THE EMBO JOURNAL

Manuscript EMBO-2017-96717

Carcinogen Susceptibility is Regulated by Genome Architecture and Predicts Cancer Mutagenesis

Pablo E. García-Nieto, Erin Schwartz, Devin A. King, Jonas Paulsen, Philippe Collas, Rafael E. Herrera & Ashby J. Morrison

Corresponding author: Ashby J. Morrison, Stanford University

Review timeline:

Submission date:	12 February 2017
Editorial Decision:	04 April 2017
Revision received:	16 May 2017
Editorial Decision:	07 June 2017
Revision received:	05 July 2017
Accepted:	14 July 2017

Editor: Anne Nielsen

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 04 April 2017

Thank you again for providing us with a preliminary point-by-point response to the concerns raised by our three referees. I have now read it and discussed it with my colleagues in the editorial team as well as consulted with referee #1 on the proposed experiments.

The outcome is that I would like to invite you to submit a revised version of the manuscript in which you address the comments of all three reviewers along the lines discussed in your point-by-point response. However, I would strongly encourage you to also include at least some of the data that you have available to show that the disruption of heterochromatin in the nuclear periphery - be it via deletion of Lamins and histone methyltransferases - affects the regional susceptibility to UV-induced damage. I realize that delineating the full interplay between all factors controlling peripheral chromatin localization may be outside the scope of the current manuscript; however, since both refs #1 and #2 point to the correlative nature of the current data this aspect will have to be strengthened in order for the revised manuscript to be a strong candidate for publication here. The need to move from correlative to causal data was re-emphasised by ref #1 during our consultation based on your point-by-point response, and I would be happy to discuss the exact nature and extent of the data that could be included. I should add that it is EMBO Journal policy to allow only a single round of revision and acceptance of your manuscript will therefore depend on the completeness of your responses in this revised version.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE REPORTS

Referee #1:

García-Nieto and colleagues aim to investigate the relationship of chromatin structure/ organization and susceptibility to DNA damage - in this case induced by u.v. This is an interesting area and the authors have made a promising start. They conclude that there is an elevated mutation rate in heterochromatin - particularly that enriched in H3K9me3 and associated with nuclear lamins. However, the current manuscript suffers from an excess of correlations and a lack of experimental manipulation to test hypotheses. I therefore rate it as preliminary as it stands.

In their experimental work, the authors expose IMR90 fibroblasts to a brief (10s) pulse of uv and then analyze uv-induced cyclobutane pyrimidine dimers (CPDs) genome-wide. Mutations represent the balance between the rates of DNA damage and DNA repair. A concern is therefore the extent to which the author's conclusions about susceptibility to DNA damage - but assayed by measuring mutation rate - is influenced by differential rates of DNA repair and particularly transcription-coupled repair (TCR) in inactive vs active chromatin. They try to mitigate against this by harvesting cells and preparing DNA for immunoprecipitation immediately after the uv exposure. However, TCR may be very rapid and there are insufficient experimental details given about how the cells were harvested to be able to judge the extent to which there might have been the opportunity for some TCR. I have three suggestions for the authors to improve this aspect of their study. Firstly, to repeat their analysis in in human cells with compromised TCR- e.g. from Cockayne or XP patients, or in cell lacking mismatch repair proteins. Secondly, if the authors are correct, and there is an excess of DNA damage per se in the heterochromatin at the nuclear periphery, then this should be revealed by simple immunostaining of damaged cells immediately after uv exposure with antibodies that detect CPDs. Finally, if the heterochromatin at the nuclear periphery has a direct, or indirect role in the differential rates of DNA damage then this pattern of mutation should be affected by the inhibition of H3K9 methylation and by perturbation of organization relative to the nuclear lamina.

Specific comments
1. To get around the problem of biased repair mechanisms, especially transcription-coupled repair, the authors state that they harvested cells 'immediately following a brief (<10s) exposure to uv'. However, there are no details in the Materials and method about the cell harvesting method. There may well be time during the cell harvesting (e.g. cell trypsinization etc) for transcription-coupled repair to take place and therefore to bias their findings.

2. (line 109) "Sequencing replicates are consistent within different bin sizes" and Fig1D. This should be elaborated on within the text, especially the weaker correlations at bin sizes of < 50kb.

3. What is the bin size for calling 'epigenomic states'. It is important in Figure 2A (and page 5) and in the associated analyses, that the bin sizes between mutation rate and epigenetic marks are matched. Epigenomic states are also very cell-type specific and so only epigenetic data from IMR90 cells should be included in this analysis.

4. The conclusion that mutation rates are highest in inactive heterochromatin is consistent with the publication from Prendergast et al., 2007 (BMC Evol. Biol.), who used evolutionary comparisons in neutral sequences to conclude that mutation rates are lowest in open regions of the genome and that regions of the genome with a closed chromatin structure have the highest background mutation rate. This work should be cited.

5. If the authors are correct that H3K9me3 marked heterochromatin at the nuclear lamina is indeed the most susceptible to DNA damage then the authors should test this hypothesis by repeating their analyses in cells treated with BIX- 01294 which both inhibits G9a/Glp (Kubicek et al., 2007, Mol Cell) and displaces lamina-associated domains (LADs) from the nuclear lamina (Kind et al., 2013, Cell).

6. The author's conclusions are consistent with the 'bodyguard hypothesis' of Hsu, 1975 (Genetics

79: 137-150) and this paper should be cited.

Referee #2:

This study uses an elegant approach to map UV-induced photoproducts across the human genome, specifically in IMR90 human fibroblasts. These maps are then combined with epigenetic data and 3D chromosome maps based on Chromosome Conformation Capture (Hi-C) to make the strong correlation that genes near the nuclear lamin are damaged more than genes undergoing transcription or located within the inside of the nucleus. It would appear that for the most part the authors take strong genomic approaches and statistical analyses to analyze their data. The authors then used mutational maps from melanoma to present the case that lesion frequency is what drives mutagenesis. However, since this information is correlational and not causal, many of the sections and phrasing of this manuscript need to be softened. The authors have generated a truly robust data set that will provide the field an important treasure trove of data for years to come. However, the authors over speculated on the causes of mutations based solely on their data set. Responding to the following points will help to strengthen this study. The authors are encouraged to put their data out there without overly hyping the results. Their work will have an important impact on the field.
General questions about the approach:
1. Is damage linear with dose? 100 J/m2 produces saturating number of photoproducts (the point where the reverse reaction becomes important. Clearly this superlethal dose is probably inducing on the order of one photoproduct every kb. The authors should cite several papers to make this lesion frequency clear.
2. How quickly is the DNA harvested after UV? Are the cells UV irradiated on ice? How long after the damage are the cell lysed? This is not a trivial concern as 6-4 photoproducts are rapidly repaired.
3. Why didn't the authors use both CPD and 6-4 photoproduct antibodies to capture the damaged DNA? While the overall yield for CPD to 6-4 photoproducts is 3-5:1, at specific sites 6-4 photoproducts can actually be induced at higher frequencies?
4. How efficient is their photolyase treatment? CC dimers while less frequent are also more difficult to repair - these would pose blocks to sequencing and thus cause an under representation of this lesion.
General comments:
1. The figure legends are too brief and not sufficiently clear for the reader to understand what is being presented. More information needs to be given about what the reader is seeing, how were these data generated, and how the data are being presented.
2. The authors do a poor job of integrating their data with those generated from the Sancar lab on the rate of photoproduct repair across the genome (Genes Dev. 2015 May 1;29(9):948-60. & PNAS E2124-E2133, March 28, 201)6) The authors of the manuscript also use different terms for the epigenetic subdomains given in Figure 2A than those in the Sancar papers. This makes it impossible for someone who is knowledgeable in DNA repair, but not in epigenetic jargon to try to integrate the two data sets. The Sancar laboratory makes a strong case that is the rates of repair that dictate the mutation frequency. None of their recent papers are discussed or even cited. This oversight needs to be corrected. Thus, the authors are strong encouraged to attempt to take their incidence data and combine it with the repair data from the Sancar laboratory - perhaps a simple correlation would be of value?
3. Photofootprinting has been around a long time and was actually used to map transcription factor binding sites to promoters. Thus while nucleosome sparse regions may be expected to not shield from UV it is well known that binding of specific proteins to DNA helps to shield UV damage to DNA.
4. Based on first principles, given the concentration of proteins and DNA within the nucleus (even with a very short pathlength) wouldn't it be expected that the most peripheral region of the nucleus would receive the most damage and due to the absorbance of the UV light? They authors talk about a Hsu, (1975) paper, but a discussion with some biophysicists could quickly help resolve this question. A quick calculation using Beers Law, the DNA and protein concentrations and their extinction coefficients even with a path length of the radius of the nucleus suggests that perhaps as much as 30% of incident UV light would be absorbed before reaching the center of the nucleus.
Specific Comments:
1. Title and Lines 35-36 last sentence of the abstract. The authors are proposing a cause and effect in their wording, yet these data are correlational at best. Clearly the phrase, "... and dicates genome instability in cancer", is not supported. This type of language needs to be softened.

2. Line 58 CPDs may or may not be more cytotoxic see: BMC Cancer. 2005; 5: 135. Which states, "after UVB, In contrast, 6-4PP lesions comprise only 18% of UVB-induced DNA lesions, but account for 70% of the apoptosis". Clearly the authors are misleading the readers.
3. Lines 71-73 - They need to cite the Hanawalt lab for the discovery of transcription-coupled repair and actually the both the Adar etal PNAS 2016 and Hu etal Genes Dev. 2015 May 1;29(9):948-60. From the Sancar lab are pertinent to this point and should be cited in the introduction and then again in the discussion.
4. Line 95 Gerd Pfeifer's lab has been working on this concept for the last decade and using LMPCR had data to support this concept.
5. Figure 5 and again in Figure 8 the authors show a small fraction of their data rendered in Chrom3D incorporates LAD positioning and Chromosome Conformation Capture (Hi-C). While I am not familiar with this software if the rendering is really in 3D and the perspective be altered - an additional supplemental movie which rotates the nucleus around so that the reader has a better view of these data would be a wonderful addition.
6. Lines 250-252. The authors try to conclude a cause from a correlation. If chromatin architecture affects repair rates than this correlational at best and not casually related as the authors are trying to suggest.

