THE EMBO JOURNAL

Manuscript EMBO-2012-81456

The Lgr5 Intestinal Stem Cell Signature: Robust Expression of Proposed Quiescent '+4' Cell Markers

Javier Muñoz, Daniel E. Stange, Arnout G. Schepers, Marc van de Wetering, Bon-Kyoung Koo, Shalev Itzkovitz, Richard Volckmann, Kevin S. Kung, Jan Koster, Sorina Radulescu, Kevin Myant, Rogier Versteeg, Owen J. Sansom, Johan H. van Es, Nick Barker, Alexander van Oudenaarden, Shabaz Mohammed, Albert J. R. Heck and Hans Clevers

Corresponding author: Hans C. Clevers, Hubrecht Institute

Review timeline:

Submission date:	20 March 2012
Editorial Decision:	25 April 2012
Revision received:	13 May 2012
Accepted:	15 May 2012

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 April 2012

Thank you for submitting your study for consideration to The EMBO Journal editorial office.

Please accept my apologies for the slight delay in getting a decision to you, based on the fact that one of the scientists that originally agreed did neither provide a report nor responded to numerous chasers. Having two rather consistent expert views at hand, I am now in a position to reach a final conclusion.

As you will see from the enclosed comments, both appreciate the quality and general interest resulting from your study, though raise some critical points that I ask you to address in a thoroughly revised manuscript. As the comments are concrete and explicit, there is no point to repeat these here. I hope that these facilitate efficient and appropriate revision of your dataset.

Please do not hesitate to get in touch in case of further questions (preferably via E-mail).

I do have to formerly remind you that The EMBO Journal considers only one round of major revisions with the ultimate decision solely depend on content and strength of the final manuscript.

Yours sincerely,

Editor
The EMBO Journal

REFEREE REPORTS:

Referee #1:

Overall, the manuscript by MuÒoz et al contains an exciting collection of data regarding the expression profile of intestinal stem cells analyzed on mRNA and protein level. It challenges recently published result on the nature and marker expression of the so called +4 cell and will have a major impact on the field. I have only few comments which may help to further improve the paper:

Major points

The data in the original Takeda paper with regards to the localization of Hopx look actually very convincing and it surprises that this cannot be confirmed in the current study. There is an antibody available against Hopx which should be used to further support the current claims and to exclude that differences in post-transcriptional regulation are causing the observed differences in expression pattern.

The mass spec data which were not confirmed by the mRNA expression profiling may indicate candidate stem cell markers which are mainly regulated by post-transcriptional mechanisms as the authors have pointed out. Since they make out ~20% of all markers identified by mass spec, it appears justified to validate at least some of these by alternative methods (immunohistochemistry etc.).

In Fig 4a, only one of the 17 identified, not stem cell restricted markers, which had been validated by in situ hybridization, is shown. Please include the remaining 16 in the supplement so that the reader can directly compare the overlap between profiling results and the actual expression patterns.

Takeda et al 2011 has performed similar analysis for Bmi1 and this paper should be cited not only for Hopx but also for this part.

Minor points
The original Capecchi paper on Bmi1 analyzes mainly the first 10cm of the small intestine. In the current manuscript it was not stated whether a similar region was used for analysis. This information should be added.

The title is not very informative with regards to the results on the +4 cell. Please change to better express the main findings of the paper.

Referee #2:

"The Lgr5 Intestinal Stem Cell Signature and its overlap with Markers of +4 cells"

In this comprehensive study, Clevers and colleagues explore the identity of Lgr5+ve cells and their descendants by transcriptomics and proteomics. The first part of the study provides a wider view of the genes and proteins that characterize crypt base columnar cells (CBCs), which constitute a bona-fide population of intestinal stem cells (ISCs). However, whether CBCs are the only ISC population in the crypt is a matter of intense debate. Over the last couple of years, other putative ISC types have been described. These accessory ISCs can be labeled with specific marker genes such as Bmi1, Hopx, Lrig1 and Tert. Apparently, they display different properties than those originally reported for Lgr5+ve ISCs including slower kinetics of cell cycle and distinct localization within the crypt axis. Thus, the second part of the manuscript is devoted to analyze the expression pattern of marker genes of these additional ISC populations. Authors provide strong evidence that Bmi1, Hopx, Lrig1 and Tert are not expressed in discrete cell populations around +4 crypt cell position but rather they show extended expression domains throughout the crypt which in some cases overlap to that of Lgr5. For the case of Bmi1, authors include tracing experiments demonstrating that the Bmi1 tracing allele previously reported (Capecchi and colleagues) does not reproduce the expression of the endogenous locus and it traces randomly from different crypt positions. Altogether, these results cast doubts about the true identity of these additional ISC populations. I found the manuscript compelling and of great interest to the stem cell community as a whole. The first part of the work expands our

knowledge on the biology of ISCs whereas the second part helps to clarify the discussion about the true nature of ISCs. I only have a few minor comments:

1. Authors state that MS analysis did not capture Lgr5 because this protein is present in particular membrane compartments. However, authors are capable of identifying many membrane proteins in their analysis. This part would be more convincing if authors could show in which particular subcellular compartment resides Lgr5 and explain why their extracts do not contain this cell fraction.

