

In-depth and 3-Dimensional Exploration of the Budding Yeast Phosphoproteome

Michael Charles Lanz, Kumar Yugandhar, Shagun Gupta, Ethan Sanford, Vitor Faça, Stephanie Vega, Aaron Joiner, J. Christopher Fromme, Haiyuan Yu, and Marcus Bustamante Smolka
DOI: [10.15252/embr.202051121](https://doi.org/10.15252/embr.202051121)

Corresponding author(s): Marcus Smolka (mbs266@cornell.edu), Haiyuan Yu (haiyuan.yu@cornell.edu)

Review Timeline:

Submission Date:	17th Jun 20
Editorial Decision:	7th Aug 20
Revision Received:	5th Oct 20
Editorial Decision:	19th Nov 20
Revision Received:	30th Nov 20
Accepted:	3rd Dec 20

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. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

Dear Dr. Smolka

Thank you for the submission of your research manuscript to our journal. I apologize for the delay in handling your manuscript, but we have only now received the full set of referee reports that is copied below.

As you will see, the referees acknowledge that the comprehensive yeast phosphosite dataset you present will be a useful and valuable resource for the community. The referees also note that the approach to map potentially functional phosphosites to protein interfaces has been reported before. These earlier studies should be referenced and used to discuss similarities and differences in the analysis and results, as suggested by the referees.

Given the positive evaluation of your dataset as a resource for the community, we would like to invite you to revise your manuscript for potential publication in our "Resource" section 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.

I notice that a small part of the dataset was previously published (Lanz et al 2018). Please specify, which part of the dataset was reused from this earlier publication in more detail in the paper. This could be done in the methods section and you can also do it in form of a "Data citation" (see point 8 below). Alternatively, the reference to the earlier dataset could be part of the Data availability section in this case with a reference to Lanz et al 2018 (see also point 7). But it needs to be clear which part of the data was generated and published earlier.

We invite you to submit your manuscript within three months of a request for revision. This would be November 7th in your case. Yet, given the current COVID-19 related lockdowns of laboratories, we have extended the revision time for all research manuscripts under our scooping protection to allow for the extra time required to address essential experimental issues. Please contact us to discuss the time needed and the revisions further.

IMPORTANT NOTE: we perform an initial quality control of all revised manuscripts before re-review. Your manuscript will FAIL this control and the handling will be DELAYED if the following APPLIES:

- 1) A data availability section is missing.
- 2) Your manuscript contains error bars based on $n=2$. Please use scatter blots showing the individual datapoints in these cases. The use of statistical tests needs to be justified.

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.

Please note that for all articles published beginning 1 July 2020, the EMBO Reports reference style will change to the Harvard style for all article types. Details and examples are provided at <https://www.embopress.org/page/journal/14693178/authorguide#referencesformat>

When submitting your revised manuscript, we will require:

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 (). 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 (). Please find instructions on how to link your ORCID ID to your account in our manuscript tracking system in our Author guidelines
()

6) Supplementary information: You have submitted five Supplementary figures in form of a .pdf file, which is fine as such. But please note the nomenclature Appendix and Appendix Figure Sx for these. The Appendix file needs a title page with a short Table of Content including page numbers.

You have submitted 13 Supplementary tables. These are rather complex datasets and should therefore be submitted as datasets (again as excel files). The nomenclature for these is "Dataset EVx". Please provide the legends for these in a separate tab of the .xls files. Alternatively, the legend can be supplied as a separate text file (README) and zipped together with the Table/Dataset file.

See also: ()

7) Before submitting your revision, primary datasets (and computer code, where appropriate) produced in this study need to be deposited in an appropriate public database (see < <https://www.embopress.org/page/journal/14693178/authorguide#dataavailability>>). Specifically, we would kindly ask you to provide public access to the phosphoproteomics datasets that were generated in this study.

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#dataavailability>>). 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) Data citations to acknowledge the re-use of published datasets: 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 .

9) We would also encourage you to include the source data for figure panels that show essential data. Numerical data should be provided as individual .xls or .csv files (including a tab describing the data). For blots or microscopy, uncropped images should be submitted (using a zip archive if multiple images need to be supplied for one panel). Additional information on source data and instruction on how to label the files are available .

10) 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 (please specify technical or biological replicates) underlying each data point 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.
- 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.

11) 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). 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.

Yours sincerely

Martina Rembold, PhD
Editor
EMBO reports

Referee #1:

Here the authors created a comprehensive budding yeast phosphoproteomic database, using 75 of their own phosphoproteomic datasets, obtained under different growth and treatment conditions, in combination with other datasets available in the YeastMine/Saccharomyces Genome Database and a recent study from Dengjel's group, to create a dataset that in total contains 46,553 phosphosites. Based on saturation analysis, they argue that their list contains a nearly complete set of possible yeast phosphosites. By using their quantitative SILAC datasets, they analyzed the consequences of DNA damage treatments (23,000 sites) and cell cycle progression (11,000 sites) for changes in phosphorylation status. They also parsed this dataset to define where these sites map in proteins with respect to the known domain architectures available from protein structural information, and found that several hundred are located in or near to protein-protein interaction interfaces, many of which would be predicted to disrupt interactions when phosphorylated. They went on to validate two PPIs they predicted would be inhibited by phosphorylation at interface sites; Rad23/Png1, where a Rad23 S270D phosphomimic mutation prevented Png1 association, and Ypt1/Trs23, where a Ypt1 S75D phosphomimic mutation reduced Trs23 interaction. They conclude that protein phosphorylation plays an important role in regulating PPIs in yeast.

This comprehensive yeast phosphosite database should be extremely valuable to the yeast community interested in phosphorylation based signaling. The major conclusion that a significant fraction of these sites could be used to regulate protein-protein interactions is of interest, and might not have been predicted a priori. However, an earlier phosphoproteomic study (Studer et al. Science 354, 229, 2016) which defined and analyzed the phosphoproteomes of 18 fungal species, including *S. cerevisiae*, had already come to a similar conclusion, finding that many conserved sites of phosphorylation are present at protein-protein interfaces, and validating two of these to demonstrate their potential to regulate these interactions. While their *S. cerevisiae* phosphosite dataset is much larger than that reported by Studer et al. and contains more examples of PPI phosphosites, it is very surprising that the authors did not acknowledge that a similar conclusion had already been reported several years ago.