Referee #3:

In this manuscript Garcia-Nieto and colleagues present a global analysis to predict susceptibility of genomic regions for UV-induced lesions. They perform extensive correlations with chromatin features and nuclear localization of such loci and identify that while the accessible genome seems overall more protected from DNA lesions the lamina-associated regions highly enriched for heterochromatic marks at the nuclear periphery are more prone to accumulate lesions. Furthermore the authors find that melanoma associated genes are more prone to be affected by UV induced lesions. Overall the authors highlight the importance of subnuclear localization rather than openness in determining susceptibility of genomic regions for UV-induced lesions. However, proper controls are missing in many datasets that underlie several important conclusions in this study. In addition, many key points must be addressed to fully appreciate the findings. Specific points of concern are as follows:

Major points:
- In their entire analysis, authors should critically analyze whether repetitive regions are not leading to any misinterpretations in observations.
- Figure 1D: include datasets that are non-replicates to predict what is indeed for a good correlation of such datasets
- Figure 2A: which cell types have been used in Roadmap studies and which ones for DNA lesion studies? If they originate from different cell lines, one needs to perform validations for accessibility and some key histone marks in the same cellular systems.
- Figure 2A: Is it possible that the DNA lesion enrichment with repressed stages is coming from repeats? If lesions occur randomly in the genome, they would be overrepresented in such regions due to their repetitive nature. In case authors have not do so, they should exclude repetitive elements from their analysis and reconfirm that the association with different chromatin states holds true at distinct loci.
- H3K9ac, H2AZ and replication time with DNA lesions. However, the same graph are presented in the EV Fig 3. In the EV fig 3, others epigenetic marks with DNA lesions are tested, but in the text, these results have not described. Can you explain them?
- Same as above for Lamin B1 association (Figure 3) as this could also be influenced by the repetitive nature of LINE elements.
- Figure 5: since the position data derived for 3D modelling is from a different cellular system, the authors should comment on the generality of such data and maybe provide additional confirmation by using another such dataset in case they exist.
- Figure 6: The authors should plot along 2-3 features that they expect to weakly correlate with mutation rates to suggest what is a randomly expected mutation rate.
- Figure 6: in the part, the authors try to correlate DNA lesions and C>T mutations in melanoma from dataset available. This analysis was performed on 25 tumors only, can you enlarge the cohort?
- In melanoma, the gene most frequently mutated by DNA lesions is TP53 (Hodin et al., 2012), which are localized on the chromosome 17? We can also cite the BRAF and CDKN2A genes, which

are localized on the chromosome 7 and 9 respectively. In their analysis, authors shown correlation between DNA lesions and C>T mutation on the chromosome 13. Could you show the results for the chromosomes containing these genes most frequently mutated by UV in melanoma (TP53, BRAF and CDKNA2)? Do you find a correlation between DNA lesions and C>T mutations for these genes?
- Figure 7: The comparisons with small gene groups such as 98 or 39 genes represent small sets. It is not clear whether the plots presented in A and B are derived from a single comparison of gene groups or the average of e.g. 100 different combinations of random 39 genes. If later is true please state in the text - otherwise consider performing such analysis to show this is a conistent finding in several random comparisons. The expression state as well as locus localization in the nucleus may strongly bias the outcome.
- Figure 8: authors correlate the abundance of DNA lesions in cancer driver genes. Surprisingly, TP53 gene is not the most mutated gene in their anaysis. However, given that 50% of the mutations in TP3 gene are associated with UV-radiation (Hodin et al., 2012), could you explain this discripancy? Same question applies for the CDKN2A gene.

Minor points:
- Figure 1A: replace <30sec with 10sec as this will be correct, remove whole genome sequencing as only IPed material was sequenced
- Figure 1E: The association with TpC is not visible and should either not be highlighted in the text or be presented better (e.g. in a zoomed in version)
- The authors should explain why they chose the dose and duration of treatment shown and maybe provide any titration experiments they might have done to determine the effective dose or cite literature in this regard if they exist.
- Figure 2A: reorder the graph according to active and inactive state (also consider histone patterns in middle panel) and label inplot accordingly. Also label chromatin states on top.
- Figure 2: in the panel C, authors shown that there is a negative correlation between DNase accessibility, H3K9ac, H2AZ and replication time with DNA lesions. However, the same graph are presented in the EV Fig 3. In the EV fig 3, others epigenetic marks with DNA lesions are tested, but in the text, these results have not described. Could you explain them?
- Figure 3B: provide correlations for green, yellow and orange bin.
- The authors should discuss the relevance for transcription coupled repair mechanisms to their findings.

1st Revision - authors' response 16 May 2017

STANFORD
UNIVERSITY

Ashby J. Morrison, Ph.D.
Assistant Professor
Department of Biology

May 15, 2017

Dear Dr. Nielsen,

Thank you for the opportunity to revise and resubmit our manuscript entitled, "**Carcinogen Susceptibility is Regulated by Genome Architecture and Dictates Cancer Mutagenesis**".

We greatly appreciate the Referees' thoughtful and insightful review of our manuscript. In the point-by-point response, I hope you will find satisfactory responses to all comments raised by the Referees. We believe that the manuscript is much improved based on these suggestions.

Revisions to the main text have been highlighted to assist in the review process.

Please feel free to contact me if you need any additional information.

Sincerely,

Referee #1:

García-Nieto and colleagues aim to investigate the relationship of chromatin structure/ organization and susceptibility to DNA damage - in this case induced by u.v. This is an interesting area and the authors have made a promising start. They conclude that there is an elevated mutation rate in heterochromatin - particularly that enriched in H3K9me3 and associated with nuclear lamins. However, the current manuscript suffers from an excess of correlations and a lack of experimental manipulation to test hypotheses. I therefore rate it as preliminary as it stands.

In their experimental work, the authors expose IMR90 fibroblasts to a brief (10s) pulse of uv and then analyze uv-induced cyclobutane pyrimidine dimers (CPDs) genome-wide. Mutations represent the balance between the rates of DNA damage and DNA repair. A concern is therefore the extent to which the author's conclusions about susceptibility to DNA damage - but assayed by measuring mutation rate - is influenced by differential rates of DNA repair and particularly transcription-coupled repair (TCR) in inactive vs active chromatin. They try to mitigate against this by harvesting cells and preparing DNA for immunoprecipitation immediately after the uv exposure. However, TCR may be very rapid and there are insufficient experimental details given about how the cells were harvested to be able to judge the extent to which there might have been the opportunity for some TCR. I have three suggestions for the authors to improve this aspect of their study. Firstly, to repeat their analysis in in human cells with compromised TCR- e.g. from Cockayne or XP patients, or in cell lacking mismatch repair proteins. Secondly, if the authors are correct, and there is an excess of DNA damage per se in the heterochromatin at the nuclear periphery, then this should be revealed by simple immunostaining of damaged cells immediately after uv exposure with antibodies that detect CPDs. Finally, if the heterochromatin at the nuclear periphery has a direct, or indirect role in the differential rates of DNA damage then this pattern of mutation should be affected by the inhibition of H3K9 methylation and by perturbation of organization relative to the nuclear lamina.

Specific comments

1. To get around the problem of biased repair mechanisms, especially transcription-coupled repair, the authors state that they harvested cells 'immediately following a brief (<10s) exposure to uv'. However, there are no details in the Materials and method about the cell harvesting method. There may well be time during the cell harvesting (e.g. cell trypsinization etc) for transcription-coupled repair to take place and therefore to bias their findings.

We have added more details regarding the cell harvest protocol to the manuscript. These changes can be found on lines 354-356 of the Materials and Methods and lines 117-121 of the main text:

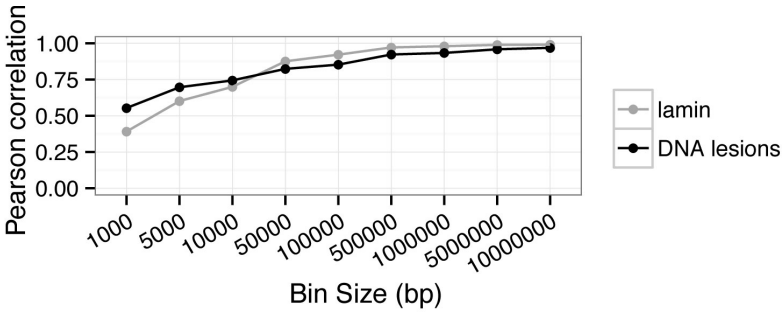
"Less than 10 seconds elapsed from the time of UV exposure to cell lysis with 1% SDS buffer. Repair of CPD lesions is marginally detectable (<5% of all lesions) within one hour following UV exposure (Moser et al, 2005; Verbruggen et al, 2014; Adar et al, 2016). Thus, these experiments were designed to assess immediate CPD formation rather than DNA repair kinetics."

2. (line 109) "Sequencing replicates are consistent within different bin sizes" and Fig1D. This should be elaborated on within the text, especially the weaker correlations at bin sizes of < 50kb.

The correlation plot (Fig 1D) is meant to illustrate the confidence that we have in our data at different window size resolutions. Even smaller bin sizes with a correlation greater than 0.7 is considered to be consistent. We have elaborated on these observations in the main text.

Lines 122-124: "Broad domains of UV lesion abundance were observed with consistency between sequencing replicates at multiple bin sizes from 1 KB to over 1 MB (Fig 1E)."