2. Authors identify ISC enriched genes with either Affymetrix or Agilent arrays, a number of which did not score positive in both platforms. Are those genes false positives or do they simply represent probes exclusive of only one of the two platforms? Perhaps authors could verify the expression of those genes by qPCR.

3. Authors comment, "Nevertheless, some genes were found enriched at the mRNA level, but not at the protein level and vice versa. For 28 genes from the "mRNA stem cell signature", no enrichment at the protein level was detected. As proteins are the main mediators of biological functions, these genes are unlikely to play a specific biological role in stem cells and were therefore subtracted from the signature". Would it be possible that no protein is detected because they are low abundant (as Bmi1 or Tert) or simply because they are not recovered in the extracts (as Lgr5). Authors should discuss this issue and reformulate their statement.

4. Authors show that Bmi1 protein is not expressed in a restricted cell population at +4 by IHC. This is an important finding and I suggest to move these data from the supplementary material to figure 8.

5. The analysis of stem cell marker genes in different Lgr5 populations shown in figure 6 suggests that Hopx, Lrig, Tert and Bmi1 have widespread expression patterns in the crypts. Data shown are extracted from microarray analysis. However, as authors report here, different platforms give rise to different measurements in differential gene expression depending on the technology and selected probes and it is also well established that microarray normalization algorithms flatten gene expression. Therefore, authors should strengthen Figure 6 by measuring +4 ISC maker gene expression by qPCR in the same set of samples.

6.Previous works showed that Tert (Montgomery et al. PNAS 2011) and Lrig1 (Powel et al. Cell. 2012) expression is enriched in individual cells disperse within different crypts. These data are in sharp contrast with authors' observations (Figure 6 and 7), which report homogenous mRNA expression throughout several crypt cell positions. However, would it be possible that authors missed individual cells with high levels of Tert or Lrig1 in given crypts? How many crypts have they observed in the ISH analysis? Authors must discuss this possibility in the revised version.

1st Revision - Authors' Response 13 May 2012

We have now addressed the points raised by the reviewers. New data include an update on the overlap between array platforms (based on new releases of the Annotation Files) and Mass Spec, immunohistochemistry for Hopx confirming the broad expression of this protein at the crypt base including the stem cell/Paneth cell zone, and qPCR to confirm the array data.

I hope that The EMBO Journal will now be able to publish our study

Referee #1:

Overall, the manuscript by Munoz et al contains an exciting collection of data regarding the expression profile of intestinal stem cells analyzed on mRNA and protein level. It challenges recently published result on the nature and marker expression of the so called +4 cell and will have a major impact on the field. I have only few comments which may help to further improve the paper:

Major points

1. The data in the original Takeda paper with regards to the localization of Hopx look actually very convincing and it surprises that this cannot be confirmed in the current study. There is an antibody available against Hopx which should be used to further support the current claims and to exclude that differences in post-transcriptional regulation are causing the observed differences in expression pattern.

Let us first say that our in situ hybridization data (figs) match very well with the array and mass spec data. As suggested by the reviewer, we have performed immunohistochemistry for Hopx. Nuclear staining was detected in a pattern that matches the mRNA expression pattern. This staining is now shown in Figure 7 together with Bmi1 and Olfm4.

2. The mass spec data which were not confirmed by the mRNA expression profiling may indicate candidate stem cell markers which are mainly regulated by post-transcriptional mechanisms as the authors have pointed out. Since they make out ~20% of all markers identified by mass spec, it appears justified to validate at least some of these by alternative methods (immunohistochemistry etc.).

A very useful remark. We examined the overlap again in detail. We realized that both Agilent as well as Affymetrix have released new updates of their microarray annotation files. We therefore re-analyzed the overlap between the two array platforms and MassSpec (Figure 1D and 2B). This increased the number of genes within the overlap significantly. All Supplementary Tables have been updated.

