

## O-glycan initiation directs distinct biological pathways and controls epithelial differentiation

Ieva Bagdonaite, Emil M. H. Pallesen, Zilu Ye, Sergey Y. Vakhrushev, Irina N. Marinova, Mathias I. Nielsen, Signe H. Kramer, Stine F. Pedersen, Hiren J. Joshi, Eric P. Bennett, Sally Dabelsteen, and Hans H. Wandall

---

### Review timeline:

Submission date:	16 July 2019
Editorial Decision:	27 August 2019
Revision received:	6 December 2019
Editorial Decision:	14 February 2020
Revision received:	3 March 2020
Accepted:	16 March 2020

---

Editor: Martina Rembold

### 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

27 August 2019

---

Thank you for the submission of your research manuscript to EMBO Reports. We have now received the full set of referee reports that is copied below.

As you will see, the referees acknowledge that the findings are potentially interesting. However, referees 1 and 2 also point out several technical concerns and have a number of suggestions for how the study should be strengthened, and I think that all of them should be addressed.

Given these constructive comments, we would like to invite you to revise your manuscript with the understanding that the referee concerns (as detailed above and in their reports) must be fully addressed and their suggestions taken on board. Please address all referee concerns in a complete point-by-point response. Acceptance of the 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 your responses included in the next, final version of the manuscript.

Revised manuscripts should be submitted within three months of a request for revision; they will otherwise be treated as new submissions. Please contact us if a 3-months time frame is not sufficient for the revisions so that we can discuss the revisions further.

When submitting your revised manuscript, please carefully review the instructions that follow below. Failure to include requested items will delay the evaluation of your revision.

1) a .docx formatted version of the manuscript text (including legends for main figures, EV figures and tables). Please make sure that the changes are highlighted to be clearly visible.

2) individual production quality figure files as .eps, .tif, .jpg (one file per figure).

Please download our Figure Preparation Guidelines (figure preparation pdf) from our Author Guidelines pages  
<https://www.embopress.org/page/journal/14693178/authorguide> for more info on how to prepare your figures.

3) a .docx formatted letter INCLUDING the reviewers' reports and your detailed point-by-point responses to their comments. As part of the EMBO Press transparent editorial process, the point-by-point response is part of the Review Process File (RPF), which will be published alongside your paper.

4) a complete author checklist, which you can download from our author guidelines (<<https://www.embopress.org/page/journal/14693178/authorguide>>). Please insert information in the checklist that is also reflected in the manuscript. The completed author checklist will also be part of the RPF.

5) Please note that all corresponding authors are required to supply an ORCID ID for their name upon submission of a revised manuscript (<<https://orcid.org/>>). Please find instructions on how to link your ORCID ID to your account in our manuscript tracking system in our Author guidelines (<<https://www.embopress.org/page/journal/14693178/authorguide#authorshipguidelines>>)

6) The Supplementary information is in the correct format (Figure EVx, Dataset EVx, Appendix) but I suggest to add a title to the first page of the Appendix.

7) Before submitting your revision, primary datasets produced in this study need to be deposited in an appropriate public database (see <<https://www.embopress.org/page/journal/14693178/authorguide#datadeposition>>).

Specifically, we would kindly ask you to provide public access to the following datasets:

- RNA seq analysis
- Glyco- and Phosphoproteomics

Please remember to provide a reviewer password if the datasets are not yet public.

The accession numbers and database should be listed in a formal "Data Availability " section (placed after Materials & Method) that follows the model below (see also <<https://www.embopress.org/page/journal/14693178/authorguide#datadeposition>>). Please note that the Data Availability Section is restricted to new primary data that are part of this study.

# Data availability

The datasets (and computer code) produced in this study are available in the following databases:

- RNA-Seq data: Gene Expression Omnibus GSE46843  
 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46843>)
- [data type]: [name of the resource] [accession number/identifier/doi] ([URL or identifiers.org/DATABASE:ACCESSION])

\*\*\* Note - All links should resolve to a page where the data can be accessed. \*\*\*

8) Our journal encourages inclusion of \*data citations in the reference list\* to directly cite datasets that were re-used and obtained from public databases. Data citations in the article text are distinct from normal bibliographical citations and should directly link to the database records from which the data can be accessed. In the main text, data citations are formatted as follows: "Data ref: Smith et al, 2001" or "Data ref: NCBI Sequence Read Archive PRJNA342805, 2017". In the Reference list, data citations must be labeled with "[DATASET]". A data reference must provide the database name, accession number/identifiers and a resolvable link to the landing page from which the data can be accessed at the end of the reference. Further instructions are available at <<https://www.embopress.org/page/journal/14693178/authorguide#referencesformat>>.

9) Regarding data quantification:

- Please ensure to specify the name of the statistical test used to generate error bars and P values, the number (n) of independent experiments underlying each data point (not replicate measures of one sample), and the test used to calculate p-values in each figure legend. Discussion of statistical methodology can be reported in the materials and methods section, but figure legends should contain a basic description of n, P and the test applied.