Points: 1. As the authors discuss, it is difficult to know a priori whether an identified phosphorylation site is functionally important, but two criteria that are commonly applied are whether the sites is conserved in evolution, and whether the stoichiometry of phosphorylation is reasonably high under at least some conditions. Apparently, the authors did not take into account either factor in their

interpretation of possible function. It would be particularly informative to determine whether sites located within PPI interfaces are conserved in other fungi or metazoans, as Studer et al. found (this could either be conservation of the residues themselves or prior evidence that the residue in question is phosphorylated in another species). In this regard, it would also be important to define in general what fraction of these ~46,000 phosphosites correspond to evolutionarily conserved Ser/Thr residues. Given that yeast lacks a canonical tyrosine kinase/PTP system, it would also be particularly interesting to determine what fraction of the (p)Tyr sites are conserved across evolution.

2. With regard to completeness of the yeast phosphosite database, it is not clear whether their list includes meiotic phosphosites, which is important because there is at least one meiosis-specific kinase, *Ime2*.

3. Page 7: Are the ~500 pY sites "real", i.e. can these sites be identified if pTyr mAb or superbinder SH2 domain enrichment is used instead of iMAC enrichment? If they are physiological, what are the "tyrosine" kinases involved in their phosphorylation and the PTPs that dephosphorylate them?

4. Page 8: Perhaps the authors could indicate that the increase in phosphosite identification observed using the chymotryptic phosphopeptide dataset is in part because basophilic kinase sites where neighboring Lys/Arg are key determinants of kinase recognition and phosphorylation generates tryptic phosphopeptides that are often too "small" to be detected by MS.

5. Page 9: T104 is at the end of the C2 calcium/phospholipid-binding domain of Rsp5 that localizes Rsp5 to membranes. Would phosphorylation of T104 affect the function of this domain, which has been proposed to act as a negative regulator of the Rsp5 E3 ligase activity, as well as be involved in membrane localization?

6. Page 10: What fraction of the DNA damage-induced phosphosites lie in Ser/Thr.Gln motifs, and would therefore presumably be direct Tel1/Mec1 targets?

7. Page 12: For substrate phosphorylation, ePKs generally require the peptide backbone around the target hydroxy-amino acid to adopt a linear conformation in the active site so that the target Ser/Thr hydroxyl can be presented in the correct orientation for phosphate transfer. This raises the question of how Ser/Thr residues in structured regions are presented for phosphorylation? Does there have to be conformational unfolding (c.f. PKR/eIF2 α)? Could some of the phosphosites found in solvent accessible regions be phosphorylated co-translationally?

8. Page 13, line 16: It is unfortunate that S (steric) and E (electrostatic) also stand for Ser and Glu!

9. Page 14: What domain does Rad23 S270 lie in? Although the S270D Rad23 phosphomimic mutation has functional consequences, and the reduced interaction of Rad23 S270D with Png1 is likely due to the negative Asp charge, as the authors imply, it remains possible that the effect of the Asp mutation (which is not a great phosphomimic due to Asp only possessing a single negative charge and a small hydrated ionic shell) is instead due some other local structural change induced by the mutation. A S270N mutant Rad23 would be an excellent control in this regard.

10. Page 16: The same issue applies to the S75D Ypt1 mutation, where a S75N mutation would be a nice control.

11. Finally, the authors might acknowledge that, despite their efforts, their yeast protein phosphosite database is likely incomplete because it does not contain any non-canonical protein phosphorylation sites, i.e. pHis, pArg, pLys, pCys, pAsp and pGlu sites, which are now being identified in large numbers in other eukaryotes.

Referee #2:

Comments

This manuscript titled "In-depth and 3-Dimensional Exploration of the Budding Yeast Phosphoproteome" described a comprehensive analysis of large-scale phosphoproteome data sets from budding yeast. The authors developed an analysis strategy employing multiple database search engines and phospho-site localization algorithms, and derived a set of high quality phosphosites. This set, combined with existing SGD forms the backbone of follow up analyses. The authors further extended analysis to understand the functional importance of these phosphosites using multiple strategies, including those in close proximity to a sensitizing mutation and those showing dynamic regulation. Beyond these, the authors mapped phosphosites to crystal structures and identified sites that sit at the protein-protein interaction (PPI) interface. An algorithm was developed to quantitative assess the impact of phosphorylation events on these sites. The predicted sites that potentially can interrupt PPI were validated with IP-MS. Overall, this is a very well-designed study and computational analyses. I would highly recommend for publication.

A few comments:

- 1) As the phosphosites predicted to decrease PPI were validated, have you looked at the phosphosites potentially can enhance protein-protein interaction?
- 2) Which score is more predictive: E score or S score?
- 3) Can you comment on proteome coverage of the phosphosite data set? Does it cover the entire yeast proteome?
- 4) in P17, text referred to Figure S2C but no S2C was present. Please correct.

Referee #3:

Summary

1. Does this manuscript report a single key finding?

YES

Combination of yeast phosphoproteomic data with structural protein data to identify potentially functional phosphorylation sites.

2. Is the reported work of significance (YES), or does it describe a confirmatory finding or one that has already been documented using other methods or in other organisms etc (NO)?

YES

3. Is it of general interest to the molecular biology community?

YES

Brings together available phosphoproteomic and protein structural data.

4. Is the single major finding robustly documented using independent lines of experimental evidence (YES), or is it really just a preliminary report requiring significant further data to become convincing, and thus more suited to a longer format article (NO)?

YES

In the Manuscript by Lanz et al the authors aggregate and reprocess phosphoproteomic data of budding yeast that they have previously generated. The dataset includes budding yeast arrested at different cell cycle stages and in combination with the DNA damaging agent MMS. They evaluate the depth of their dataset, assess different phospho site localization score cutoffs, compare their data to a few available datasets and perform basic enrichment analyses for cell cycle and DNA damage dependent sites. They then proceed to map phosphorylation sites onto available protein structures to derive potentially important sites for protein-protein interactions which they summarize in an interactive online platform. They validate the importance of two phosphorylation sites in disrupting protein complex formation.

The paper is well written and the subject under investigation highly relevant. The results are presented in a clear way. The study provides a comprehensive novel online resource that will be useful for the community. However, I have some reservations with the claims of the authors about the depth of their phosphoproteomic analysis, the comprehensiveness of the analyzed conditions and the analysis regarding cell cycle regulation and DNA damage signaling. The aggregated phosphoproteomic data is mostly qualitative and not quantitative and should be treated that way. The coherence of the study could benefit from shortening the cell cycle/ DNA damage part.

I support the publication of the manuscript after the authors address the following concerns and recommendations:

1) The phosphoproteomic data comes from many independent experiments that were designed to answer focused questions relevant to DNA damage and cell cycle. Cells were exposed to different treatments (0.02% MMS, alpha factor, low glucose, nocodazole) and different yeast strains were used (including kinase knockouts and kinase overexpression strains). The yeast phosphoproteome is highly dynamic and strongly dependent on activated kinases and the background proteome. The presented dataset probes a small and biased set of biological conditions and is therefore not sufficient to assess the overall phosphoproteomic space in budding yeast. In this light, statements like "Saturation analysis suggests that most tryptic phosphopeptides in this organism have now been identified" seem like a large overstatement.