As a comparison, we include correlations between previously published lamin A sequencing replicates (Lund *et al*, 2014) and our DNA lesion sequencing replicates within the same bin sizes (Figure Below).

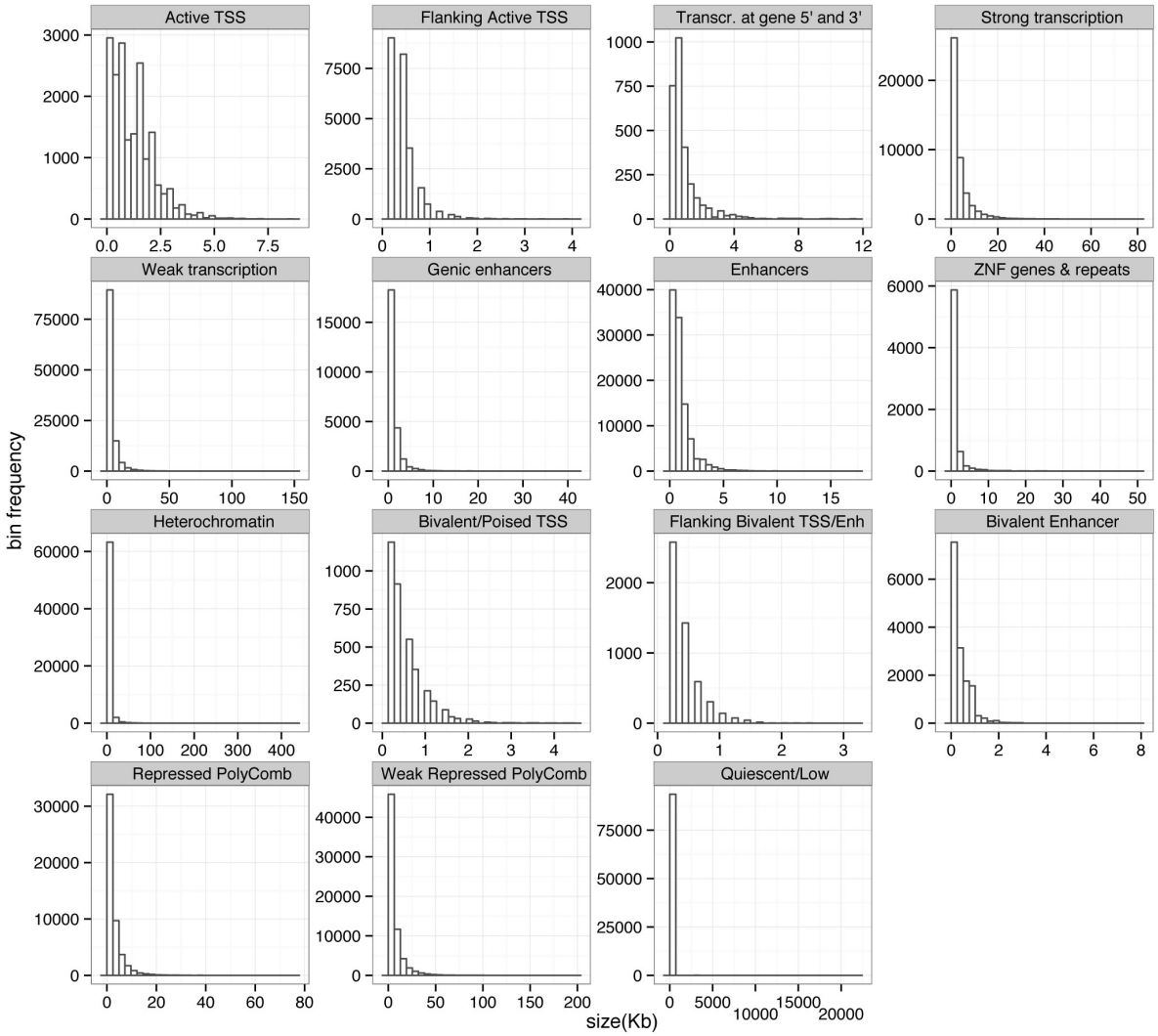


3. What is the bin size for calling 'epigenomic states'. It is important in Figure 2A (and page 5) and in the associated analyses, that the bin sizes between mutation rate and epigenetic marks are matched. Epigenomic states are also very cell-type specific and so only epigenetic data from IMR90 cells should be included in this analysis.

The epigenomic states that we refer to were obtained by the NIH Roadmap Epigenomics Consortium using the ChromHMM program (Ernst & Kellis, 2012). This software was fed chromatin marks measured in IMR90 cells and it partitions the genome into chromatin different states of variable sizes. This is now described in the text:

Lines 396-398: "Bin sizes are variable for each state and were determined by the Roadmap Epigenomics Consortium project (Roadmap Epigenomics Consortium et al, 2015)."

Below is the distribution of bin sizes for each of the 15 states used in Fig. 2. If the Reviewer requests, we can put this in the Expanded View Figures.



The use of IMR90 cells is detailed in the Materials and Methods section and has been added to the figure legend.

Lines 697-698: "Top panel, boxplots of DNA lesion abundance within 15 previously defined chromatin states in IMR90 cells (Roadmap Epigenomics Consortium et al, 2015)."

4. The conclusion that mutation rates are highest in inactive heterochromatin is consistent with the publication from Prendergast et al., 2007 (BMC Evol. Biol.), who used evolutionary comparisons in neutral sequences to conclude that mutation rates are lowest in open regions of the genome and that regions of the genome with a closed chromatin structure have the highest background mutation rate. This work should be cited.

This work has been cited in lines 339-343: *"Moreover, evolutionary comparisons of neutral DNA sequences have also found higher mutation rates in heterochromatin compared to euchromatin (Prendergast et al, 2007), demonstrating that damage in heterochromatin can be maintained during evolution without significant detriment to survival."*

5. If the authors are correct that H3K9me3 marked heterochromatin at the nuclear lamina is indeed the most susceptible to DNA damage then the authors should test this hypothesis by repeating their analyses in cells treated with BIX- 01294 which both inhibits G9a/Glp (Kubicek et al., 2007, Mol Cell) and displaces lamina-associated domains (LADs) from the nuclear lamina (Kind et al., 2013, Cell).

We have performed the suggested experiment and have also utilized another H3K9me3 inhibitor and included the results in the manuscript.

Lines 209-216: "Cells were treated with methyltransferase inhibitors for H3K9me3 (Chaetocin, a Suvar3-9 inhibitor) and H3K9me2 (BIX01294, a G9a methyltransferase inhibitor). These inhibitors alter the formation of LADs (Kind et al, 2013; Kubicek et al, 2007; Illner et al, 2010; Greiner et al, 2005) and as expected, result in decreased H3K9me3 levels (Fig 3E). Strikingly, we further find that either inhibitor also leads to a dramatic reduction in the accumulation of UV lesions (Fig 3F). These results indicate that lamin-associated heterochromatin plays a causal role in carcinogen susceptibility."

Regarding the suggestion of IF to assess the spatial distribution of CPD in the nucleus. This is a technique we have tried to develop in the past but have been unsuccessful due to experimental limitations. This is primarily because the only commercial CPD antibody available recognizes CPDs in single-stranded DNA, not double-stranded DNA (Cosmos Bio). The DNA denaturation step in the IF protocol calls for treatment of cells with 2N HCl for 30 minutes. From our analysis, this dramatically distorts the chromatin structure within the nucleus. Decreasing the concentration or time of HCl helps to preserve chromatin structure, but significantly diminishes CPD signal, thus confounding our ability to use this technique.

Nevertheless, we hope we have compiled sufficient data to suggest that CPDs are enriched in lamin-associated regions at the nuclear periphery. Specifically, we show that lamina-associated domains (LADs), which are abundant at the nuclear periphery (Guelen *et al*, 2008) have elevated DNA lesions (Fig 3 and 4). In addition, chromosomes that have been mapped to the nuclear periphery (Bolzer *et al*, 2005) absorb more DNA lesions than ones in the interior (Fig 5A). Furthermore, using 3D genome modeling we can predict that the regions of the genome most susceptible to acquire DNA lesions are closer to the nuclear periphery (Fig 5B and C). Finally, we have now added the suggested experiments (Fig 3E-F) that demonstrate a causal role for lamin-associated heterochromatin in carcinogen susceptibility.

6. The author's conclusions are consistent with the 'bodyguard hypothesis' of Hsu, 1975 (Genetics 79: 137-150) and this paper should be cited.

This reference is cited on line 326-328 of the manuscript:
"*Alternatively, a "bodyguard hypothesis" has previously been proposed (Hsu, 1975), whereby chromatin at the nuclear periphery absorbs genetic injuries from exogenous sources to 'shield' the nuclear interior."*

Referee #2:

This study uses an elegant approach to map UV-induced photoproducts across the human genome, specifically in IMR90 human fibroblasts. These maps are then combined with epigenetic data and 3D chromosome maps based on Chromosome Conformation Capture (Hi-C) to make the strong correlation that genes near the nuclear lamin are damaged more than genes undergoing transcription or located within the inside of the nucleus. It would appear that for the most part the authors take strong genomic approaches and statistical analyses to analyze their data. The authors then used mutational maps from melanoma to present the case that lesion frequency is what drives mutagenesis. However, since this information is correlational and not causal, many of the sections and phrasing of this manuscript need to be softened. The authors have generated a truly robust data set that will provide the field an important treasure trove of data for years to come. However, the authors over speculated on the causes of mutations based solely on their data set. Responding to the following points will help to strengthen this study. The authors are encouraged to put their data out there without overly hyping the results. Their work will have an important impact on the field.

General questions about the approach:
1. Is damage linear with dose? 100 J/m2 produces saturating number of photoproducts (the point where the reverse reaction becomes important. Clearly this superlethal dose is probably inducing on the order of one photoproduct every kb. The authors should cite several papers to make this lesion frequency clear.