IMPORTANT: Please note that error bars and statistical comparisons may only be applied to data obtained from at least three independent biological replicates. If the data rely on a smaller number of replicates, scatter blots showing individual data points are recommended.

- Graphs must include a description of the bars and the error bars (s.d., s.e.m.).

- Please also include scale bars in all microscopy images.

10) As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. This File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

You are able to opt out of this by letting the editorial office know ([emboreports@embo.org](mailto:emboreports@embo.org)). If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

We would also welcome the submission of cover suggestions, or motifs to be used by our Graphics Illustrator in designing a cover.

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:

In this manuscript the authors use the spontaneously immortalized human epidermal keratinocyte, HaCaT, to study the role of GalNAc-T's in epithelial homeostasis and differentiation. They show that GalNAc-T1, T2, and T3 are the three predominantly expressed GalNAc-T's expressed in HaCaT and primary keratinocytes. They subsequently generate knockouts of each using both ZFN nuclease and Crispr-Cas9 (GalNAc-T3 only) methods. The KO cells were induced to form a stratified squamous epithelium in culture, and each showed distinct phenotypes. To better understand the underlying pathways affected by deletion of the individual GalNAc-T's, the authors performed global transcriptomic, differential glycoproteomic, and differential phosphoproteomic analyses on the KO cells. The glycoproteomic analysis revealed the enzymes modified distinct substrates with a small overlap, and the transcriptomic and phosphoproteomic analyses revealed that different underlying pathways were being affected. Gene ontology analysis suggested that the GalNAc-T1 KO affected mainly members of the endomembrane system, GalNAc-T2 KO affected mainly cell-ECM adhesion, while GalNAc-T3 KO affected pathways involved in epithelial differentiation. Significantly, the authors validated the effect of loss of a specific GalNAc-T on the function of specific target proteins.

These are well done studies that confirm the distinct function of three GalNAc-T's in a variety of biological functions which will be of interest to a wide variety of researchers from multiple fields. A few concerns need to be addressed:

1. Figure EV3: The lysotracker staining in the T1 KO cells is very hard to see, so it is difficult to conclude from this data that there is less co-localization of pHrodo-labeled LDL and lysosomes. Is it possible that there is a defect in lysosomal pH in the T1 KO cells?
2. Figures 4L versus EV5: Why does T3 KO result in increased NICD staining in Figure 4L, but

decreased NICD staining in Figure EV5?

3. Top of page 14: The authors suggest that there is significant overlap between the glycosites identified by the two methods (dimethyl labeling and TMT) and they refer to "Fig S6", which this reviewer could not find. In contrast, Dataset EV4 (second tab) shows that only 13 glycopeptides were identified by both methods out of nearly 400 total. The authors suggest that, "This can be partly explained by potential over glycosylation in cells with truncated O-glycans", although if this were the case, more peptides would be identified with the dimethyl labeling (SC samples) than TMT, but the opposite happened with the T1 samples. Unless there is more data suggesting that the two methods are identifying more common sites (i.e. the missing Fig S6), the authors should revise this section to indicate that the methods are different and seeing different things.

Minor comments:

4. Figure 1C: Were none of the other GALNTs detected?

5. Figure 1D: The upper (skin) and lower (HaCaT organ culture) images look quite different, not only the intensity of GalNAc-T staining, but also DAPI staining. Are there better images, or can the authors explain the differences? Also, unlike the description in the text, GalNAc-T2 appears to be basal in the organ culture, but is more widespread in the skin.

6. Dataset EV1: The GALNT3 KO tab shows that levels of GALNT3 are increased. Based on the text, I assume this should be GALNT5? The authors should check for other typos in the datasets.

7. Figure EV1 legend: It would be helpful to define "sc" and "+ST" in the legend.

8. Figure 2B: The Venn diagram needs to be explained in the legend.

9. Dataset EV3: Are there any significant/interesting changes in the proteome that may implicate additional pathways altered in the KO cells?

10. Page 12, last 2 sentences of Results: "Appendix Figure S1" should be "Figure EV5".

11. Page 13, bottom: "With these techniques, we unambiguously identified 67, 88, and 63 single-GalNAc/GalGalNAc glycosites for GalNAc-T1, GalNAc-T2, and GalNAc-T3, respectively." Figure 3D suggests 49, 60, and 50, respectively.

## Referee #2:

The manuscript O-glycan initiation directs distinct biological pathways and controls epithelial differentiation provides in-depth analysis of the impact of loss of GalNAc-T1-3 on epithelial cell function and gene expression. It represents a systems biology approach to an important problem. The data are well presented. Some specific deficiencies are noted below:

1. There are no labels on the figures in the pdf making them hard to evaluate.

1. morphological phenotyping would be improved with arrows and annotations. Even a trained pathologist would have trouble identifying the features presented in figure 1 without arrows and enlargements.