I suggest that the authors define a core set of basal phosphorylation sites and sets of condition-specific sites. They can then perform saturation analysis of the basal phosphoproteome. They should check if including condition-specific phosphoproteome sets in their saturation analysis are expanding the phosphoproteomic space.

2) The DNA damage and cell cycle analysis (Figure 3) is not coherent with the story and does

provide only limited new insights relevant to this study as it stands. Figure 3A-D shows the activation profile of Tel1, Mec1, Rad53 and CDK kinases by monitoring the phosphorylation of known substrates of these kinases, which is interesting, but unrelated to the main focus of the study, i.e. to reveal unknown regulated phosphorylation sites to identify potentially functional sites. I suggest the authors focus on identifying significantly regulated sites rather than monitoring the behavior of known kinase substrates. They should report if they can find something novel about the regulated phosphosites. They should find a way to present these regulated sites in the broader context of their overall dataset.

3) The SILAC experiments in Figure 3 completely lack any statistical control. Rather than just showing replicate correlation of SILAC phosphosite abundances the authors should perform statistical testing to determine significantly regulated sites (e.g t-test or ANOVA).

4) The cell cycle analysis is not quantitative and not statistically controlled. The authors use a spectral counting approach to infer phospho site abundance, which has many weaknesses, in particular because it is applied to many independent experiments. It is not clear what data was used to perform this analysis since it is not indicated in table S1. It is also not clear if and how many biological replicates were used. It seems that this approach is highly dependent on the efficiency of the phosphopeptide enrichment, the reproducibility of the HILIC fractionation and the quality of the LC-MS/MS measurement. Therefore the authors need to provide more rigorous quality control. They need to prove that this approach is robust, reproducible and has enough sensitivity to identify cell cycle-dependent phosphorylation events. They need to show biological replicate reproducibility across independent experiments. Apart from identifying 5 PSMs for a site to be considered for this analysis, they should also include a filter requiring the phosphosite to be identified in more than one biological replicate.

Overall the cell cycle analysis is of much lesser quality and resolution than comparable studies (eg. Swaffer et al, 2018 in *S. pombe*). I suggest eliminating Figure S5 and toning down the description and discussion of the cell cycle results. The authors should definitely tone down their claim on having created "the most extensive catalog of cell cycle-dependent phosphorylation events in this organism".

5) The authors need to prove that they are not inflating the final number of phosphosites in their dataset by incorrect FDR filtering. Specifically they need to revisit their search strategy using multiple different search engines. In general it is not ok to combine results of two different search engines that have been individually filtered for FDR, since every spectrum has double the chance of getting called. The only time this works (without another tool) is if the two searches look for distinct sets of peptides, such that the same peptide is not considered twice

(<https://pubs.acs.org/doi/abs/10.1021/pr501173s>). If the authors want to combine results of multiple search engines they need to use a tool that allows them to do that (e.g. iProphet <https://www.mcponline.org/content/10/12/M111.007690.long>).

Furthermore it is unclear why they use a very high initial FDR filter of >10% for their SEQUEST search.

Minor comments:

1) Several supplementary tables are wrongly referenced in the text, please check all references. Eg. Table S11 on bottom of page 12 is wrong (should be Table S9). Table S10 page 43 line 2 is wrong (should be Table S11).

2) For the 3D analysis: Other studies have conducted similar analysis looking at phospho sites in

protein interfaces (e.g. Ochoa et al. Nat Biotechnol 2020 10.1038/s41587-019-0344-3, Beltrao et al. Cell 2012 10.1016/j.cell.2012.05.036 and Studer et al. Science 2016 0.1126/science.aaf2144). The authors should reference those studies and discuss similarities and differences in their analysis and results.

3) It is not clear from the text if the information derived from the computationally positioned phospho sites onto the PDB structures is also made available on superPhos.

4) For the Yen1 and Mrc1 analysis: Define in which condition these sites were found by AP-MS. If the proteins or the reported phospho sites were DNA damage or cell cycle induced this may justify why the sites were found in this study and underrepresented phosphoproteomic datasets.

5) The link to the superPhos online database should be indicated in the abstract.

6) Figure 1: Add the % symbol to the localization probability values.

Reviewer 1

...an earlier phosphoproteomic study (Studer et al. Science 354, 229, 2016) which defined and analyzed the phosphoproteomes of 18 fungal species, including S. cerevisiae, had already come to a similar conclusion, finding that many conserved sites of phosphorylation are present at protein-protein interfaces, and validating two of these to demonstrate their potential to regulate these interactions...it is very surprising that the authors did not acknowledge that a similar conclusion had already been reported several years ago.

As part of a broader study on the conservation of phosphorylated residues across multiple yeast species, Studer et al. also assessed the prevalence of phosphorylation at protein-protein interfaces. Though their analysis contains ~10-fold less budding yeast phosphosites, Studer et al. implemented a similar methodology for mapping phosphorylation to protein interfaces and predicting the impact on protein-protein interactions. This absence of this citation was an oversight on our part. The work is now referenced and discussed throughout the text:

“This concept was previously explored in budding yeast by Studer et al., albeit on a much smaller scale. To systematically identify phosphorylation that would result in “clashes” between interacting proteins, we devised a minimal scoring system based on the steric and electrostatic environment surrounding phosphosites near a protein interface region (see methods for detailed explanation of how the scores were calculated). Our method is similar, in essence, to the approach employed by Studer et al.”

“Overall, together with previous work, these examples further support the notion that it is possible to systematically predict the impact of phosphorylation on the regulation of protein-protein interactions based on the structural context of its occurrence.”

In addition to the larger size of the database being analyzed, our work is also unique because it uses Interactome Insider, which expands the analysis of phosphorylation at interaction interfaces to proteins without crystal structure (e.g. proteins with only homology models, or interaction interfaces predicted based on conservation).

1) ...two criteria that are commonly applied are whether the sites is conserved in evolution, and whether the stoichiometry of phosphorylation is reasonably high under at least some conditions. Apparently, the authors did not take into account either factor in their interpretation of possible function. It would be particularly informative to determine whether sites located within PPI interfaces are conserved in other fungi or metazoans.

The reviewer highlights two features of a phosphorylation event that can potentially contribute to a prediction of functional importance - #1 Conservation and #2 Stoichiometry. In our revised manuscript, we more deeply explore these two important features in our dataset.