Figure 1A demonstrates the dose dependent response of CPD lesions with UV exposure. We have now included a description of the lesion frequency in the text.

Lines 111-115: "Figure 1A demonstrates the dose dependent accumulation of CPD lesions in response to UV exposure. The 100 J/m^2 dose was chosen for subsequent analysis because it is well below the level of genome saturation and estimated to induce one DNA lesion every 534-672 base pairs given previous quantifications of pyrimidine dimer frequency that assume homogenous genome distribution (van Zeeland et al, 1981; Mitchell et al, 1989; 1991)."

2. How quickly is the DNA harvested after UV? Are the cells UV irradiated on ice? How long after the damage are the cell lysed? This is not a trivial concern as 6-4 photoproducts are rapidly repaired.

We have added more details regarding the cell harvest protocol to the manuscript. These changes can be found on lines 354-356 of the Materials and Methods and lines 117-121 of the main text:

"Less than 10 seconds elapsed from the time of UV exposure to cell lysis with 1% SDS buffer. Repair of CPD lesions is marginally detectable (<5% of all lesions) within one hour following UV exposure (Moser et al, 2005; Verbruggen et al, 2014; Adar et al, 2016). Thus, these experiments were designed to assess immediate CPD formation rather than DNA repair kinetics."

Cells were not irradiated on ice to prevent any effects of cold shock that may alter nuclear architecture. We did not map 6-4 photoproducts, thus our analysis should not be affected by repair.

3. Why didn't the authors use both CPD and 6-4 photoproduct antibodies to capture the damaged DNA? While the overall yield for CPD to 6-4 photoproducts is 3-5:1, at specific sites 6-4 photoproducts can actually be induced at higher frequencies?

We have added text in lines 57-59 to clarify the analysis of CPD versus 6-4PPs: *"CPDs are by far the most abundant UV-induced DNA lesion more causal to mutagenesis and malignant transformation (Jans et al, 2005; Brash, 2015; You, 2001)."*

4. How efficient is their photolyase treatment? CC dimers while less frequent are also more difficult to repair - these would pose blocks to sequencing and thus cause an under representation of this lesion.

Indeed, the photolyase repair is a critical component of the experimental protocol. Thus, this was very rigorously optimized to ensure >95% repair of photoproduct lesions that was determined using slot blot. An example of repair is currently shown in Figure 1. We have added this description to the manuscript.

Lines 121-122: "DNA pyrimidine dimers were then immunoprecipitated and repaired in vitro with >95% repair efficiency before sequencing (Fig 1B-D)."

General comments:

1. The figure legends are too brief and not sufficiently clear for the reader to understand what is being presented. More information needs to be given about what the reader is seeing, how were these data generated, and how the data are being presented.

All figure legends have been reviewed with the expectation that they provide sufficient detail. Please let us know if anything remains unclear.

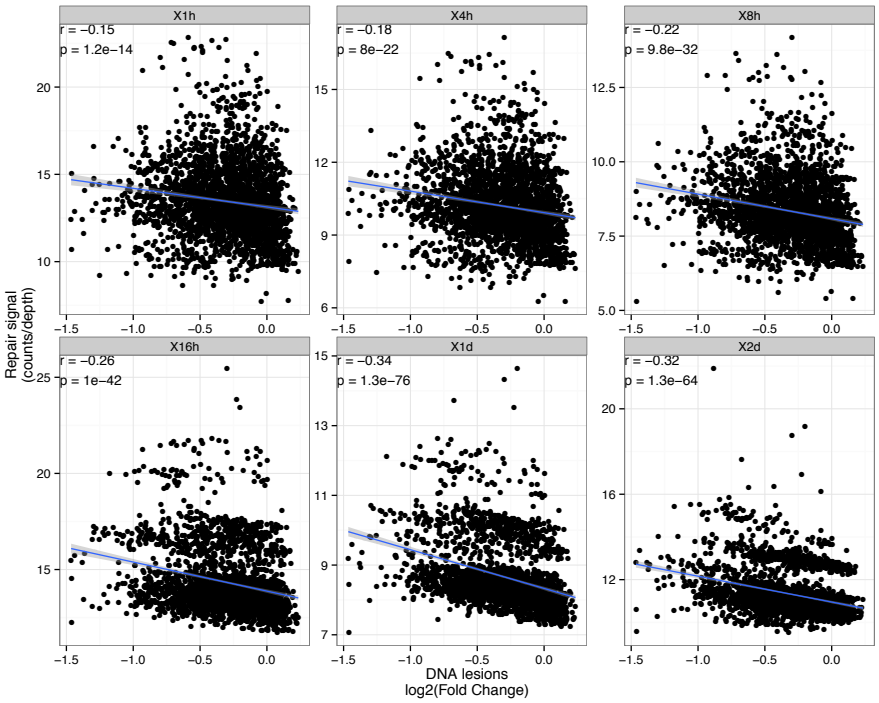
2. The authors do a poor job of integrating their data with those generated from the Sancar lab on the rate of photoproduct repair across the genome (Genes Dev. 2015 May 1;29(9):948-60. & PNAS E2124-E2133, March 28, 201)6) The authors of the manuscript also use different terms for the epigenetic subdomains given in Figure 2A than those in the Sancar papers. This makes it impossible for someone who is knowledgeable in DNA repair, but not in epigenetic jargon to try to integrate the two data sets. The Sancar laboratory makes a strong case that is the rates of repair that dictate the mutation frequency. None of their recent papers are discussed or even cited. This oversight needs to be corrected. Thus, the authors are strong encouraged to attempt to take their incidence data and combine it with the repair data from the Sancar laboratory - perhaps a simple correlation would be of value?

Papers from the Sancar lab used ENCODE data to acquire chromatin states, whereas we used the newer and more comprehensive data from the Roadmap Epigenomics Consortium (Roadmap Epigenomics Consortium *et al*, 2015). As such, there is no one-to-one correspondence between their chromatin states (ENCODE) and ours (Roadmap).

The Sancar papers also used skin fibroblasts whereas we used lung fibroblasts. For this reason, we are hesitant to directly compare our sequencing results to theirs. Nevertheless, we compared XR-seq data from PNAS E2124-E2133, March 28, 2016 with our DNA lesion data and have included the correlations (figure below). Repair signal is shown as XR-seq reads for each timepoint. For example, X1h compares the XR-seq data at 1 hour post UV treatment with our CPD abundance analyses. As can be seen there is a negative correlation for all time points indicating that regions with more CPD accumulation, such as LADs, are repaired with less efficiency.

We have also added a text to the manuscript noting that heterochromatin regions are repaired with less efficiency, thus contributing to elevated mutation frequency in these regions.

Lines 337-339: "Indeed, previous studies have shown that heterochromatin is more refractory to UV-induced repair than euchromatin (Hu et al, 2015; Hanawalt & Spivak, 2008; Adar et al, 2016; Bohr et al, 1985; Mellon et al, 1987)."



3. Photofootprinting has been around a long time and was actually used to map transcription factor binding sites to promoters. Thus while nucleosome sparse regions may be expected to not shield from UV it is well known that binding of specific proteins to DNA helps to shield UV damage to DNA.

We have added text to acknowledge the use of photofootprinting in lines 81-83 of the introduction*: "nucleosomes and DNA bound transcription factors influence susceptibility to UV-induced DNA lesions (Mao et al, 2016; Selleck & Majors, 1987)."*

4. Based on first principles, given the concentration of proteins and DNA within the nucleus (even with a very short pathlength) wouldn't it be expected that the most peripheral region of the nucleus would receive the most damage and due to the absorbance of the UV light? They authors talk about a Hsu, (1975) paper, but a discussion with some biophysicists could quickly help resolve this question. A quick calculation using Beers Law, the DNA and protein concentrations and their extinction coefficients even with a path length of the radius of the nucleus suggests that perhaps as much as 30% of incident UV light would be absorbed before reaching the center of the nucleus.

We agree with the reviewer's assumptions, which is a primary reason the Hsu, 1975 paper was mentioned. Indeed, we could include more theoretical quantitative measurements of UV absorption throughout the nucleus. However, it would be difficult to accurately measure protein and DNA concentrations in different subnuclear compartments. Moreover, because of the differences in absorption spectra among amino acids and the different propensities to acquire UV lesions due to chromatin structure and abundance of DNA binding proteins, absolute quantification of intra-radial nuclear absorption would be exceedingly difficult to predict.

Specific Comments:

1. Title and Lines 35-36 last sentence of the abstract. The authors are proposing a cause and effect in their wording, yet these data are correlational at best. Clearly the phrase, "... and dicates genome instability in cancer", is not supported. This type of language needs to be softened.

We agree with the reviewer, that in the original version of the manuscript the title and last sentence of the abstract overstated the results. As requested by Reviewer 1, we have now included causal data that shows disruption of H3K9me3 and lamin function alters susceptibility to UV lesion accumulation. We hope this new data now supports the sentences in question.

Lines 209-216: "Cells were treated with methyltransferase inhibitors for H3K9me3 (Chaetocin, a Suvar3-9 inhibitor) and H3K9me2 (BIX01294, a G9a methyltransferase inhibitor). These inhibitors alter the formation of LADs (Kind et al, 2013; Kubicek et al, 2007; Illner et al, 2010; Greiner et al, 2005) and as expected, result in decreased H3K9me3 levels (Fig 3E). Strikingly, we further find that either inhibitor also leads to a dramatic reduction in the accumulation of UV lesions (Fig 3F). These results indicate that lamin-associated heterochromatin plays a causal role in carcinogen susceptibility."

2. Line 58 CPDs may or may not be more cytotoxic see: BMC Cancer. 2005; 5: 135. Which states, "after UVB, In contrast, 6-4PP lesions comprise only 18% of UVB-induced DNA lesions, but account for 70% of the apoptosis". Clearly the authors are misleading the readers.