2. The enrichment of protein targets in figure 3 are interesting but hard to assess:  
Are the protein abundances taken into account?  
Are the number of sites/protein considered in this analysis?  
What is the enrichment if protein domains are considered? i.e., EGF-repeats etc.

3. The quality of individual micrographs (fluorescence) is poor. The enlarged version help but without proper labelling (see above) it is hard to point out the poor ones.

Since figures are not labelled other problems may be present that I cannot properly decipher.

In summary, this is nice work but without figure legends, hard to assess.

1st Revision - authors' response

6 December 2020

## Point-by-point response

### Referee #1:

In this manuscript the authors use the spontaneously immortalized human epidermal keratinocyte, HaCaT, to study the role of GalNAc-T's in epithelial homeostasis and differentiation. They show that GalNAc-T1, T2, and T3 are the three predominantly expressed GalNAc-T's expressed in HaCaT and primary keratinocytes. They subsequently generate knockouts of each using both ZFN nuclease and Crispr-Cas9 (GalNAc-T3 only) methods. The KO cells were induced to form a stratified squamous epithelium in culture, and each showed distinct phenotypes. To better understand the underlying pathways affected by deletion of the individual GalNAc-T's, the authors performed global transcriptomic, differential glycoproteomic, and differential phosphoproteomic analyses on the KO cells. The glycoproteomic analysis revealed the enzymes modified distinct substrates with a small overlap, and the transcriptomic and phosphoproteomic analyses revealed that different underlying pathways were being affected. Gene ontology analysis suggested that the GalNAc-T1 KO affected mainly members of the endomembrane system, GalNAc-T2 KO affected mainly cell-ECM adhesion, while GalNAc-T3 KO affected pathways involved in epithelial differentiation. Significantly, the authors validated the effect of loss of a specific GalNAc-T on the function of specific target proteins.

These are well done studies that confirm the distinct function of three GalNAc-T's in a variety of biological functions which will be of interest to a wide variety of researchers from multiple fields. A few concerns need to be addressed:

Query 1. Figure EV3: The lysotracker staining in the T1 KO cells is very hard to see, so it is difficult to conclude from this data that there is less co-localization of pHrodo-labeled LDL and lysosomes. Is it possible that there is a defect in lysosomal pH in the T1 KO cells?

Response 1. We agree that the illustrations could be optimized, as the laser intensity has been set not to oversaturate the brightest images, making LysoTracker poorly visible in GALNT1 KO. LysoTracker does, however, label small vesicles in GALNT1 KO. With that said, we should point out, that we only observed LDL localizing in a proportion of small vesicles in GALNT1 KO making them more difficult to visualize. This is in contrast to the larger vesicles observed in wild type cells.

We acknowledge that the lysosomal pH could be affected. We would, however, not expect that the decrease in intracellular pH observed in GALNT1 KO cells affects the lysosomal pH and hence the detection of LysoTracker.

Action 1. We have now increased the brightness of the red channel in Figure EV3 to enhance visibility.

Query 2. Figures 4L versus EV5: Why does T3 KO result in increased NICD staining in Figure 4L, but decreased NICD staining in Figure EV5?

Response 2. We apologize that we failed to describe these differences clearly. The text is now updated to include an improved description and discussion of the differences between Figure 4L and EV5.

There are a few major differences between the two experiments. Firstly, different keratinocyte cell lines are used to monitor NICD cleavage. In Fig 4L, non-labeled HaCaT keratinocytes are co-cultured with immortalized primary keratinocytes (N/TERT1) labeled with cell-tracker (red). As seen in these cultures, HaCaT GALNT3 KO cells do not display increased Notch1 cleavage (increase in NICD), when compared to HaCaT WT cells. In fact, the proportion of HaCaT GALNT3

KO cells undergoing Notch1 activation is somewhat lower compared to HaCaT WT (see the updated Fig 4L with boundaries between two cell types marked). This is consistent with the decreased NICD cleavage observed when HaCaT cells are grown in 3D in Fig EV5. In Fig 4L we, however, observed increased Notch1 cleavage in the N/TERT1 cells when co-cultured with HaCaT GALNT3 KO cells. This co-culture experiment is designed to present the Notch ligand on HaCaT cells and induce trans activation of Notch in N/TERT1 cells. The results suggest that absence of GalNAc-T3 in the ligand presenting cells (HaCaT) causes premature activation of Notch1 in the receiving cell (N/TERT1 cells).

Fig EV5 shows 3D differentiated epidermis created with HaCaT cells alone. In this situation, we again observe decreased Notch1 cleavage in GalNAc-T3 KO cells, but we do not see any increased activation of Notch1 as observed in the N/TERT1 cells. As mentioned in the text (page 17, line 8), we do not have a clear explanation for this, but speculate that it could be due to a GalNAc-T3 mediated effect on the balance between cis and trans-activation of Notch1. In this context, it is important to remember that human keratinocytes express multiple Notch isoforms and several different ligands, and it is therefore difficult with the current setup to assess the mechanistic details of how GalNAc-T3 affects differentiation.

