The chromatin landscape of the euryarchaeon *Haloferax volcanii*
Response to reviewer comments

```
Reviewer Expertise:

Referee #1: archaea cell and molecular biology, genetics, extremophiles
Referee #2: archaeal chromatin, cell and molecular biology
Referee #3: archaeal histones, genetics, extremophiles
```

Reviewer 1:

> The study by Marinov et al. identified the regions of the Haloferax volcanii genome accessible to several different chemical and enzymatic probes at single-nucleotide resolution. The analyses, along with patterns and variations emerging from them, are described in an engaging style. New and informative molecular properties of this organism are less obvious, and the value for the target audience (microbiologists) has important limitations, as described below.

We thank the reviewer for the comments.

> The radically different molecular environments in Haloferax cells vs. eukaryotic nuclei do not seem to have factored significantly in setting up the analyses. Aside from demonstrating that formaldehyde cross-linking is essential to detect significant levels of DNA protection, it seems that no functional impacts of salt, formaldehyde, or other components carrying over into the various probing reactions were considered. Similarly, basic information such as the chemical composition of the DNA-probing reactions, principles of the techniques (how unprotected DNA is marked and mapped) or the influence of the DNA itself is not provided.

The ATAC reactions are not happening in the original media, but in ATAC buffer. Thus salt does not really play a major role. Neither formaldehyde crosslinking, not kethoxal labeling are expected to be influenced by the presence of salt in the media.

We have provided expanded explanations regarding how the assays work in the revised version of the text.

> Similarly, the observed genomic patterns of DNA that is accessible under the probing conditions are rarely interpreted with respect to what is known about halophilic archaea. Several genome-scale analyses of transcription in Haloferax and related species or other archaea (DOI: 10.3390/genes12071018; DOI: 10.1371/journal.pone.0215986; DOI: 10.1080/15476286.2020.1723328; DOI: 10.1186/s12864-016-2920-y) were not considered, despite the relevant data they provide. Similarly, the study made no attempt to relate any observed pattern of accessibility to the local sequence, despite the potential relevance for chromatin proteins (see Nalabothula et al, 2013).

Our study focuses on chromatin and active transcription. We were aware of the studies listed by the reviewer, but they did not in fact provide us with as much relevant information as it might seem as they measured steady-state transcription levels using various forms of RNA-seq, while our focus with KAS-seq was on active transcription.

We have related sequence to accessibility in the revised version of the manuscript

> Especially revealing is the lack of concern that after three months at room temperature the majority of cells in a culture might be dead and thus irrelevant for chromatin analysis.

Only one experiment was performed under such conditions, and it is how we made the discovery of the strong single-stranded signal around CRISPR arrays. That was a curiosity experiment, and part of the very first KAS-seq we ever did – we wanted to see what it would look like under such conditions. After we made that discovery based on that experiment, we had the choice of either honestly reporting how we made the discovery, or not being honest about it. We are of the very strong opinion that the former is the right way to go, for scientific integrity's sake.

Chromatin analysis on standing cultures was carried out a week after active growth, and was necessitated by the 3-month KAS experiment and the need to honestly report it.

Several terms or phrases seem to have meanings unique to the authors or are otherwise unclear or confusing in their context. These include "gene body" (used throughout the manuscript), "individual promoters" (p.5: does this genome have group promoters?), "uniform KAS-seq profile" (p.8: uniform in what sense? How is this significant?), "typical more symmetrical changes" (p. 8: in what sense are any of the changes symmetrical?). With respect to Figure 7, it is not clear what correlation or lack thereof can be inferred simply by looking at these heat maps.

Many of these terms are strandard in the genomics literature. We have revised the manuscript to make it more clear for people not familiar with it.

The distinctions on p. 5 regarding "bulk methods", "absolute occupancy", "single-molecule" method, and measuring "the act of active transcription" seem particularly non-sensical. According to the Methods, all data presented in this paper come from populations of DNA molecules extracted from multiple cells. The reported occupancy is sequence-specific, but it is measured by comparing the number of reads mapped to that position to the numbers mapped to other sites, not by limiting the analysis to one DNA molecule. Similarly, the results are not time-resolved; as with other whole-genome methods, they provide a global assessment at a given time, not a temporal sequence of events. Also, one could argue that analyzing the RNAs from live cells (RNAseq) gives a more direct view of transcription than mapping accessible sites in DNA from fixed cells.

XXX HAVEN'T RESPONDED TO THIS AS THIS IS A WHOLE LOT OF OUT-OF-DOMAIN MISUNDERSTANDING AND THIS ISN'T AN ACTUAL RESUBMISSION XXX.

Reviewer 2:

In this manuscript the authors describe studies of chromatin organization using genomic tools in the eurarchaeote Haloferax volcanii. The manuscript is timely in that it addresses a topic of increasing interest in the field. More and more studies on in vivo organization of archaeal chromatin have recently seen the light, but much remains to be understood. Apart from relevance to chromatin researchers in the archaeal field, the work is also of potential interest to researchers interested in eukaryotic chromatin, due to the recently re-enforced close evolutionary relationship between archaea and eukaryotes, and the notion that in context of chromatin organization, histone proteins are conserved.