We thank the referee for pointing out this oversight. The sentence in question was meant to read: "The two major classes of mutagenic DNA lesions induced by UV radiation are cis-syn-cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts (6–4PPs*). "CPDs are by far the most abundant UV-induced DNA lesion more causal to mutagenesis and malignant transformation (Jans et al, 2005; Brash, 2015; You, 2001)."* This oversight has been corrected on lines 57-59 of the manuscript.

3. Lines 71-73 - They need to cite the Hanawalt lab for the discovery of transcription-coupled repair and actually the both the Adar etal PNAS 2016 and Hu etal Genes Dev. 2015 May 1;29(9):948-60. From the Sancar lab are pertinent to this point and should be cited in the introduction and then again in the discussion.

These references have been added to the text. We apologize for the not including in the original manuscript.

Lines 71-73: "In addition, transcriptional activity is strongly inversely correlated with mutation frequency, owing to the activity of transcription-coupled repair (Hu et al, 2015; Hanawalt & Spivak, 2008; Adar et al, 2016; Bohr et al, 1985; Mellon et al, 1987)."

Lines 337-339: "Indeed, previous studies have shown that heterochromatin is more refractory to UV-induced repair than euchromatin (Hu et al, 2015; Hanawalt & Spivak, 2008; Adar et al, 2016; Bohr et al, 1985; Mellon et al, 1987)."

4. Line 95 Gerd Pfeifer's lab has been working on this concept for the last decade and using LMPCR had data to support this concept.

A reference to Gerd Pfeifer's contribution to the field has been cited in lines 84-87: *"The non-stochastic distribution of DNA lesions may also influence acquisition of other carcinogen-induced adducts. For example, cytosine methylation at the p53 gene locus correlates with acquisition of lesions formed by benzo(a)pyrene diolepoxide (BPDE), a potent carcinogen in tobacco smoke (Denissenko et al, 1997; 1996)."*

5. Figure 5 and again in Figure 8 the authors show a small fraction of their data rendered in Chrom3D incorporates LAD positioning and Chromosome Conformation Capture (Hi-C). While I am not familiar with this software if the rendering is really in 3D and the perspective be altered -

an additional supplemental movie which rotates the nucleus around so that the reader has a better view of these data would be a wonderful addition.

Movies have been added in the Expanded View section.

6. Lines 250-252. The authors try to conclude a cause from a correlation. If chromatin architecture affects repair rates than this correlational at best and not casually related as the authors are trying to suggest.

As previously mentioned, we hope new data in Figure 3 now supports these statements.

Referee #3:

In this manuscript Garcia-Nieto and colleagues present a global analysis to predict susceptibility of genomic regions for UV-induced lesions. They perform extensive correlations with chromatin features and nuclear localization of such loci and identify that while the accessible genome seems overall more protected from DNA lesions the lamina-associated regions highly enriched for heterochromatic marks at the nuclear periphery are more prone to accumulate lesions. Furthermore the authors find that melanoma associated genes are more prone to be affected by UV induced lesions. Overall the authors highlight the importance of subnuclear localization rather than openness in determining susceptibility of genomic regions for UV-induced lesions. However, proper controls are missing in many datasets that underlie several important conclusions in this study. In addition, many key points must be addressed to fully appreciate the findings. Specific points of concern are as follows:

Major points:
In their entire analysis, authors should critically analyze whether repetitive regions are not leading to any misinterpretations in observations.

Indeed, reads ambiguously mapped to repetitive and low complexity regions can bias ChIP-seq signal. To alleviate the effects of these potential artifacts, all sequencing data was subjected to stringent processing filters as described in the Roadmap Epigenomics Consortium (Roadmap Epigenomics Consortium *et al*, 2015). Specifically, multi-mapping reads were discarded, and remaining uniquely-mapped reads were de-duplicated such that at most one read was retained at each position in the genome. Reads were then filtered by a consensus unique 101-bp mappability track which was obtained from https://github.com/kundajelab/chipseq_pipeline. These steps prevent the over-representation of sequencing reads at repetitive genome regions, and therefore help control for in silico artifacts associated with short-read sequencing of repetitive DNA. The uniquely mapped, de-duplicated, and filtered reads were then input into MACS2.0 to generate the final signal tracks that were used for all downstream analyses.

Figure 1D: include datasets that are non-replicates to predict what is indeed for a good correlation of such datasets.

The correlation plot (Fig 1D) is meant to illustrate the confidence that we have in our data at different window size resolutions. Even smaller bin sizes with a correlation greater than 0.7 is considered to be consistent. We have elaborated on these observations in the main text.

Lines 122-124: "Broad domains of UV lesion abundance were observed with consistency between sequencing replicates at multiple bin sizes from 1 KB to over 1 MB (Fig 1E)."

As a comparison, we include correlations between previously published lamin A sequencing replicates (Lund *et al*, 2014) and our DNA lesion sequencing replicates within the same bin sizes (Figure Below).

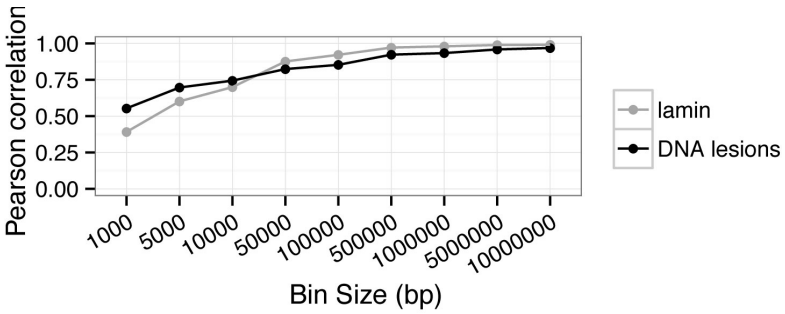


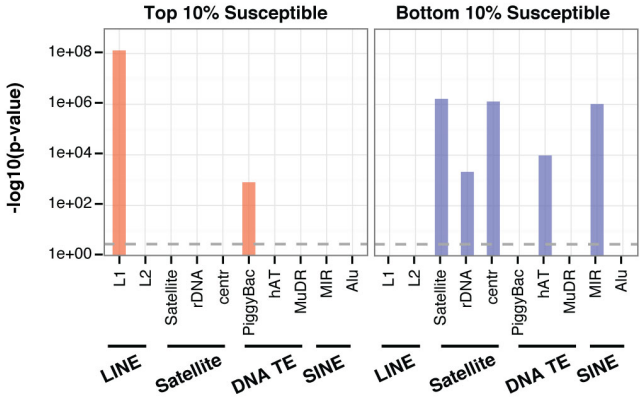
Figure 2A: which cell types have been used in Roadmap studies and which ones for DNA lesion studies? If they originate from different cell lines, one needs to perform validations for accessibility and some key histone marks in the same cellular systems.

IMR90 cells were used for both DNA lesion susceptibility studies and epigenetic marks from the Roadmap Epigenomics Consortium. This has been highlighted in both the figure legend and Materials and Methods.

Figure 2A: Is it possible that the DNA lesion enrichment with repressed stages is coming from repeats? If lesions occur randomly in the genome, they would be overrepresented in such regions due to their repetitive nature. In case authors have not do so, they should exclude repetitive elements from their analysis and reconfirm that the association with different chromatin states holds true at distinct loci.

Please refer to the first comment regarding processing of sequencing data to avoid bias due to repetitive elements.

Figure 8 illustrates how genes located in LADs acquire more lesions on average than those genes located away from LADs. This analysis excludes repeats and shows that our original observations are maintained. We also have analyses that we did not include in the manuscript, showing that not all repetitive elements are enriched in DNA lesions. For example, LINE L1 elements show the greatest enrichment, whereas most of the rest (including L2) show no enrichment or are depleted. If the reviewer wishes, we can include this in the Expanded View section.



UV susceptibility is differential in repeat sequences. The genome was classified as highly (top 10%) or lowly (bottom 10%) susceptible to DNA lesions, using 5kb windows, and then enrichment of different repeat classes within those categories was calculated. Centromere is abbreviated as centro. Half a tetratricopeptide repeat is hAT. Mammalian-wide interspersed repeat is MIR.

H3K9ac, H2AZ and replication time with DNA lesions. However, the same graph are presented in the EV Fig 3. In the EV fig 3, others epigenetic marks with DNA lesions are tested, but in the text, these results have not described. Can you explain them?

Lines 166-171 detail the histone modifications shown in EV Fig 3: *"In addition, genomic regions enriched in euchromatin and enhancer function histone modifications are depleted in UV lesions and include: acetylated H3K4, H3K23, H3K27, and H2AK9; H3K36me3; H4K20me1; and H3K79me1,2,3 (EV Fig 3). Regions with H3K56 acetylation, which is associated with genome stability during replication (Masumoto et al, 2005), have the weakest correlation of all histone marks investigated (r = 0.08)."*

Same as above for Lamin B1 association (Figure 3) as this could also be influenced by the repetitive nature of LINE elements.

Please refer to response above to Figure 2A.

Figure 5: since the position data derived for 3D modelling is from a different cellular system, the authors should comment on the generality of such data and maybe provide additional confirmation by using another such dataset in case they exist.

The 3D modeling was performed with IMR90 Hi-C data, thus is directly applicable to our lesion mapping in IMR90 cells. This has been noted in the figure legend and Materials and Methods.

Figure 6: The authors should plot along 2-3 features that they expect to weakly correlate with mutation rates to suggest what is a randomly expected mutation rate.

The only histone mark that does not have a strong correlation (either positive or negative) with mutation rate is H3K56 acetylation. The genome wide correlation comparing H3K56 acetylation and C>T mutations is *r* =-0.09 (pvalue = <0.0001) and is shown for chromosome 13 below. We can include this figure in the Expanded View section if the Reviewer wishes.