Action 2. Text is now updated to include a description of the differences between Figure 4L and EV5. In addition, we marked the boundaries between HaCaT and N/TERT-1 cells in the cleaved Notch1 staining images in Fig 4L, to better see the level of Notch1 activation in the two cell types. The text on page 12 was updated to the following:

“In order to address the potency of Jagged-1<sup>T3KO</sup> to induce Notch1 signalling, we co-cultured HaCaT WT or GALNT3 KO keratinocytes with N/TERT1 immortalized keratinocytes, and assessed Notch1 activation by nuclear expression of cleaved Notch1 intracellular domain (NICD). In addition, we evaluated differentiation by cytokeratin 10 staining. Co-culture of N/TERT1 WT and HaCaT GALNT3 KO cells induced robust differentiation of N/TERT1 cells, which were mostly organized into compact piles, strongly positive for nuclear cleaved NICD, and increased keratin 10 expression (Fig 4L-N). In contrast, co-culture of WT HaCaT and N/TERT1 keratinocytes only induced limited keratin 10 expression and failed to induce organized piles of keratinocytes positive for NICD (Fig 4L-N). Furthermore, a lower level of Notch1 intracellular domain cleavage was observed in HaCaT WT, with fewer cells positive for cleaved NICD in HaCaT GALNT3 KO cells. The combined results support the potential role of GalNAc-T3-specific O-glycosylation on differentiation possibly through fine-tuning Jagged-Notch interactions. We also investigated Notch1 signaling in the organotypic models by visualizing localization of cleaved NICD, which was primarily detectable in the nuclei of suprabasal cell layers of WT HaCaT skin (Fig EV5). Compared to WT, little Notch1 signaling was observed in suprabasal layers of GALNT3 KO organotypic models (Fig EV5). This is consistent with the delayed differentiation phenotype, and likely signifies the contribution of multiple GalNAc-T3 targets to tissue development.”

Query 3. Top of page 14: The authors suggest that there is significant overlap between the glycosites identified by the two methods (dimethyl labeling and TMT) and they refer to "Fig S6", which this reviewer could not find. In contrast, Dataset EV4 (second tab) shows that only 13 glycopeptides were identified by both methods out of nearly 400 total. The authors suggest that, "This can be partly explained by potential over glycosylation in cells with truncated O-glycans", although if this were the case, more peptides would be identified with the dimethyl labeling (SC samples) than TMT, but the opposite happened with the T1 samples. Unless there is more data suggesting that the two methods are identifying more common sites (i.e. the missing Fig S6), the authors should revise this section to indicate that the methods are different and seeing different things.

Response 3. We thank the reviewer for the comment and apologize for the wrong reference – it should refer to Appendix Figure S1. We do refer to overlap between protein substrates, because the overlap at individual glycosite level is indeed very small. In fact, GalNAc-T1 is the only case where more substrates were identified using the TMT method. It is the opposite for GalNAc-T2 and -T3 where we also see around 30 % of protein substrate overlap. As the reviewer points out, it is important to acknowledge the methodological differences, including the chemical properties of the two labels differently affecting the resulting sample complexity as well as the dynamic range of detected glycopeptides, as well as the different lectin enrichment approaches.

Action 3. The text on page 14 was corrected to the following:

“We found limited overlap between the identified glycosites between each GalNAc-T isoform, supporting the validity of the strategy and identified candidates. Comparison of glycoprotein and glycosite identities for each isoform between the two methods has shown variable overlap (Appendix Fig S1, Dataset EV4). Almost a third of glycoprotein substrates for GalNAc-T2 and GalNAc-T3 identified with TMT labeling were also identified with dimethyl labeling, whereas very little overlap was observed for GalNAc-T1 (Appendix Fig S1). However, the majority of commonly identified protein substrates were glycosylated at different positions using the two methods. This is not surprising given the different genetic backgrounds of the cells used for the analysis, as well as the distinct methodological approaches affecting the chemical properties and the dynamic range of detected glycopeptides. Transcriptomic and proteomic differences from WT cells should also be taken into account.”

Minor comments:

Query 4. Figure 1C: Were none of the other GALNTs detected?

Response 4. Yes, that is correct – RPKM expression values for GALNT4, GALNT8, GALNT9, GALNT13, GALNT15, GALNT16, and GALNT19 were well below 1 RPKM. The individual RPKM values for the different GALNT genes can be found in Dataset EV1.

Query 5. Figure 1D: The upper (skin) and lower (HaCaT organ culture) images look quite different, not only the intensity of GalNAc-T staining, but also DAPI staining. Are there better images, or can the authors explain the differences? Also, unlike the description in the text, GalNAc-T2 appears to be basal in the organ culture, but is more widespread in the skin.

Response 5.

The figure has now been improved by including more directly comparable images. The differences in the submitted figure are mainly due to different imaging techniques, with skin imaged using widefield microscopy and organotypic models using confocal microscopy, as well as different magnifications. Regarding the DAPI stain, we occasionally see poor dye penetrance in frozen tissue sections, which can be improved with sucrose treatment prior to freezing.