Conservation: We performed an alignment analysis similar to Studer *et al.* We calculated a conservation 'score' for every phosphosite that could be aligned to multiple yeast species. Like Studer *et al.*, we found that phosphorylation sites that localize to interface residues are significantly more conserved on average (new Figure S6). We distilled our conservation analysis into a single value, the conservation score (see methods section), and have incorporated that information into the final dataset.

“Studer et al. also reported that phosphorylation sites that lie at interface residues tend to exhibit more conservation throughout multiple fungal species, a finding supported by our own investigation of phosphosite conservation in our dataset (see methods for full description). We distilled our conservation analysis into a single score and incorporated into our final dataset (Table S2).”

Stoichiometry: Though some methods have been described to determine phosphorylation stoichiometry using phosphoproteomic mass spec, the absence of stoichiometric information from the vast majority of phosphoproteomic datasets is a fundamental limitation of the technology. In acquiring our dataset, we did not employ a specific approach to calculate phosphosite stoichiometry, which would be very challenging given the extent of our analysis. That said, we were able to perform an imperfect, yet informative analysis to infer high phosphorylation stoichiometry for phosphosites in low abundant proteins (new Fig 3B, Supplemental Table 2). By comparing the number of phosphosite PSMs with the predicted copy number of the associated phosphoprotein, we now highlight a set of phosphorylation events that likely occur at higher stoichiometry. This figure also highlights dynamic ranges of PSM numbers for the phosphosites identified in our study, as some phosphosites have thousands of independent identifications.

“While the ease with which a phosphorylated peptide is detected depends upon multiple factors (i.e. sequence-specific chemical properties that influence digestion, solubility, enrichment, ionization, etc.), its abundance in the injected mixture is a major contributing determinant. With this in mind, we plotted the number of PSMs for each phosphosite as a function of the harboring protein’s estimated copy number (Fig 3B). As an imperfect means to infer phosphorylation stoichiometry, we highlight 500 phosphosites with highest PSM# -to- protein abundance ratios (Fig 3B, see “#identifications” and “ProteinAbundance” columns in Table S2).”

2) *...it is not clear whether their list includes meiotic phosphosites, which is important because there is at least one meiosis-specific kinase, Ime2”*

The reviewer is correct to point out that our dataset does not contain experiments on meiotic cells. So, we do not capture meiotic-specific phosphorylation events like those mediated by Ime2. This fact is now acknowledged in the text:

“We also note that our dataset lacks spectra acquired from meiotic conditions and, therefore, may not contain phosphorylation events mediated by meiosis-specific kinases, like Ime2 (Foiani et al, 1996; Guttmann-Raviv et al, 2002).”

3) *...are the ~500 pY sites "real", i.e. can these sites be identified if pTyr mAb or superbinder SH2 domain enrichment is used instead of iMAC enrichment? If they are physiological, what are the "tyrosine" kinases involved in their phosphorylation and the PTPs that dephosphorylate them?*

While we have not attempted to enrich for phospho-tyrosine using the methods mentioned from the reviewer, we are confident that there are hundreds of phospho-tyrosine peptides in our dataset. These include events with previously established regulatory roles (e.g. Swe1 phosphorylation of CDK pY19). This may be because many budding yeast kinases, like Swe1 and MAKK, are thought to have dual specificity.

After receiving this feedback from the reviewer, we decided to emphasize the following point more clearly in the revised manuscript: phospho-tyrosine sites are more likely to be falsely localized. This is clearly demonstrated by the fact that decreasing phosphosite localization tolerance nearly doubles the number of identified pTyr sites in our dataset (highlighted in the new Fig 2A).

The explanation for this observation is that Tyrosine phosphorylation is much rarer than Serine/Threonine phosphorylation. Thus, given that most sites have some chance of mis-localization, the probability that a true Serine/Threonine phosphosite is mis-localized to a Tyrosine is much higher than a true Tyrosine phosphosite mis-localizing to a Serine/Threonine. Consistent with this notion, we performed a quality control analysis for the group of phosphopeptides with assigned pTyr and found that phospholocalization scores are lower compared to pSer and pThr.

We now explicitly encourage the use of higher localization scores for users interested in pTyr phosphorylation:

“Because sites identified as phospho-tyrosine in our study (and possibly YeastMine) are prone to represent mis-localized phospho-serine or -threonine, we encourage the careful consideration of the PSM quality metrics when investigating tyrosine phosphorylation. We found that filtering based on the number of phosphosite identifications (PSMs) dramatically increases overall data quality (Fig 2C and D) and reduces the overall false discovery rate (Fig 2E, Fig S2).”

4) ...increase in phosphosite identification observed using the chymotryptic phosphopeptide dataset is in part because basophilic kinase sites where neighboring Lys/Arg are key determinants of kinase recognition...?

The reviewer is probably referring to the fraction of phosphorylation sites that are directly adjacent to a Lysine/Arginine in the Tryptic VS Chymotryptic dataset. It is possible that phosphorylation directly adjacent to lysine/arginine may impede digestion by trypsin, which would be less of an issue for chymotrypsin. In addition, we think the use of chymotrypsin allows us to sample regions of the proteome that are depleted of lysine/arginine (sometimes referred to as lysine/arginine deserts) and that phosphopeptides from these regions provide the ‘boost’ to our identification numbers.

5) T104 is at the end of the C2 calcium/phospholipid-binding domain of Rsp5 that localizes Rsp5 to membranes. Would phosphorylation of T104 affect the function of this domain, which has been proposed to act as a negative regulator of the Rsp5 E3 ligase activity, as well as be involved in membrane localization?

Interesting point! This is exactly the goal of making this resource available for the community. We hope other researchers can pursue these biological questions further. We feel that addressing this question is beyond our current capabilities and beyond the scope of the current manuscript.

6) Page 10: What fraction of the DNA damage-induced phosphosites lie in Ser/Thr.Gln motifs, and would therefore presumably be direct Tel1/Mec1 targets?

In the original manuscript, every highlighted Mec1/Tel1 substrate (blue) harbored the consensus S/T-Q phosphorylation motif, which represents a significant fraction of the significantly changing sites in Figure 4B. We designed Table S7 in a way that enables the reader to filter for kinase recognition motifs-of-interest.

7) Page 12: For substrate phosphorylation, ePKs generally require the peptide backbone around the target hydroxy-amino acid to adopt a linear conformation in the active site so that the target Ser/Thr hydroxyl can be presented in the correct orientation for phosphate transfer. This raises the question of how Ser/Thr residues in structured regions are presented for phosphorylation? Does there have to be conformational unfolding (c.f. PKR/eIF2 α)? Could some of the phosphosites found in solvent accessible regions be phosphorylated co-translationally?