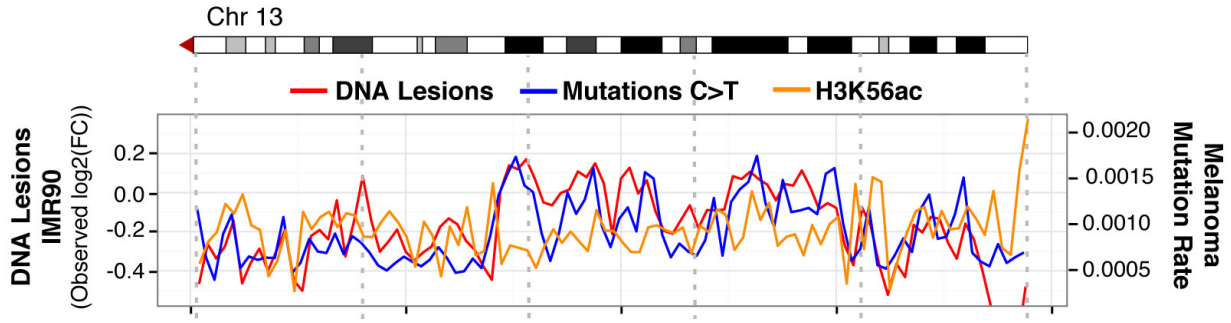


Figure 6: in the part, the authors try to correlate DNA lesions and C>T mutations in melanoma from dataset available. This analysis was performed on 25 tumors only, can you enlarge the cohort?

We utilized all the genome-wide melanoma sequencing we know to be publicly available.

In melanoma, the gene most frequently mutated by DNA lesions is TP53 (Hodin et al., 2012), which are localized on the chromosome 17? We can also cite the BRAF and CDKN2A genes, which are localized on the chromosome 7 and 9 respectively. In their analysis, authors shown correlation between DNA lesions and C>T mutation on the chromosome 13. Could you show

the results for the chromosomes containing these genes most frequently mutated by UV in melanoma (TP53, BRAF and CDKNA2)? Do you find a correlation between DNA lesions and C>T mutations for these genes?

Using our current method of IP'ing DNA lesions on DNA strands of 200-300 bp, we lack single base pair resolution of susceptibility along genic regions. However, in Fig. 8 we plot the relationship between mutation frequency in melanoma and the mean CPD lesion accumulation of cancer driver genes. It should be also noted, that these chromosome maps are modeled in 1Mb bins, thus lack high resolution. We are currently working on a method to perform high resolution single base pair UV lesions mapping. However, this method still requires further optimization.

Figure 7: The comparisons with small gene groups such as 98 or 39 genes represent small sets. It is not clear whether the plots presented in A and B are derived from a single comparison of gene groups or the average of e.g. 100 different combinations of random 39 genes. If later is true please state in the text - otherwise consider performing such analysis to show this is a conistent finding in several random comparisons. The expression state as well as locus localization in the nucleus may strongly bias the outcome.

We thank the Reviewer for the suggestion. The analysis has been recalculated using 1000 different random groups. The figure and figure legend have been updated to reflect the adjusted p-values.

Figure 8: authors correlate the abundance of DNA lesions in cancer driver genes. Surprisingly, TP53 gene is not the most mutated gene in their anaysis. However, given that 50% of the mutations in TP3 gene are associated with UV-radiation (Hodin et al., 2012), could you explain this discripancy? Same question applies for the CDKN2A gene.

The size of the gene name in Fig. 8 represents the mutation frequency among melanomas (i.e. how many tumors have a mutated copy of that gene) and not how many individual mutations exist within a particular gene. This has been clarified in the figure legend.

As previously mentioned, using our current method of IP'ing DNA lesions on DNA strands of 200-300 bp, we lack single base pair resolution of CPD lesions along genic regions.

Minor points:
Figure 1A: replace <30sec with 10sec as this will be correct, remove whole genome sequencing as only IPed material was sequenced

We have corrected the figure as suggested.

Figure 1E: The association with TpC is not visible and should either not be highlighted in the text or be presented better (e.g. in a zoomed in version)

This figure has been updated with the use of box plots to better visualize the differences in IP versus input material.

The authors should explain why they chose the dose and duration of treatment shown and maybe provide any titration experiments they might have done to determine the effective dose or cite literature in this regard if they exist.

Figure 1A demonstrates the dose dependent response of CPD lesions with UV exposure. We have now included a description of the lesion frequency in the text.

Lines 111-115: "Figure 1A demonstrates the dose dependent accumulation of CPD lesions in response to UV exposure. The 100 J/m^2 dose was chosen for subsequent analysis because it is well below the level of genome saturation and estimated to induce one DNA lesion

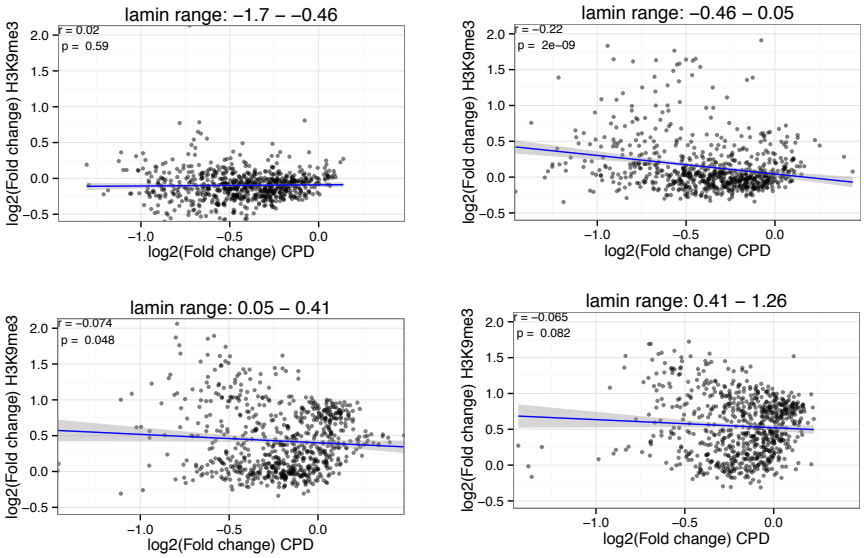
every 534-672 base pairs given previous quantifications of pyrimidine dimer frequency that assume homogenous genome distribution (van Zeeland et al, 1981; Mitchell et al, 1989; 1991).”

Figure 2A: reorder the graph according to active and inactive state (also consider histone patterns in middle panel) and label inplot accordingly. Also label chromatin states on top.

We believe it is easier to visualize propensities of CPD accumulation when ordering the plot from low to high of DNA lesions. However, we have color-coded active (gene-rich, green) and inactive (gene-depleted, red) chromatin states to make the distinction easily discernable. We hope this adjustment is sufficient to achieve the reviewer’s request.

Figure 3B: provide correlations for green, yellow and orange bin.

After dividing the H3K9me3 data based on lamin signal, H3K9me3 correlation with DNA lesion signal is lost (figures below). This proves that the weak correlation we observe between H3K9me3 and DNA lesion signal is primarily driven by lamin B1 abundance. Upon re-evaluation, the results of Fig 3B were not well described in the manuscript. This has now been updated in the revised manuscript.



The authors should discuss the relevance for transcription coupled repair mechanisms to their findings.

These references have been added to the text. We apologize for the oversight.

Lines 71-73: *“In addition, transcriptional activity is strongly inversely correlated with mutation frequency, owing to the activity of transcription-coupled repair (Hu et al, 2015; Hanawalt & Spivak, 2008; Adar et al, 2016; Bohr et al, 1985; Mellon et al, 1987).”*

Lines 337-339: “Indeed, previous studies have shown that heterochromatin is more refractory to UV-induced repair than euchromatin (Hu et al, 2015; Hanawalt & Spivak, 2008; Adar et al, 2016; Bohr et al, 1985; Mellon et al, 1987).”

Adar S, Hu J, Lieb JD & Sancar A (2016) Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proceedings of the National Academy of Sciences* **113:** E2124–33

Bohr VA, Smith CA, Okumoto DS & Hanawalt PC (1985) DNA repair in an active gene: Removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* **40:** 359–369

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR & Cremer T (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* **3:** e157

Brash DE (2015) UV Signature Mutations. *Photochem. Photobiol.* **91:** 15–26

Denissenko MF, Chen JX, Tang MS & Pfeifer GP (1997) Cytosine methylation determines hot spots of DNA damage in the human P53 gene. *Proc Natl Acad Sci USA* **94:** 3893–3898

Denissenko MF, Pao A, Tang M & Pfeifer GP (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science* **274:** 430–432

Ernst J & Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9:** 215–216

Greiner D, Bonaldi T, Eskeland R, Roemer E & Imhof A (2005) Identification of a specific inhibitor of the histone methyltransferase SU(VAR)3-9. *Nat Chem Biol* **1:** 143–145

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W & van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453:** 948–951

Hanawalt PC & Spivak G (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Reviews Molecular Cell Biology* **9:** 958–970

Hsu TC (1975) A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics* **79 Suppl:** 137–150

Hu J, Adar S, Selby CP, Lieb JD & Sancar A (2015) Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev* **29:** 948–960

Illner D, Zinner R, Handtke V, Rouquette J, Strickfaden H, Lanctôt C, Conrad M, Seiler A, Imhof A, Cremer T & Cremer M (2010) Remodeling of nuclear architecture by the thiodioxoxpiperazine metabolite chaetocin. *Exp Cell Res* **316:** 1662–1680

Jans J, Schul W, Sert Y-G, Rijksen Y, Rebel H, Eker APM, Nakajima S, van Steeg H, de Gruijl FR, Yasui A, Hoeijmakers JHJ & van der Horst GTJ (2005) Powerful skin cancer protection by a CPD-photolyase transgene. *Current Biology* **15:** 105–115

Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore WA & van Steensel B (2013) Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153:** 178–192

Kubicek S, O'Sullivan RJ, August EM, Hickey ER, Zhang Q, Teodoro ML, Rea S, Mechtler K, Kowalski JA, Homon CA, Kelly TA & Jenuwein T (2007) Reversal of H3K9me2 by a Small-Molecule Inhibitor for the G9a Histone Methyltransferase. *Mol Cell* **25:** 473–481

Lund E, Oldenburg AR & Collas P (2014) Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Research* **42:** e92

Mao P, Smerdon MJ, Roberts SA & Wyrick JJ (2016) Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings of the National Academy of Sciences* **113:** 9057–9062

Masumoto H, Hawke D, Kobayashi R & Verreault A (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436:** 294–298

Mellon I, Spivak G & Hanawalt PC (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51:** 241–249

Mitchell DL, Jen J & Cleaver JE (1991) Relative induction of cyclobutane dimers and cytosine photohydrates in DNA irradiated in vitro and in vivo with ultraviolet-C and ultraviolet-B light. *Photochem. Photobiol.* **54:** 741–746

Mitchell DL, Vaughan JE & Nairn RS (1989) Inhibition of transient gene expression in Chinese hamster ovary cells by cyclobutane dimers and (6-4) photoproducts in transfected ultraviolet-irradiated plasmid DNA. *Plasmid* **21:** 21–30

Moser J, Volker M, Kool H, Alekseev S, Vrieling H, Yasui A, van Zeeland AA & Mullenders LHF (2005) The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA Repair* **4:** 571–582

Prendergast JGD, Campbell H, Gilbert N, Dunlop MG, Bickmore WA & Semple CAM (2007) Chromatin structure and evolution in the human genome. *BMC Evol. Biol.* **7:** 72

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, et al (2015) Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317–330

Selleck SB & Majors J (1987) Photofootprinting in vivo detects transcription-dependent changes in yeast TATA boxes. *Nature* **325:** 173–177

van Zeeland AA, Smith CA & Hanawalt PC (1981) Sensitive determination of pyrimidine dimers in DNA of UV-irradiated mammalian cells. Introduction of T4 endonuclease V into frozen and thawed cells. *Mutat Res* **82:** 173–189

Verbruggen P, Heinemann T, Manders E, Bornstaedt von G, van Driel R & Höfer T (2014) Robustness of DNA Repair through Collective Rate Control. *PLoS Computational Biology* **10:** e1003438–14

You YH (2001) Cyclobutane Pyrimidine Dimers Are Responsible for the Vast Majority of Mutations Induced by UVB Irradiation in Mammalian Cells. *Journal of Biological Chemistry*

276: 44688–44694

2nd Editorial Decision 07 June 2017

Thank you for submitting a revised version of your manuscript to The EMBO Journal. It has now been seen by all three original referees and their comments are shown below.

As you will see, ref #1 is satisfied with the revision as it is now while refs #2 and #3 ask for a few minor clarifications and additional analysis. The two referees basically point to the same thing: the correlative nature of the data (as we also discussed with ref #1 based on your preliminary point-by-point response). I would therefore invite you to submit a final revised version of the manuscript in which you moderate the title and abstract as suggested by ref #2 and include the additional analysis requested by ref #3. For the latter points, I would encourage you to focus on the control set for mutation rate vs lesion frequency that the referees suggests. The idea of expanding to further analysis of highly mutated regions in melanoma is also interesting but will not be an absolute requirement from our side at this point. I would also ask you to discuss/clarify the additional minor points raised by this referee.

In addition, I have to ask you to address the following editorial issues concerning text and figures:

-> Please move the EV Movie Legends from the main manuscript file to individual word documents. These should preferably be zipped together with the individual EV movie files but we can also do that part in-house.

-> Please rename the manuscript items called 'Additional figure 5 data' and 'Additional Figure 7 and 8 data' to Dataset EV1 and Dataset EV2 and change the callouts in the manuscript text accordingly. In addition, I would ask you to include an EV legend in each of them as an extra tab within the Excel sheet.

-> Please make sure that both the number of replicas used for calculating statistics and the nature of the error bars are indicated in all relevant figure legends

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal, I look forward to receiving your final revision.

REFEREE REPORT

Referee #1:

I am stifled with the revisions that the authors have made to their manuscript and I think that it is now suitable for publication in EMBOJournal.

Referee #2:

This revised manuscript deals with a study that mapped the genomic location of UV-induced cyclobutane pyrimidine dimers. This is the first of its kind, and the authors have shown that there is a strong correlation with increased lesion frequency in IMR90 fibroblasts and mutation frequencies in specific genes associated with melanoma. For the most part the authors have responded well to the concerns raised by each reviewer, including new experimental data and key additions to the wording in the manuscript, as well as, the addition of citations to important previous work. The inclusion of movies that display the 3-dimensional positions of the UV-induced photoproducts is a great addition, as they give a much better view of the spatial distribution within the nucleus.
The analysis of the Sancar repair data with the authors lesion frequency data is very important (even with the caveats raised in the rebuttal) and adding it to the supplemental data would be helpful to the reader.
However, as pointed out by the reviewers, these studies are correlational at best (mapping lesions in lung fibroblasts and comparing to melanoma). The new experimental data supports the hypothesis that heterochromatic regions in LADs is more susceptible to UV-induced CPD, however the correlation to mutagenesis in melanoma is still just that. Thus, the title of the paper implies cause

and effect and since this is not the case needs to be softened appropriately. Also the last line of the abstract is just not a true statement, as pointed out in the first review, several groups (particularly Gerd Pfeifer) have made the argument that the carcinogen lesion frequency at the nucleotide level is well correlated with mutagenesis, the new layer the authors have added is the nuclear architecture. This line needs to be better written to reflect this point.
On lines 274-276, the authors state: "Collectively, these results suggest that the origin of mutational heterogeneity in melanoma genomes is largely specified by the intrinsic properties of carcinogen susceptibility regulated by epigenome architecture." This either needs to be softened or the new information regarding the Sancar repair data mentioned here. The problem is the phrase, "largely specified" which is misleading.
Again in the discussion lines 337-343 the authors have an opportunity to mention the Sancar repair data in this context. This should be added.

Referee #3:

- general summary and opinion about the principle significance of the study, its questions, and findings
By mapping the UV-induced photoproducts across the human genome the authors provide the first genome-wide distribution map of early target loci of UV in IMR90 human fibroblast what provides a useful resource to the research community for further mechanistic studies. Their data suggest that transcribed chromatin is overall less susceptible to UV-induced damage while loci within H3K9me3 marked heterochromatin as well as close to the nuclear lamina are more prone to accumulate such DNA alterations. This is a significant finding that will have to be proven by other researchers in different cellular background in order to see whether this is a general phenomenon upon UV exposure. Furthermore the authors correlate lesion frequency to mutation frequency in melanoma and conclude that lesion frequency is causal for mutations that ultimately lead to melanoma. As this is averaged over a huge region (entire right arm of chromosome 13) and real fluctuations in the mutational rate are not recapitulated in the lesion rate it is difficult to judge the relevance of the correlation (see below). The authors also relate to melanoma genes but they are not further specified in nature and seam to represent many genes (figure 7) and cancer driver genes that are also not specified further (figure 8). Overall the authors follow a very interesting question and present a useful resource dataset to the community but the conclusions drawn on the correlations are not validated to the best extend and still lack important controls to be fully convincing.

- specific major concerns essential to be addressed to support the conclusions
In figure 6 the authors predict lesion rate in melanoma cells based on chromatin marks and correlate to the mutation rate and also correlate real lesion rate in IMR90 fibroblasts to mutation rate. It will be very much needed to have some control sets in which you will not expect a correlation to melanoma mutation rate in order to judge the relevance of the correlation. For example, the regions that are not highly mutated still have high predicted lesion frequency i.e. the up and down within the graphs is not following what suggests that the precise region will never be predictable just based on the distribution of lesions as we have the entire right arm of chromosome 13 plotted. Furthermore to prove the clinical relevance of their data the authors could look at the most highly reported mutated regions for melanoma and repeat their analysis also for this subset of regions as well as for a subset of never mutated regions.
In figure 8 the authors conclude that cancer driver genes have increased UV susceptibility. First, how does this fit their observations of UV targeting more heterochromatic and distal regions? The authors should provide genome browser tracks for accessibility and H3K9me3 of these genes in precursor cells in comparison to non-damaged loci. Furthermore, a comparison to the average of all other genes is misleading. A box plot of lesion frequency of top 20 drivers should be compared to random sets of other 20 genes with respect to lesion frequency, H3K9me3, accessibility and nuclear localization (partly addressed in figure 7).

- minor concerns that should be addressed
In figure 7 it is not clear why there is such a dramatic number of genes in several bins. How are melanoma genes defined? If there is e.g. 16987 in bin 2?
In figure 8 it would be helpful to explain how driver genes have been selected as e.g. Myc is not mentioned.

- any additional non-essential suggestions for improving the study (which will be at the author's/editor's discretion)
N/A

2nd Revision - authors' response 05 July 2017

Dear Dr. Nielsen,

Thank you for the opportunity to revise and resubmit our manuscript entitled, "Carcinogen **Susceptibility is Regulated by Genome Architecture and Predicts Cancer Mutagenesis".**

We greatly appreciate the Referees' thoughtful and insightful review of our manuscript. In the point-by-point response, I hope you will find satisfactory responses to all comments raised by the Referees. We believe that the manuscript is much improved based on these suggestions.

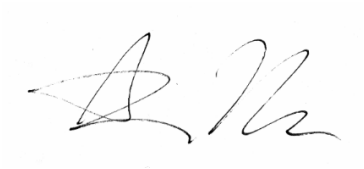
In addition, we have performed the requested tasks, listed below:
-> Please move the EV Movie Legends from the main manuscript file to individual word documents. These should preferably be zipped together with the individual EV movie files but we can also do that part in-house.