Concerning the 59 proteins, we examined now in detail the mRNA expression levels and found that the vast majority of them (78% (46/59)) were enriched in stem cells, but not high enough to be included in the signature. This analysis is added now to the text (page 6, paragraph 2, line 16). The mRNA expression data of the 59 proteins can now be found in Supplementary Table S11. These genes originally seemed to be enriched only at the protein level, indicating post-transcriptional modifications, but there exclusion was mainly caused by the necessity of using thresholds for enrichment. We would therefore like to take back our initial statement and now state that post-transcriptional regulation does “not appear to represent a major mechanism regulating protein levels in intestinal stem cell related genes”. There might be post-translational modifications which lead to a further enrichment of these proteins compared to their mRNA levels in stem cells, but we feel that this does not need to be addressed in this manuscript.

3. In Fig 4a, only one of the 17 identified, not stem cell restricted markers, which had been validated by in situ hybridization, is shown. Please include the remaining 16 in the supplement so that the reader can directly compare the overlap between profiling results and the actual expression patterns.

We could detect a specific in-situ signal in 17 genes, a gradient in nine genes and a specific expression at the crypt bottom in eight genes (page 7, paragraph 1, line 5). A new supplementary figure on the nine genes with gradient has been included (S6). It shows that these genes are indeed expressed at the crypt bottom in a gradient that extends to varying degrees above the stem cell/Paneth cell zone.

4. Takeda et al 2011 has performed similar analysis for Bmi1 and this paper should be cited not only for Hopx but also for this part.

Thanks for this hint. We have included now a reference to the Bmi1 IHC (page 10, paragraph 2, line 11) shown by Takeda et al (Supporting Figure 7 in their publication). This IHC indeed nicely illustrates the extended expression domain of Bmi1, which matches to the starting position of LacZ tracings as documented by us and –independently- Tian et al (Figure 4j in their publication, also referenced). We would like to note here that neither Takeda nor Tian et al explicitly mention these observations on broad Bmi1 expression in the respective papers.

Minor points

The original Capecchi paper on Bmi1 analyzes mainly the first 10cm of the small intestine. In the current manuscript it was not stated whether a similar region was used for analysis. This information should be added.

We have also analyzed only the first 10cm of the small intestine and state so now in the text (page 8, last paragraph, line 6).

The title is not very informative with regards to the results on the +4 cell. Please change to better express the main findings of the paper.

We have changed the title of the manuscript which now stresses the point that Lgr5 cells robustly express markers of proposed quiescent/+4 cells.

Referee #2:

"The Lgr5 Intestinal Stem Cell Signature and its overlap with Markers of +4 cells"

In this comprehensive study, Clevers and colleagues explore the identity of Lgr5+ve cells and their descendants by transcriptomics and proteomics. The first part of the study provides a wider view of the genes and proteins that characterize crypt base columnar cells (CBCs), which constitute a bona-fide population of intestinal stem cells (ISCs). However, whether CBCs are the only ISC population in the crypt is a matter of intense debate. Over the last couple of years, other putative ISC types have been described. These accessory ISCs can be labeled with specific marker genes such as Bmi1, Hopx, Lrig1 and Tert. Apparently, they display different properties than those originally reported for Lgr5+ve ISCs including slower kinetics of cell cycle and distinct localization within the crypt axis. Thus, the second part of the manuscript is devoted to analyze the expression pattern of marker genes of these additional ISC populations. Authors provide strong evidence that Bmi1, Hopx, Lrig1 and Tert are not expressed in discrete cell populations around +4 crypt cell position but rather they show extended expression domains throughout the crypt which in some cases overlap to that of Lgr5. For the case of Bmi1, authors include tracing experiments demonstrating that the Bmi1 tracing allele previously reported (Capecchi and colleagues) does not reproduce the expression of the endogenous locus and it traces randomly from different crypt positions. Altogether, these results cast doubts about the true identity of these additional ISC populations. I found the manuscript compelling and of great interest to the stem cell community as a whole. The first part of the work

expands our knowledge on the biology of ISCs whereas the second part helps to clarify the discussion about the true nature of ISCs. I only have a few minor comments:

1. Authors state that MS analysis did not capture Lgr5 because this protein is present in particular membrane compartments. However, authors are capable of identifying many membrane proteins in their analysis. This part would be more convincing if authors could show in which particular subcellular compartment resides Lgr5 and explain why their extracts do not contain this cell fraction.