Co-translational phosphorylation is a very interesting possibility and provides a very logical explanation for how a kinase is able to access these inaccessible regions. We thank the reviewer for raising this interesting point. However, we think providing direct evidence for this phenomenon is beyond the scope of this work, but we will keep it in mind for future investigations.

8) It is unfortunate that S (steric) and E (electrostatic) also stand for Ser and Glu!

We agree. We now use STE and ELE as shorthand for steric and electrostatic, respectively.

9 / 10) Page 14: *What domain does Rad23 S270 lie in? Although the S270D Rad23 phosphomimic mutation has functional consequences, and the reduced interaction of Rad23 S270D with Png1 is likely due to the negative Asp charge, as the authors imply, it remains possible that the effect of the Asp mutation (which is not a great phosphomimic due to Asp only possessing a single negative charge and a small hydrated ionic shell) is instead due some other local structural change induced by the mutation. A S270N mutant Rad23 would be an excellent control in this regard. : The same issue applies to the S75D Ypt1 mutation, where a S75N mutation would be a nice control.*

The reviewer suggests a good experiment to define the relative contribution of electrostatics versus steric determinants. As proposed by the reviewer, we performed an IP-MS of the Rad23 S270N mutant and found that the S270N mutation did not impair the Rad23-Png1 interaction to the same extent as S270D (new Fig S7). This result suggests that the electrostatics is the major determinant impairing this interaction, with a potential low or no effect of steric constraints. This finding may reveal that ELE scores overall may be a better predictor of phosphosites that impair interactions compared to STE scores, although a more systematic study comparing S-D to S-N mutations in several interactions would be necessary to fully address this point. Unfortunately, this is beyond our current capabilities. However, our current data with Rad23 should motivate future studies to undertake a more thorough and focused investigations in this precise question.

11) *...the authors might acknowledge that, despite their efforts, their yeast protein phosphosite database is likely incomplete because it does not contain any non-canonical protein phosphorylation sites, i.e. pHis, pArg, pLys, pCys, pAsp and pGlu sites, which are now being identified in large numbers in other eukaryotes.*

This is now acknowledged in the text:

“Moreover, our search pipeline does not capture phosphorylation that occurs on non-canonical residues, which has recently been identified in other eukaryotes (Hardman et al, 2019).”

In case the reviewer is curious – we did try searching for these non-canonical sites in our yeast dataset. Because these alternative phosphorylation events were exceedingly rare in our dataset (likely due to our sample preparation / fractionation protocol, which was different than other studies focused on identifying non-canonical residues) when compared to the prevalence of STY phosphorylation, every MS2 spectra that ‘matched’ to a non-canonical phosphopeptide was of dubious quality (manual inspection). We therefore did not pursue these non-canonical sites further.

Reviewer 2

...I would highly recommend for publication.

Thank you for the positive feedback!

As the phosphosites predicted to decrease PPI were validated, have you looked at the phosphosites potentially can enhance protein-protein interaction?

Very interesting point - we plan to address this point in future work. There are several examples in the literature of phosphorylation being required for the physical interaction of two proteins. These interactions are mediated by protein domains that recognize and bind phosphorylated proteins (BRCT domains are one prominent example). In our analysis, it would be reasonable to predict that a phosphorylation event which lies in a PPI region with compatible electrostatics (and does not interfere sterically) may strengthen the PPI. This hypothesis would require further testing. That said, it will likely be more difficult to find examples of PPI enhancing-phosphorylation from our analysis, since proteins that are crystalized together for structural analysis are not typically phosphorylated.

Which score is more predictive: E score or S score?

This is an important question. Please also see our response to Reviewer #1's point 9/10 above. In summary, it is difficult to say without testing more examples, although our examples with Rad23 suggests that ELE (electrostatic) score seems more predictive. It is important to note that the quality of any STE or ELE score prediction will always depend on the quality of the crystal structure, which can vary significantly (even between crystal structures of similar resolution).

Can you comment on proteome coverage of the phosphosite data set? Does it cover the entire yeast proteome?

The saturation analysis in Figure 2E suggests about 2/3 of yeast proteins harbor at least a single phosphorylation site. As pointed out by reviewer 1, our dataset does not contain any experiments performed with meiotic yeast, so we do not expect complete coverage of the yeast proteome.

Reviewer 3

I have some reservations with the claims of the authors about the depth of their phosphoproteomic analysis, the comprehensiveness of the analyzed conditions and the analysis regarding cell cycle regulation and DNA damage signaling. The aggregated phosphoproteomic data is mostly qualitative and not quantitative and should be treated that way. The coherence of the study could benefit from shortening the cell cycle/ DNA damage part.

Reviewer 3 highlighted important parts of our manuscript that needed improvement/clarity. As described in our responses below, we performed several analyses requested by the reviewer and adjusted the text and figures according to many of his/her suggestions. We believe these changes helped improve the manuscript.

1.1) ...The yeast phosphoproteome is highly dynamic and strongly dependent on activated kinases and the background proteome. The presented dataset probes a small and biased set of biological conditions and is therefore not sufficient to assess the overall phosphoproteomic space in budding yeast. In this light, statements like "Saturation analysis suggests that most tryptic phosphopeptides in this organism have now been identified" seem like a large overstatement.

We agree that our claims of 'saturation' should be toned down. We have altered our claims as follows:

We removed the sentence - "*We next sought to estimate whether our ability to discover new budding yeast phosphoproteins or phosphosites using mass spectrometry is reaching saturation.*"

We removed this phrase - "*...while our ability to detect novel tryptic phosphopeptides in budding yeast is reaching saturation...*"

We added the sentence - "*It is likely that the phosphoproteome can also be expanded by exploring a more diverse set of cellular states. For example, our dataset lacks spectra acquired from meiotic conditions and, therefore, may not contain phosphorylation events mediated by meiosis-specific kinases, like Ime2 (Foiani et al, 1996; Guttmann-Raviv et al, 2002).*"

1.2) ...I suggest that the authors define a core set of basal phosphorylation sites and sets of condition-specific sites. They can then perform saturation analysis of the basal phosphoproteome. They should check if including

condition-specific phosphoproteome sets in their saturation analysis are expanding the phosphoproteomic space.

We agree that this analysis suggested by the reviewer would better frame the current saturation of the phosphoproteome. However, in our case, defining a “basal” set of phosphorylation events is difficult since most experiments in our dataset are performed in different DNA damaging conditions. Also, to attempt this analysis we would have to re-distribute all our experiments into “condition-specific” groups and then re-search all our spectral data (which requires hundreds of hours of search time).