GalNAc-T2 staining in fact seems a bit more widespread in the organ culture, although primarily located in the basal layer, where most intense staining is observed (lower panel), and more selective in skin (upper panel).

Action 5. The images of GalNAc-T stainings in skin previously obtained with a widefield fluorescent microscope have been replaced with confocal micrographs in Figure 1.

The text on page 4 has been corrected to the following:

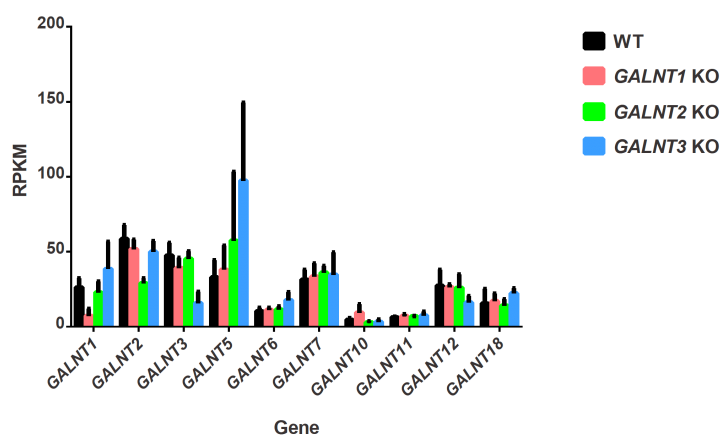
“Immunocytochemistry showed the localization of GalNAc-T1, -T2, and -T3; human skin and HaCaT 3D models expressed GalNAc-Ts in a similar expression pattern, with GalNAc-T2 primarily expressed in basal cells and broader expression of GalNAc-T1 and -T3 in all epithelial layers (Fig 1D).”

Query 6. Dataset EV1: The GALNT3 KO tab shows that levels of GALNT3 are increased. Based on the text, I assume this should be GALNT5? The authors should check for other typos in the datasets.

Response 6. The GALNT3 tab in fact shows decreased expression of GALNT3 (log(fold change) = -1.629412048). That is because the knock out procedure truncates and functionally inactivates genes of interest, however, truncated mRNA is still expressed and not completely gone. Regarding GALNT5, it does not show up as statistically significant in the analysis when all three GALNT3 KO clones are considered, therefore is not present in the respective tab. The mention in the text is based on pulling out all GALNT genes from the dataset without doing statistical analysis.

Action 6. We have included the graph reflecting increase in GALNT5 expression in the appendix. In addition, we have updated the figure legend to better describe the statistical analysis.

“Dataset EV1. RNA Seq. Gene expression data in RPKM is presented. In the “all genes normalized” tab, missing values were replaced with 0.001, and data analyzed using the CLC Genomics Workbench (Qiagen). In the individual KO tabs, statistically significant gene expression changes considering all three knock out clones of individual GALNTs (R package EdgeR analysis) are shown (filtered for at least 1 RPKM in at least one of the compared entries).  $\log_2(\text{KO}/\text{WT})$ .”



Appendix Figure S2. Expression of GALNT genes (> 1 RPKM) in GALNT KO cell lines.

Query 7. Figure EV1 legend: It would be helpful to define "sc" and "+ST" in the legend.

Action 7. The figure legend was corrected to the following:

Figure EV1. Characterization of GALNT KO cell lines. GALNT KO cell lines in WT or COSMC KO (sc, “Simple Cell”) background were stained for GalNAc-T1, GalNAc-T2, GalNAc-T3, as well as T (with (ST (sialyl-T) + T) or without (T) neuraminidase treatment), Tn, and STn (sialyl-Tn) glycoforms using monoclonal antibodies and lectins. Scale bar – 10  $\mu\text{m}$ .

Query 8. Figure 2B: The Venn diagram needs to be explained in the legend.

Action 8. The figure legend was corrected to the following:

Figure 2. Phosphoproteomic analysis. (A) Strategy for differential phosphoproteomics of HaCaT GALNT KO cell lines, where the Gaussian distributions and the Venn diagram represent selection criteria. (B) Analysis of differentially phosphorylated peptides. Upregulated or downregulated phosphopeptides in GALNT KOs were considered significant, if phosphopeptides were identified in both clones more than 2x SD away from the median, and given the same phosphopeptides were within the normal variation of two wild type biological replicates. The Venn diagram illustrates the selection criteria, where the color scheme corresponds to the Gaussian distribution illustration above. The bar graph depicts numbers of unique differential phosphosites in the individual GALNT KOs. (C) The Venn diagram represents overlap between differential phosphosites identified in the different GALNT isoform KOs. (D) GO terms (DAVID) associated with phosphoproteins identified in the 3 differential datasets. p value (Benjamini multiple comparison adjusted) of 0.05 is marked with a dashed line.

Query 9. Dataset EV3: Are there any significant/interesting changes in the proteome that may implicate additional pathways altered in the KO cells?