-> Please rename the manuscript items called 'Additional figure 5 data' and 'Additional Figure 7 and 8 data' to Dataset EV1 and Dataset EV2 and change the callouts in the manuscript text accordingly. In addition, I would ask you to include an EV legend in each of them as an extra tab within the Excel sheet.

-> Please make sure that both the number of replicas used for calculating statistics and the nature of the error bars are indicated in all relevant figure legends

Please feel free to contact me if you need any additional information.

Sincerely,



Referee #1:

I am stifled with the revisions that the authors have made to their manuscript and I think that it is now suitable for publication in EMBOJournal.

Referee #2:

This revised manuscript deals with a study that mapped the genomic location of UV-induced cyclobutane pyrimidine dimers. This is the first of its kind, and the authors have shown that there is a strong correlation with increased lesion frequency in IMR90 fibroblasts and mutation frequencies in specific genes associated with melanoma. For the most part the authors have responded well to the concerns raised by each reviewer, including new experimental data and key additions to the wording in the manuscript, as well as, the addition of citations to important previous work. The inclusion of movies that display the 3-dimensional positions of the UV-induced photoproducts is a great addition, as they give a much better view of the spatial distribution within the nucleus.

The analysis of the Sancar repair data with the authors lesion frequency data is very important (even with the caveats raised in the rebuttal) and adding it to the supplemental data would be helpful to the reader.

- *We have added this data to Expanded View Fig 7.*

However, as pointed out by the reviewers, these studies are correlational at best (mapping lesions in lung fibroblasts and comparing to melanoma). The new experimental data supports the hypothesis that heterochromatic regions in LADs is more susceptible to UV-induced CPD, however the correlation to mutagenesis in melanoma is still just that. Thus, the title of the paper implies cause and effect and since this is not the case needs to be softened appropriately.

- *In the title, we have changed the word "dictates" to "predicts" to reflect the utility of CPD lesions in the computational prediction of cancer mutations (Fig 6b and EV Fig 6).*

Also the last line of the abstract is just not a true statement, as pointed out in the first review, several groups (particularly Gerd Pfeifer) have made the argument that the carcinogen lesion frequency at the nucleotide level is well correlated with mutagenesis, the new layer the authors have added is the nuclear architecture. This line needs to be better written to reflect this point.

- *We have altered the last line of the abstract and changed "dictates" to "mirrors".*

On lines 274-276, the authors state: "Collectively, these results suggest that the origin of mutational heterogeneity in melanoma genomes is largely specified by the intrinsic properties of carcinogen susceptibility regulated by epigenome architecture." This either needs to be softened or the new information regarding the Sancar repair data mentioned here. The problem is the phrase, "largely specified" which is misleading.

- *This sentence has been changed as follows: "Collectively, these results suggest that the origin of mutational heterogeneity in melanoma genomes is significantly contributed by the intrinsic properties of carcinogen susceptibility, which is regulated by epigenome architecture."*

Again in the discussion lines 337-343 the authors have an opportunity to mention the Sancar repair data in this context. This should be added.

- *The XR-seq studies from the Sancar lab have been added.*

Referee #3:

- general summary and opinion about the principle significance of the study, its questions, and findings

By mapping the UV-induced photoproducts across the human genome the authors provide the first genome-wide distribution map of early target loci of UV in IMR90 human fibroblast what provides a useful resource to the research community for further mechanistic studies. Their data suggest that transcribed chromatin is overall less susceptible to UV-induced damage while loci within H3K9me3 marked heterochromatin as well as close to the nuclear lamina are more prone to accumulate such DNA alterations. This is a significant finding that will have to be proven by other researchers in different cellular background in order to see whether this is a general phenomenon upon UV exposure. Furthermore the authors correlate lesion frequency to mutation frequency in melanoma and conclude that lesion frequency is causal for mutations that ultimately lead to melanoma. As this is averaged over a huge region (entire right arm of chromosome 13) and real fluctuations in the mutational rate are not recapitulated in the lesion rate it is difficult to judge the relevance of the correlation (see below). The authors also relate to melanoma genes but they are not further specified in nature and seam to represent many genes (figure 7) and cancer driver genes that are also not specified further (figure 8). Overall the authors follow a very interesting question and present a useful resource dataset to the community but the conclusions drawn on the correlations are not validated to the best extend and still lack important controls to be fully convincing.

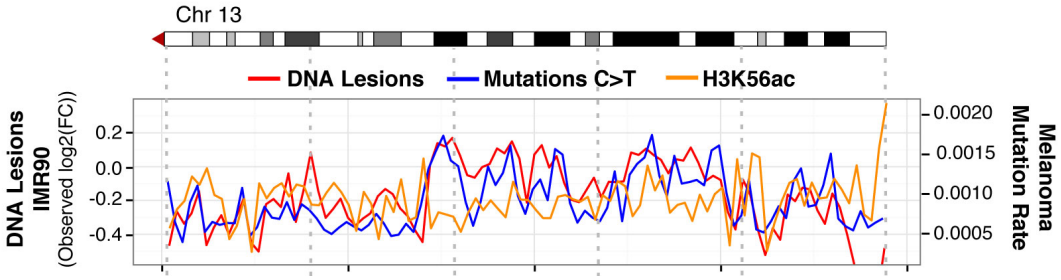
- specific major concerns essential to be addressed to support the conclusions

In figure 6 the authors predict lesion rate in melanoma cells based on chromatin marks and correlate to the mutation rate and also correlate real lesion rate in IMR90 fibroblasts to mutation rate. It will be very much needed to have some control sets in which you will not expect a correlation to

melanoma mutation rate in order to judge the relevance of the correlation. For example, the regions that are not highly mutated still have high predicted lesion frequency i.e. the up and down within the graphs is not following what suggests that the precise region will never be predictable just based on the distribution of lesions as we have the entire right arm of chromosome 13 plotted.

- *This comment was addressed in the previous response, which is pasted below:*

- "Figure 6: The authors should plot along 2-3 features that they expect to weakly correlate with mutation rates to suggest what is a randomly expected mutation rate."
- The only histone mark that does not have a strong correlation (either positive or negative) with mutation rate is H3K56 acetylation. The genome wide correlation comparing H3K56 acetylation and C>T mutations is r =-0.09 (pvalue = <0.0001) and is shown for chromosome 13 below.



Furthermore to prove the clinical relevance of their data the authors could look at the most highly reported mutated regions for melanoma and repeat their analysis also for this subset of regions as well as for a subset of never mutated regions.

- *This data is already presented in Figure 7. The leftmost box plots compare enrichment of UV-induced DNA lesions in genes that are not mutated in melanoma.*

In figure 8 the authors conclude that cancer driver genes have increased UV susceptibility. First, how does this fit their observations of UV targeting more heterochromatic and distal regions?

- *Distance to LAD (a marker of the nuclear periphery) is shown for each gene as dots under the gene name. In addition, 3 cancer driver genes and their positioning is shown in Fig.8b.*

The authors should provide genome browser tracks for accessibility and H3K9me3 of these genes in precursor cells in comparison to non-damaged loci. Furthermore, a comparison to the average of all other genes is misleading. A box plot of lesion frequency of top 20 drivers should be compared to random sets of other 20 genes with respect to lesion frequency, H3K9me3, accessibility and nuclear localization (partly addressed in figure 7).

- Lamin association is better correlated with susceptibility and melanoma mutations than H3K9me3 (Figs 3 and Fig7), thus we focused on highlighting lamin association with the cancer driver genes. Lamin association is shown in Fig. 8b. Also, the statistical power of the whole genome is stronger than 20 genes, in providing a reference point for susceptibility.

- minor concerns that should be addressed

In figure 7 it is not clear why there is such a dramatic number of genes in several bins. How are melanoma genes defined? If there is e.g. 16987 in bin 2?

- *As stated in the Methods and figure legend, mutation frequency for **all** genes was determined using the COSMIC database. Bin 2, 0.1-5% mutation frequency, has the largest number of genes.*

In figure 8 it would be helpful to explain how driver genes have been selected as e.g. Myc is not mentioned.

- Myc is not characterized as a melanoma cancer driver. The reference (Gonzalez-Perez et al, 2013) provided in the figure legend and Methods describes how cancer driver genes were determined.

- any additional non-essential suggestions for improving the study (which will be at the author's/editor's discretion)
N/A

3rd Editorial Decision 14 July 2017

Thank you for submitting the final revision, I am pleased to inform you that your manuscript has now been officially accepted for publication in The EMBO Journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name:	Ashby J. Morrison
Journal Submitted to:	EMBO Journal
Manuscript Number:	EMBOJ-2017-96717

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

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

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

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

2. Captions

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

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ 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.

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

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

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?	N/A
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	N/A
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	N/A
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.	N/A
For animal studies, include a statement about randomization even if no randomization was used.	N/A
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.	N/A
4.b. For animal studies, include a statement about blinding even if no blinding was done	N/A
5. For every figure, are statistical tests justified as appropriate?	For each figure, statistical tests are described in figure legend or materials and methods.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	For each statistical test, the data meets the assumption criteria.
Is there an estimate of variation within each group of data?	Data distribution is shown as scatter plots or box plots, where appropriate.
Is the variance similar between the groups that are being statistically compared?	Samples that are being statistically compared have similar variance.

C- Reagents

USEFUL LINKS FOR COMPLETING THIS FORM

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

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-tun

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://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

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).	Anti-CPD antibody (Cosmos Bio, clone TDM-2), anti- H3K9me3 (Abcam, cat. no. 8898)
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	IMR90 primary cells were obtained from Coriell Repository.

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

E- Human Subjects

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

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Sequencing data is available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94434 under accession number GSE94434.
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).	N/A
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).	N/A
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	N/A
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	N/A

G- Dual use research of concern

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