We think our adjustments to the text (see above point), which acknowledge this important point raised by the reviewer, sufficiently addresses this concern. Our more tempered interpretation of the saturation analysis can be found in the abstract and the first paragraph of the Discussion session – *“Importantly, the analysis in Figure 2E suggests that the size of the phosphoproteome can be expanded further using alternative digestive enzymes.”*

2) The DNA damage and cell cycle analysis (Figure 3) is not coherent with the story and does provide only limited new insights relevant to this study as it stands. Figure 3A-D shows the activation profile of Tel1, Mec1, Rad53 and CDK kinases by monitoring the phosphorylation of known substrates of these kinases, which is interesting, but unrelated to the main focus of the study, i.e. to reveal unknown regulated phosphorylation sites to identify potentially functional sites. I suggest the authors focus on identifying significantly regulated sites rather than monitoring the behavior of known kinase substrates. They should report if they can find something novel about the regulated phosphosites. They should find a way to present these regulated sites in the broader context of their overall dataset. The SILAC experiments in Figure 3 completely lack any statistical control. Rather than just showing replicate correlation of SILAC phosphosite abundances the authors should perform statistical testing to determine significantly regulated sites (e.g t-test or ANOVA).

To address several valid criticisms from the reviewer, we completely revised our DNA damage analysis (new Figure 4). To address the reviewer’s concern about statistical control, we performed a volcano analysis to derive a p-value for the changes seen in the behavior of individual phosphosites. The statistical derivation of significance is outlined in the Table S7, along with a list of all significantly changing phosphosites. Volcano plots now highlight the breadth of significant changes in the phosphoproteome for the conditions tested.

The reviewer also suggested that we *“present these regulated sites in the broader context of their overall dataset.”* To address this concern from the reviewer, we now plot the kinase regulated sites within the broader context of all significantly “regulated” (or changing) sites within the phosphoproteome (Figure 4, Volcano plots). We also added violin plots to visualize the distribution of these kinases-regulated sites and how they change, both between experiments and relative to the rest of the phosphoproteome.

Thanks to the reviewer’s suggestions, we think our analysis now better facilitates the exploration of the dataset, and we hope the readers will other uncover other interesting trends in the data.

4.1) The cell cycle analysis is not quantitative and not statistically controlled. The authors use a spectral counting approach to infer phospho site abundance, which has many weaknesses, in particular because it is applied to many independent experiments. It is not clear what data was used to perform this analysis since it is not indicated in table S1. It is also not clear if and how many biological replicates were used. It seems that this approach is highly dependent on the efficiency of the phosphopeptide enrichment, the reproducibility of the HILIC fractionation and the quality of the LC-MS/MS measurement. Therefore the authors need to provide more rigorous quality control. They need to prove that this approach is robust, reproducible and has enough sensitivity to identify cell cycle-dependent phosphorylation events. They need to show biological replicate reproducibility across independent experiments. Apart from identifying 5 PSMs for a site to be considered for this analysis, they should also include a filter requiring the phosphosite to be identified in more than one biological replicate. Overall the cell cycle analysis is of much lesser quality and resolution than comparable

studies (eg. Swaffer et al, 2018 in S. pombe). I suggest eliminating Figure S5 and toning down the description and discussion of the cell cycle results. The authors should definitely tone down their claim on having created "the most extensive catalog of cell cycle-dependent phosphorylation events in this organism".

The reviewer raises a valid concern. The strength of our cell cycle analysis comes mostly from its size. But as the reviewer points out, this analysis is only semi-quantitative, lacks statistical control, and has many inherent weaknesses. We have explicitly stated these concerns in the text and moved all cell cycle-related analyses to the supplemental data. We still believe that the general information may be useful for some readers that are searching for clues pointing to potential regulation of different phosphosites in the context of the cell cycle. We have added text to explicitly state the caveats of this analysis and caution the readers would need to take when interpreting the results (which are now moved to supplementary). For example, while not often the case, the stochastic issue becomes more relevant for phosphopeptides with lower overall number of PSMs, and whose detection may be more subjected to the stochastic nature of the whole procedure.

In the table provided in Figure S5 we tried to highlight that, despite the many limitations of the analysis, the sites with cell cycle "enrichment" make sense in the context of cell cycle biology. We acknowledge this attempted "validation" was far more qualitative than quantitative, and moved this analysis entirely to the supplement.

As suggested by the reviewer, we re-wrote the text to emphasize the limitations of our cell cycle dataset:

"Though our cell cycle analysis constitutes a large catalog of phosphorylation events, we caution that due to technical and experimental limitations, our approach using "spectral counting" for this dataset lacks the quantitative accuracy and temporal resolution achieved by more focused investigations of cell cycle phosphorylation dynamics in yeast (Swaffer et al, 2018; Touati et al, 2018)."

Concerning: "It is not clear what data was used to perform this analysis since it is not indicated in table S1. It is also not clear if and how many biological replicates were used."

Sorry, that column was missing from Table S1. We fixed this issue.

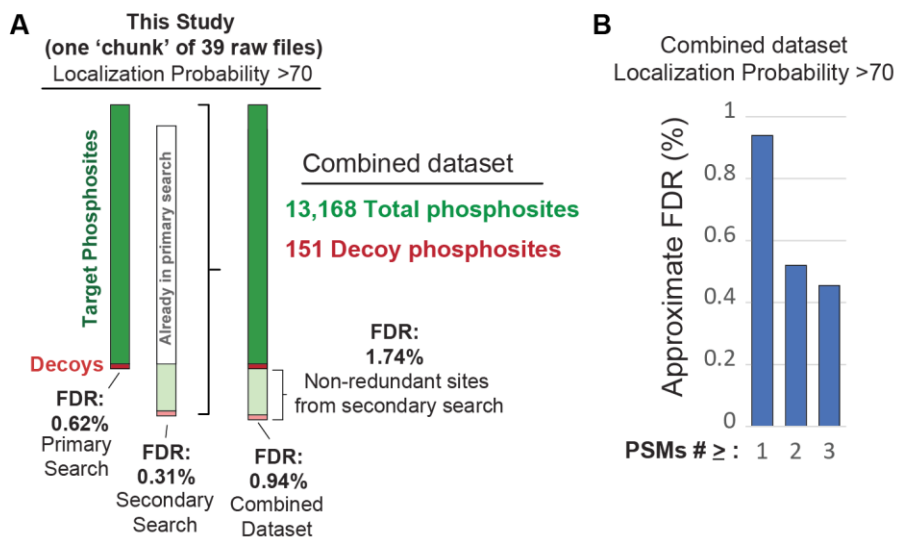
5) The authors need to prove that they are not inflating the final number of phosphosites in their dataset by incorrect FDR filtering. Specifically they need to revisit their search strategy using multiple different search engines. In general it is not ok to combine results of two different search engines that have been individually filtered for FDR, since every spectrum has double the chance of getting called. The only time this works (without another tool) is if the two searches look for distinct sets of peptides, such that the same peptide is not considered twice (<https://pubs.acs.org/doi/abs/10.1021/pr501173s>). If the authors want to combine results of multiple search engines they need to use a tool that allows them to do that (e.g. iProphet <https://www.mcponline.org/content/10/12/M111.007690.long>).