Response 9. We do see very limited proteome changes, if we stick with the identical analysis strategy used for the phosphoproteomic dataset. However, if we choose less strict criteria for the analysis, we see additional associations such as cellular response to external stimuli. For example, we find upregulation of proteins involved in wound healing and platelet activation in GALNT1 KO,



whereas type I interferon/antiviral response and antigen presentation were downregulated both in GALNT2 KO and GALNT3 KO. In addition, we observe some feedback responses to pathways affected by the individual GalNAc-Ts. An illustrative example is the upregulation of proteins involved in cell to matrix adhesion in GALNT2 KO that display compromised cell-matrix interactions. However, we have refrained from such assumptions in the manuscript, given the lack of statistical power and further experimental investigation.

Query 10. Page 12, last 2 sentences of Results: "Appendix Figure S1" should be "Figure EV5".

Action 10. Corrected, thank you.

Query 11. Page 13, bottom: "With these techniques, we unambiguously identified 67, 88, and 63 single- GalNAc/GalGalNAc glycosites for GalNAc-T1, GalNAc-T2, and GalNAc-T3, respectively." Figure 3D suggests 49, 60, and 50, respectively.

Response 11. The numbers mentioned in the text are correct, as are the ones in the figure. The latter numbers reflect the targets exclusively found in individual isoforms, as suggested in the Venn diagram. If the targets overlapping with those found in other isoforms are included, the numbers add up to those referred in the text.

## Referee #2:

The manuscript O-glycan initiation directs distinct biological pathways and controls epithelial differentiation provides in-depth analysis of the impact of loss of GalNAc-T1-3 on epithelial cell function and gene expression. It represents a systems biology approach to an important problem. The data are well presented. Some specific deficiencies are noted below:

Query 1. There are no labels on the figures in the pdf making them hard to evaluate.

Response 1. We apologize for the inconvenience. It was wrong of us to presume the figures would be labelled in the automatic generation of the PDF.

Query 2. Morphological phenotyping would be improved with arrows and annotations. Even a trained pathologist would have trouble identifying the features presented in figure 1 without arrows and enlargements.

Response 2. We have included annotations in Figure 1E with relevant links in the legend and the main text.

Action 2. The text on page 5 was updated to include links to marked elements in Figure 1E, and the figure legend was updated to the following:

“(E) Phenotypic characterization of organotypic models made with HaCaT WT or GALNT1/2/3 KO keratinocytes. IHC of tissue sections stained for differentiation marker keratin 10 (upper panel) or proliferation marker Ki67 (lower panel). Scale bar – 50  $\mu$ m. Red arrows – flattened cells; red asterisks – K10-negative region in suprabasal/granular layers; purple asterisks – pyknotic nuclei; green asterisks – increase in Ki67 positive cells.”

Query 3. The enrichment of protein targets in figure 3 are interesting but hard to assess:

Are the protein abundances taken into account?

Are the number of sites/protein considered in this analysis?

What is the enrichment if protein domains are considered? i.e., EGF-repeats etc.

Response 3. Figure 3 summarizes results from lectin enrichment experiments, where only glycosylated peptides are captured, and thus protein abundances are not considered, however, peptide data prior to Jacalin enrichment is available for TMT-labelled samples, which shows mostly unchanged protein levels of top targets presented in Table 1.

We have in previous exploratory O-glycoproteomic studies (Steentoft et al., EMBO J 2013, King et al., Blood Adv 2017) looked at number and distribution of glycosites on proteins, and found that the largest proportion of proteins are modified with one or several glycosites. Compared to broad mapping studies, with characterization of hundreds of glycosites, differential glycoproteomic datasets contain much lower numbers of reported targets. Therefore, we did not find such analysis relevant in the current study.

As discussed in the text, proteins containing fibronectin type 3 domains were more likely to be glycosylated by GalNAc-T2, whereas proteins containing EGF-like domains were over-represented in GalNAc-T1 and GalNAc-T3 targets. The associations are based on specific targets containing such domains. However, the location of glycosite in respect to the domain was not consistent. Of the six GalNAc-T2 targets containing FN3 domains, five targets contained O-glycosites between or next to the first/last FN3 domain, whereas one glycosite was located within the FN3 domain. In contrast, GalNAc-T1 and GalNAc-T3 glycosites were located far away from the EGF-like domains. In general, it should be mentioned that the proportion of targets is too low to make a definitive conclusion.

Action 3. We have analyzed the proteome of the TMT-labelled sample prior to Jacalin lectin enrichment, and included the average protein quantification in the 3 knock out clones compared to wild type for top targets presented in Table 1, as well as protein coverage by peptides. We excluded targets with protein ratios below 0.75.

Query 4. The quality of individual micrographs (fluorescence) is poor. The enlarged version help but without proper labelling (see above) it is hard to point out the poor ones.

Response 4. We agree that certain micrograph image elements could benefit from higher resolution, however, it is challenging to maintain the original quality given the file size requirements for initial submission.

Since figures are not labelled other problems may be present that I cannot properly decipher.

