutic targets in breast cancer. In light of the observed alterations of these proteins during breast cancer progression, the authors should more adequately address the potential mechanism for therapeutic intervention. There are new reports (Han, ACS Med. Chem. Lett. 2014, 5, 1219) characterizing inhibitors of TTSPs, including Hepsin activity, and their potential as anti-cancer agents. Would these data suggest that these may be detrimental to breast cancer patients, especially in terms of treatment as it relates to disease stage of the patient? In addition, previous reports have demonstrated that endogenous inhibitors of TTSPs, HAI-1 and HAI-2, are down-regulated in advanced breast cancer specimens (Clin. Cancer Res., 10:202). It is not intuitive, why both target and endogenous inhibitor would be decreased during breast cancer progression.

Thank you very much for this comment and it is true that we are not able to address the exact potential therapeutic intervention mechanisms for hepsin and TMPRSS3 in the light of our results. Based on your comment we have decided to present the studied TTSPs particularly as potential biomarkers and not as therapeutic targets in the revised manuscript.

Minor essential revisions:

1. The axis for Supplementary Figure S2D needs to be changed in order to be consistent with the other figures within the manuscript.

Thank you for pointing this out. We have now changed the axis for the figure.

The editorial office:
1. Please reformat the title page. The title page should: provide the title of the article, list the full names, institutional addresses and email addresses for all authors and indicate the corresponding author. Please note: the title should include the study design, for example "A versus B in the treatment of C: a randomized controlled trial X is a risk factor for Y: a case control study". Abbreviations within the title should be avoided.

We have re-formed the title as follows: “Low expression levels of hepsin and TMPRSS3 are associated with poor breast cancer survival”. We think that the re-formed title is more powerful regarding our findings. The title page is also re-formed as requested. We removed the running title section from the title page and re-formed indication of the corresponding author.

2. Financial support section in the title page should be inserted in the acknowledgment section of the manuscript.

Thank you for pointing this out. We have inserted the financial support section in the acknowledgment section in the revised version of the manuscript.

We are looking forward to hear about your response soon,

Sincerely Yours,

Arto Mannermaa, Ph.D.  Mikko Pelkonen, B.M.
Associate Professor, Research Director  Doctoral Student

Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland
P.O. Box 1627, FI-70211 Kuopio, Finland

E-mail: arto.mannermaa@uef.fi

Phone: +358-40-3552752

Fax: +358-17-162753