This is a very important point raised by the reviewer. We looked into the tools suggested by the reviewer. Unfortunately, we found they are not compatible with Maxquant's default output. So, we opted to manually investigate our FDR, specifically accounting for the problem mentioned by the reviewer.

By researching a chunk of our dataset (~40 raw files) with a manually embedded decoy database, we were able to monitor FDR at every step of the search process and to separately assign an FDR to the primary and secondary searches, both before and after combining them into the final dataset. This allowed us to monitor the FDR within the subgroup of phosphosites that were only identified after adding in the secondary search, and thus account for the problem mentioned by the reviewer.

We found that we are not dramatically inflating the number of phosphosites using the secondary search. The overall FDR for the final dataset is less than 1%. Even when considering only the non-redundant phosphosites

that are contributed from the secondary search, the FDR was less than 2% (Fig S2A, figure below - not drawn to scale). We also note that the FDR for the secondary search is overestimated, since decoys in Proteome Discoverer are not assigned a localization score and cannot be filtered by localization confidence (the Target sites are >70). Additionally, this analysis allowed us to demonstrate that the FDR in our dataset dramatically decreases when phosphosites with only a single PSM are excluded (Fig 2E).



Furthermore it is unclear why they use a very high initial FDR filter of >10% for their SEQUEST search.

This was a typo in the supplemental figure. We used an FDR of lower than 1% for the PD searches. We did use a 10% FDR for the SORCERER search, for which we only considered peptide backbones, followed by a stringent filter in which we only consider hits overlapping with PD results (the FDR for the Sorc/PD secondary search was 0.31%), so final FDR is kept way below 1%. The reason for using a higher FDR filter for SORCERER was based on our observation that the SORCERER engine is not great at assigning proper Peptide Prophet Scores for phosphopeptides, but that the dual SORCERER-PD approach does add much increased confidence to PD results.

Minor points:

Several supplementary tables are wrongly referenced in the text, please check all references. Eg. Table S11 on bottom of page 12 is wrong (should be Table S9). Table S10 page 43 line 2 is wrong (should be Table S11).

Thankyou. All tables should now be referenced correctly.

For the 3D analysis: Other studies have conducted similar analysis looking at phospho sites in protein interfaces (e.g. Ochoa et al. Nat Biotechnol 2020 10.1038/s41587-019-0344-3, Beltrao et al. Cell 2012 10.1016/j.cell.2012.05.036 and Studer et al. Science 2016 0.1126/science.aaf2144). The authors should reference those studies and discuss similarities and differences in their analysis and results.

These papers are acknowledged and discussed in both the results and discussion section.

It is not clear from the text if the information derived from the computationally positioned phospho sites onto the PDB structures is also made available on superPhos.

Currently, only the information from Interactome Insider is embedded into the superPhos website. We are working on a future project that

For the Yen1 and Mrc1 analysis: Define in which condition these sites were found by AP-MS. If the proteins or the reported phospho sites were DNA damage or cell cycle induced this may justify why the sites were found in this study and underrepresented phosphoproteomic datasets.

Like the majority of our dataset, both of these studies were performed in the presence of DNA damaging agents. The point of this figure is more to highlight the overlap between our dataset and studies using affinity purified protein rather than to claim deeper coverage.

Dear Dr. Smolka

Thank you for the submission of your revised manuscript to EMBO reports. Your manuscript was re-reviewed by former referee #1 and #3 and we have now received their reports (copied below).

As you will see, both referees are very positive about the study and request only minor changes to clarify text and figures. Please clarify whether Figure 2E and S2B are indeed reporting the same data.

From the editorial side, there are also a few things that we need before we can proceed with the official acceptance of your study:

1) Please add up to five keywords

2) The manuscript sections are currently not in the correct order and some headings are missing (e.g., Figure Legends).

Please see <https://www.embopress.org/page/journal/14693178/authorguide#textformat> for further information.

3) Please add a 'Conflict of Interest' paragraph

4) Appendix:

- Appendix table of content: Please add page numbers
- Please remove the Dataset legends from the Appendix and include them in a separate sheet of the Dataset .xls files called 'Legend'
- Please change the nomenclature to 'Appendix Figure Sx'
- In Appendix Figure S1 you refer to "Supplemental Table 2" and 'Supplemental Table 3', which do not exist. Please update this information.
- In Appendix Figure S4 you refer to "Supplemental Table 7", which does not exist. Please update this information
- Appendix Figure S6: please define the error bars and the number of experiments the quantification is based on in the legend.
- Please change the header of the references to e.g. 'Appendix references'

5) Figure callouts:

- Please add a callout for Fig. 3A and for Appendix Fig S6 in the text where appropriate
- Please update the callout to Fig S7, which does not exist.
- Please correct the nomenclature for Appendix figure callouts. They are missing the word 'Appendix' (e.g. Appendix Fig S1).

6) Figure legends:

- Figure 6B, E, H: the number of replicates has not been specified in the figure legend. Can this be added?

7) Please note that the abstract should describe new findings in present tense.

8) Data deposition in PRIDE: Please make sure that the dual deposition of the dataset from Lanz et al 2018 as part of PXD009734 and of PXD012395 is not against the policies of the database.

9) Please note that all corresponding authors are required to supply an ORCID ID for their name (<<https://orcid.org/>>). This information is still missing for Dr. Haiyuan Yu. 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>>)

10) All funding info needs to be entered into the relevant fields in our online submission system. We note that "Equipment Supplement R01GM097272-07S1" is missing.

11) 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 550x200-600 pixels large (width x height) in .png format. 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.

Kind regards,

Martina Rembold, PhD
Senior Editor
EMBO reports

Referee #1:

This is improved, and the authors have addressed most of my concerns, in some cases by including new data.

To address the question of phosphorylation stoichiometry, Figure 3B was added, and describes an indirect but reasonable method to infer stoichiometry of phosphorylation at all the sites in their database.

To address the question of whether the inhibitory effect of the S270D mutation on the binding of Rad23 to Png1 was dependent on the negative charge or rather the size/hydrophilicity of the Asp at position 270, Figure S6 was added and examines a Rad23 S270N mutant on binding of Png1 binding, showing that the reduction in binding to Png1 was significantly less than that of the S270D Rad23 mutant. However, it is interesting that now it appears that Rad4 binding is affected by the S270N mutation, whereas it was not by the S270D mutation.