In summary, this is nice work but without figure legends, hard to assess.

2nd Editorial Decision

14 February 2020

Thank you for the submission of your revised manuscript to EMBO reports. I apologize for the unusual delay in handling your manuscript. We have now received the report from former referee #1 who supports publication of the manuscript in EMBO reports and I have editorially checked your response to the concerns from referee 2, who was not available anymore.

Given the support from referee 1 on the revised version and the fact that referee 2 has supported publication of the first version after minor revision, I thus decided to move forward with publication.

Browsing through the manuscript myself, I noticed a few editorial things that we need before we can proceed with the official acceptance of your study.

- 1) Your manuscript will be published as Report. Therefore, please combine the Results and Discussion sections. This will also help to shorten the manuscript text to get closer to our character limit of 25,000 plus/minus 2,000 characters (excl. materials and methods and references) by eliminating some redundancy that is inevitable when discussing the same experiments twice.
- 2) Fig 4I+N: please provide scale bars for the magnification boxes.
- 3) I attach to this email a related manuscript file with comments by our data editors. Please address all comments and upload a revised file with tracked changes with your final manuscript submission. I have also taken the liberty to make some changes to the Abstract. Could you please review it?

4) Finally, EMBO reports papers are accompanied online by A) a short (1-2 sentences) summary of the findings and their significance, B) 2-3 bullet points highlighting key results and C) a synopsis image that is 550x200-400 pixels large (width x height). You can either show a model or key data in the synopsis image. Please note that the size is rather small and that text needs to be readable at the final size. Please send us this information along with the revised manuscript.

We look forward to seeing a final version of your manuscript as soon as possible.

\*\*\*\*\*

## **REFEREE REPORT**

### **Referee #1:**

The authors have addressed all of the concerns I raised in the initial review. I have no further concerns.

2nd Revision - authors' response

3 March 2020

---

The authors performed all minor editorial changes.

**YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND** ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Hans H. Wandall

Journal Submitted to: EMBO Reports

Manuscript Number: EMBOR-2019-48885V2

### Reporting Checklist For Life Sciences Articles (Rev. June 2017)

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  $\chi^2$  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.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

## B- Statistics and general methods

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

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	For organotypic tissue section quantification, sample size was chosen to ensure sufficient representation of morphological feature variation throughout the section length.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
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.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
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.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Statistical test were chosen according to numbers of different groups tested, as well as data distribution. Where needed, corrections for multiple testing were applied.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Data from figures containing statistical analysis were analysed using ANOVA followed by a post hoc Dunnett's multiple comparison test using GraphPad Prism 7, as indicated in Figure Legends. In most instances the sample size was not large enough to estimate normal distribution. Data from high-throughput datasets was analyzed using R software or QluCore Omics Explorer, as indicated in Methods.
Is there an estimate of variation within each group of data?	All data is presented as means + standard deviation, unless specified differently in the figure legend.
Is the variance similar between the groups that are being statistically compared?	Not in all instances. Where parameters of statistical analysis software allowed, built-in correction for uneven variance was applied (Geisser-Greenhouse correction).