Plasma membrane proteins are typically characterized by the presence of hydrophobic regions which are embedded in the lipid bilayer. Such regions make the identification of plasma membrane proteins by typical MS workflows quite challenging due to their inaccessibility for proteolytic enzymes and the insolubility of their transmembrane regions. Nevertheless, as pointed out by the reviewer, our data contains a high number of proteins with predicted trans-membrane domains which would indicate only a small bias in the detection limit towards these proteins in our analyses. However, Lgr5 encodes a 7-transmembrane receptor which is expressed in the plasma membrane compartment at low levels (van der Flier *et al.* 2009). The low expression and the fact that Lgr5 is a 7TM protein both probably contribute to the absence of detection of Lgr5 in our MS survey. Indeed, Gene Ontology analyses on the 7,967 proteins identified in our data set showed a clear under-representation for 7TM proteins (G-protein coupled receptors; p=8.4E-54) supporting this notion. This point has now been added to the manuscript (page 5, paragraph 2, line 15).

2. Authors identify ISC enriched genes with either Affymetrix or Agilent arrays, a number of which did not score positive in both platforms. Are those genes false positives or do they simply represent probes exclusive of only one of the two platforms? Perhaps authors could verify the expression of those genes by qPCR.

A helpful comment. As mentioned above (Referee 1, point 2), we noticed while looking at the non-overlapping genes, that both Agilent as well as Affymetrix have updated their annotation files since our last analysis and therefore re-analyzed the overlap data, which lead to changed numbers (Figure 1D, 2B).

Concerning the non-overlapping genes form the two array platforms, the majority of these genes, 72% (51/71) and 57% (31/54), do show a below-threshold enrichment in the other platform. This analysis is now described in the text (page 5, paragraph 1, line 15) and details are given in Supplementary Table S4 and S5. As for the 59 proteins that originally seemed to be uniquely enriched at the protein level, most of these genes only seem to be unique to one platform. The enrichment was in many cases just below the set threshold. The new Supplementary Tables make it easy to look up even genes which are found by only one platform. We therefore feel that performing qPCRs for a few of these non-overlapping genes is less helpful to the reader then the possibility that Table S4 and S5 offer for the analysis of the behavior of individual candidate genes in each of the two platforms.

3. Authors comment, "Nevertheless, some genes were found enriched at the mRNA level, but not at the protein level and vice versa. For 28 genes from the "mRNA stem cell signature", no enrichment at the protein level was detected. As proteins are the main mediators of biological functions, these genes are unlikely to play a specific biological role in stem cells and were therefore subtracted from the signature". Would it be possible that no protein is detected because they are low abundant (as Bmi1 or Tert) or simply because they are not recovered in the extracts (as Lgr5). Authors should discuss this issue and reformulate their statement.

We have clarified this point now in the text (page 6, paragraph 2, line 11). The protein product for the 28 genes only found in the "mRNA signature" was in fact detected by MS. However, no enrichment was found at the protein level in the Lgr5+ve stem cells (Supplementary Table S6). Furthermore, a new supplementary table has been added (S10) which gives details of these genes in the three individual analyses.

4. Authors show that Bmi1 protein is not expressed in a restricted cell population at +4 by IHC. This is an important finding and I suggest to move these data from the supplementary material to figure 8.

Agreed. Together with new IHC data for Hopx, the Bmi1 IHC has been moved to the main text (Figure 7).

5. The analysis of stem cell marker genes in different Lgr5 populations shown in figure 6 suggests that Hopx, Lrig, Tert and Bmi1 have widespread expression patterns in the crypts. Data shown are extracted from microarray analysis. However, as authors report here, different platforms give rise to different measurements in differential gene expression depending on the technology and selected probes and it is also well established that microarray normalization algorithms flatten gene expression. Therefore, authors should strengthen Figure 6 by measuring +4 ISC maker gene expression by qPCR in the same set of samples.

We have now performed qPCRs for the four genes on cDNA isolated from an independent sorting experiment (Figure 5). In most cases the ratio differences between the 5 fractions were, as predicted by the reviewer, higher than the array ratios.

6. Previous work showed that Tert (Montgomery et al. PNAS 2011) and Lrig1 (Powel et al. Cell. 2012) expression is enriched in individual cells dispersed within different crypts. These data are in sharp contrast with authors' observations (Figure 6 and 7), which report homogenous mRNA expression throughout several crypt cell positions. However, would it be possible that authors missed individual cells with high levels of Tert or Lrig1 in given crypts? How many crypts have they observed in the ISH analysis? Authors must discuss this possibility in the revised version.

Our immunohistochemical analyses on Bmi1 and HopX do not show such rare high expressors. We have now substantially increased the number of crypts analyzed by single molecule mRNA ISH to at least 30 for each gene. The numbers of analyzed crypts per gene are given in the Material and Method section for the single molecule in situs. Of note, we have observed a high level of homogeneity between the analyzed crypts, making it unlikely that we missed a specific enrichment in individual cells in a subset of crypts. This point has been added to the text now (page 8, paragraph 3, line 7).