Point: On page 15 the authors refer to this as Figure S7, and this should be corrected.

Referee #3:

The authors answered my concern and improved the clarity of their manuscript. I support the presented restructuring of the DNA damage and cell cycle analysis.

Minor comments:

- Page 9, paragraph 11 Fig 3E is wrongly referenced as Fig 2E.
- Figure 2E appears to be the same as Figure S2B.
- Second last response of the authors to reviewer 3 is incomplete.

Response to editor:

Please clarify whether Figure 2E and S2B are indeed reporting the same data.

These panels reported the same data, so we removed S2B from the appendix.

1) Please add up to five keywords

Five keywords have been added to the title page.

2) The manuscript sections are currently not in the correct order and some headings are missing (e.g., Figure Legends).

Please

see <https://www.embopress.org/page/journal/14693178/authorguide#textformat> for further information.

We added a "Figure Legends" heading. Also, we re-ordered the sections so that they align with the guidelines in the link above.

3) Please add a 'Conflict of Interest' paragraph

This section has been added.

4) Appendix:

- Appendix table of content: Please add page numbers
- Please remove the Dataset legends from the Appendix and include them in a separate sheet of the Dataset .xls files called 'Legend'
- Please change the nomenclature to 'Appendix Figure Sx'
- In Appendix Figure S1 you refer to "Supplemental Table 2' and 'Supplemental Table 3', which do not exist. Please update this information.
- In Appendix Figure S4 you refer to "Supplemental Table 7', which does not exist. Please update this information
- Appendix Figure S6: please define the error bars and the number of experiments the quantification is based on in the legend.
- Please change the header of the references to e.g. 'Appendix references'

These requested changes to the appendix have been made.

5) Figure callouts:

- Please add a callout for Fig. 3A and for Appendix Fig S6 in the text where appropriate
- Please update the callout to Fig S7, which does not exist.
- Please correct the nomenclature for Appendix figure callouts. They are missing the word 'Appendix' (e.g. Appendix Fig S1).

The figure callouts have been included and corrected.

6) Figure legends:

- Figure 6B, E, H: the number of replicates has not been specified in the figure legend. Can this be added?

The number of biological replicates is now included in the legends for these figure panels.

7) Please note that the abstract should describe new findings in present tense.

We made minor edits to the abstract.

8) Data deposition in PRIDE: Please make sure that the dual deposition of the dataset from Lanz et al 2018 as part of PXD009734 and of PXD012395 is not against the policies of the database.

I contacted PRIDE about whether this violated their policy. They said: "No, it doesn't. But it would be good if you can mention somewhere in the dataset that it is an extended study of your previous data." This information is already included in the text and the data availability section.

9) Please note that all corresponding authors are required to supply an ORCID ID for their name (<<https://orcid.org/>>). This information is still missing for Dr. Haiyuan Yu. 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>>);)

We could not change the ORCID info on the uploader. We will send an accompanying email with the ORCID info for Haiyuan and Michael.

10) All funding info needs to be entered into the relevant fields in our online submission system. We note that "Equipment Supplement R01GM097272-07S1" is missing.

The funding section is now updated.

11) 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 550x200-600 pixels large (width x height) in .png format. 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.

The summary sentences are now including on the title page. The graphical abstract is uploaded separately as a PDF.

Dr. Marcus Smolka
Cornell University
Weill Institute for Cell and Molecular Biology
339 Weill Hall
Ithaca, NY 14853-7202
United States

Dear Marcus,

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

At the end of this email I include important information about how to proceed. Please ensure that you take the time to read the information and complete and return the necessary forms to allow us to publish your manuscript as quickly as possible.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. As you are aware, 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.

If you do NOT want this File to be published, please inform the editorial office within 2 days, if you have not done so already, otherwise the File will be published by default [contact: 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."

Should you be planning a Press Release on your article, please get in contact with emboreports@wiley.com as early as possible, in order to coordinate publication and release dates.

Thank you again for your contribution to EMBO reports and congratulations on a successful publication. Please consider us again in the future for your most exciting work.

Kind regards,
Martina

Martina Rembold, PhD
Senior Editor
EMBO reports

THINGS TO DO NOW:

You will receive proofs by e-mail approximately 2-3 weeks after all relevant files have been sent to

our Production Office; you should return your corrections within 2 days of receiving the proofs.

Please inform us if there is likely to be any difficulty in reaching you at the above address at that time. Failure to meet our deadlines may result in a delay of publication, or publication without your corrections.

All further communications concerning your paper should quote reference number EMBOR-2020-51121V3 and be addressed to emboreports@wiley.com.

Should you be planning a Press Release on your article, please get in contact with emboreports@wiley.com as early as possible, in order to coordinate publication and release dates.

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Haiyuan Yu and Marcus Bustamante Smolka

Journal Submitted to: EMBO Reports

Manuscript Number: EMBOR-2020-51121V3

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 χ^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?	SILAC experiments (Figure 4) were all performed in biological replicate
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?	No data was excluded from the quantitative experiments in the manuscript
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.	No
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?	The readers can review the derivation of the significance values derived for Figure 4 (SILAC quantitation) in Dataset EV7. As a default, we assume a small amount of analytical noise should be expected for each quantified value. Readers can eliminate or adjust this value in the table.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	The quantitative data in Figure 4 distributes normally. This can be seen in panels with the violin plots.
Is there an estimate of variation within each group of data?	Yes

USEFUL LINKS FOR COMPLETING THIS FORM

<http://www.antibodypedia.com>
<http://1degreebio.org>
<http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor>

<http://grants.nih.gov/grants/olaw/olaw.htm>
<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm>
<http://ClinicalTrials.gov>
<http://www.consort-statement.org>
<http://www.consort-statement.org/checklists/view/32-consort/66-title>

<http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tum>

<http://datadryad.org>

<http://figshare.com>

<http://www.ncbi.nlm.nih.gov/gap>

<http://www.ebi.ac.uk/ega>

<http://biomodels.net/>

<http://biomodels.net/miriam/>
<http://jij.biochem.sun.ac.za>
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
<http://www.selectagents.gov/>

Is the variance similar between the groups that are being statistically compared?	The variance is roughly similar. See the violin plots in Figure 4.
---	--

C- Reagents

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).	EZview™ Red ANTI-FLAG® M2 Affinity Gel
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	NA

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

D- Animal Models

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

E- Human Subjects

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

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & 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 Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	See Data Availability Section
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).	NA
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).	NA
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 Biocompare (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.	NA

G- Dual use research of concern

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.	No
---	----