## C- Reagents

## USEFUL LINKS FOR COMPLETING THIS FORM

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

<p>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).</p>	<p>rabbit anti-clathrin: <a href="https://www.antibodypedia.com/gene/4164/CLTC/antibody/107453/4796">https://www.antibodypedia.com/gene/4164/CLTC/antibody/107453/4796</a>  rabbit anti-Rab5: <a href="https://www.antibodypedia.com/gene/3882/RAB5A/antibody/4504202/46449">https://www.antibodypedia.com/gene/3882/RAB5A/antibody/4504202/46449</a>  rabbit anti-Rab11: <a href="https://www.antibodypedia.com/gene/4588/RAB11A/antibody/166650/5589">https://www.antibodypedia.com/gene/4588/RAB11A/antibody/166650/5589</a>  rabbit anti-claudin 1: <a href="https://www.antibodypedia.com/gene/3613/CLDN1/antibody/154026/13050-1-AP">https://www.antibodypedia.com/gene/3613/CLDN1/antibody/154026/13050-1-AP</a>  mouse anti-occludin: <a href="https://www.ptglab.com/products/Occludin-Antibody-66378-1-Ig.htm">https://www.ptglab.com/products/Occludin-Antibody-66378-1-Ig.htm</a>  mouse anti-JAM-A: <a href="https://www.bio-rad-antibodies.com/monoclonal/human-cd321-antibody-m-ab-f11-mca2270.html?#purified">https://www.bio-rad-antibodies.com/monoclonal/human-cd321-antibody-m-ab-f11-mca2270.html?#purified</a>  mouse anti-integrin alpha 3: <a href="https://www.antibodypedia.com/gene/18000/ITGA3/antibody/700930/MAB1345">https://www.antibodypedia.com/gene/18000/ITGA3/antibody/700930/MAB1345</a>  mouse anti-integrin alpha 6: <a href="https://www.bio-rad-antibodies.com/monoclonal/human-cd49f-antibody-nki-goh3-mca699.html?#purified">https://www.bio-rad-antibodies.com/monoclonal/human-cd49f-antibody-nki-goh3-mca699.html?#purified</a>  rat anti-integrin beta 4: <a href="https://www.lsbio.com/antibodies/ihc-plus-itgb4-antibody-integrin-beta-4-antibody-clone-439-9b-flow-ihc-wb-western-is-b3778/111797">https://www.lsbio.com/antibodies/ihc-plus-itgb4-antibody-integrin-beta-4-antibody-clone-439-9b-flow-ihc-wb-western-is-b3778/111797</a>  rabbit anti-collagen XVII <math>\alpha</math>1: <a href="https://www.abcam.com/collagen-xvii-antibody-epr18614-ab184996.html#description_images_1">https://www.abcam.com/collagen-xvii-antibody-epr18614-ab184996.html#description_images_1</a>  goat anti-desmoglein 1: <a href="https://www.antibodypedia.com/gene/3998/DSG1/antibody/697893/AF944">https://www.antibodypedia.com/gene/3998/DSG1/antibody/697893/AF944</a></p>
	<p>rabbit anti-desmocollin 2: <a href="https://www.antibodypedia.com/gene/2782/DSC2/antibody/1455754/13876-1-AP">https://www.antibodypedia.com/gene/2782/DSC2/antibody/1455754/13876-1-AP</a>  goat anti-E-cadherin: <a href="https://www.antibodypedia.com/gene/1347/CDH1/antibody/697431/AF648">https://www.antibodypedia.com/gene/1347/CDH1/antibody/697431/AF648</a>  mouse anti CD44: <a href="https://www.antibodypedia.com/gene/804/CD44/antibody/699015/BBA10">https://www.antibodypedia.com/gene/804/CD44/antibody/699015/BBA10</a>  mouse anti-cytokeratin 10: <a href="https://www.aglient.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cytokeratin-10-(concentrate)-76617">https://www.aglient.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cytokeratin-10-(concentrate)-76617</a>  mouse anti-cyclin A: Cat. No. NCL-CYCLIN A, clone 6E6 Novocastra Laboratories, UK (<a href="https://www.leicabiosystems.com/fileadmin/img_uploads/novocastra_reagents/2013/9512_027_Rev_A_2013_USA_IHC_and_ISH_product_catalog.pdf">https://www.leicabiosystems.com/fileadmin/img_uploads/novocastra_reagents/2013/9512_027_Rev_A_2013_USA_IHC_and_ISH_product_catalog.pdf</a>)  rabbit anti-cleaved Notch1 IC: <a href="https://www.antibodypedia.com/gene/8424/NOTCH1/antibody/107054/4147">https://www.antibodypedia.com/gene/8424/NOTCH1/antibody/107054/4147</a></p>
<p>7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.</p>	<p>HaCaT and N/TERT-1 keratinocyte cell lines were obtained from JG Rheinwalds lab, Harvard Institute of Medicine, Brigham &amp; Womens Hospital. Cultured cells had routinely been tested for Mycoplasma contamination using the MycoAlert Kit from Lonza.</p>

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

#### D- Animal Models

<p>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.</p>	<p>NA</p>
<p>9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.</p>	<p>NA</p>
<p>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.</p>	<p>NA</p>

#### E- Human Subjects

<p>11. Identify the committee(s) approving the study protocol.</p>	<p>NA</p>
<p>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.</p>	<p>NA</p>
<p>13. For publication of patient photos, include a statement confirming that consent to publish was obtained.</p>	<p>NA</p>
<p>14. Report any restrictions on the availability (and/or on the use) of human data or samples.</p>	<p>NA</p>
<p>15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.</p>	<p>NA</p>
<p>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.</p>	<p>NA</p>
<p>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.</p>	<p>NA</p>

#### F- Data Accessibility

<p>18. Provide a "Data Availability" section at the end of the Materials &amp; Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Depositor'.</p> <p>Data deposition in a public repository is mandatory for:</p> <ol style="list-style-type: none"> <li>Protein, DNA and RNA sequences</li> <li>Macromolecular structures</li> <li>Crystallographic data for small molecules</li> <li>Functional genomics data</li> <li>Proteomics and molecular interactions</li> </ol>	<p>The RNAseq data from this publication have been deposited to the Gene Expression Omnibus GEO database (<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141387">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141387</a>) and assigned the identifier GSE141387. The mass spectrometry-based data from this publication (phosphoproteomics, glycoproteomics, and proteomics) have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (<a href="https://www.ebi.ac.uk/pride/">https://www.ebi.ac.uk/pride/</a>) and assigned the identifier PXD016618.</p>
<p>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).</p>	<p>See 18.</p>
<p>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).</p>	<p>NA</p>
<p>21. 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 Biomodols (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.</p>	<p>NA</p>

#### G- Dual use research of concern

<p>22. 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.</p>	<p>NA</p>
--	-----------