We thank the reviewer for the helpful comments and suggestions.

1. The abstract says, "These results provide a foundation for the future functional studies of archaeal chromatin." Is this the main finding of this work that it forms a foundation for future work? This manuscript is data-rich, but is lacking in extracting meaningful information that is of biological significance. The submitted manuscript includes ATAC-seq, KAS-seq, and NOMe-Seq/dSMF profiles of up to 7 different conditions. The quality of the downstream analysis performed in this paper is poor in light of the potential that the data submitted with this manuscript has.

This is the first time the multiple genomics methods used in the paper have been adapted and optimized to and tehn applied to archaea. It was not the goal of the study to dive deep into specific changes in accessibility relating to a particular biological context, but to make global observations about the structure of *Haloferax* chromatin.

2. ATAC-seq experiments described in this manuscript involve using formaldehyde to freeze the nucleosomes in place. The TSS ratio improves when ATAC-seq is performed with 0.1% formaldehyde, and improves further with 1% formaldehyde. Were higher concentrations of formaldehyde tested before a choice was made to proceed with a final concentration of 1%?

We did not test higher concentrations because 1% FA is already a very strong fixation condition and it provided a signal-to-noise ratio comparable to that usually seen in yeast (a conventional eukaryote). It is possible that e.g. 4% FA would provide an additional boost, but that might also introduced artifacts (past experience has shown that it does have that effect in the context of ChIP-seq).

3. Figure 2 shows the NOMe-Seq and dSMF metaprofiles for exponential phase, but only the NOMe-Seq profile for the stationary phase. What does the dSMF metaprofile for stationary phase look like?

We only carried out an SMF experiment (with both CpG and GpC enzymes) in exponential cells.

4. Figure 2 shows single molecule DNA protection/methylation maps. The only description of these maps in the manuscript is from an ensemble perspective and the 'single molecule' information provided by these experiments has been ignored. For instance, the single molecule maps show that in some cells, the region that is examined is completely protected, while in others it is partially protected or unprotected. The authors should use these maps to extract information about different H. volcanii populations in culture.

The main point of the single-molecule analysis is that distinct populations of single-molecules are not readily identifiable (they are in eukaryotes). Protection states are only visible on very localized scales, i.e. several tens of base pairs but not on scales of several hundred base pairs.

5. The authors suggest that "...large-scale domains might be a feature associated with lack of nucleosomal chromatin in prokaryotes..." I disagree with this suggestion. The organisation of the chromosome does not depend on the protein that organises DNA but how it organises the DNA. For instance, Sulfolobus acidocaldarius lacks histone proteins, and therefore, has a non-nucleosomal chromosome. But, its chromosome contact map shows A/B compartmentalisation – a feature characteristic of eukaryotic (nucleosomal) chromosomes. ATAC-seq data should not be used to directly comment on the domain organisation of the chromosome. The chromatin accessibility data would be better interpreted in light of chromosome conformation capture maps.

The phrase "large-scale domains" does not refer to interaction domains, but to domains of elevated accessibility/expression. We have corrected manuscript in that section to better clarify that point.

6. The authors say in the results, "We also reexamined the Sulfolobus islandicus dataset and found that it displays a much more modest TSS enrichment than that seen in H. volcanii..." The TSS enchrichment value is 1.25 ± 0.1 in H. volcanii, and 1.11 in Sulfolobus islandicus. Since the error in the TSS enrichment value for Sulfolobus islandicus has not been provided, it is difficult to evaluate the significance or insignificance of this value, and hence the validity of the statement in the results.

There is only a single dataset provided in the *Sulfolobus* study.

7. Supplementary figure 1 shows ATAC-seq metaprofiles of the region around the TSS in Haloferax volcanii. The graph shows that in fixed cells, the TSS is not protected, but it is protected in unfixed cells. On the basis of this, does formaldehyde fixation introduce artefacts? Especially when considering the ATAC-seq metaprofile of gDNA SRR18275628.

Supplementary Figure 1 shows metaprofiles for *Sulfolobus* from an externally published dataset that we have no control over, not for *Haloferax volcanii*.

We have considered the possibility that fixation might produce artifacts, but consider that unlikely, because it produces enrichment precisely around TSSs (and not in a way that can be explained by sequence composition; see further below), and the detailed profile actually matches the previously published MNase-seq dataset as the inverse of it (as it should).

8. In describing the advantages of the KAS-seq assay in studying active transcription, the authors write, that KAS-seq "...also identifies other ssDNA structures, such as those resulting from paused polymerase molecules, G-quadruplexes, and others." While discussing the strong KAS-seq signal upstream of Cas6 in the CRISPR2 array, the authors do not explain why they consider the signal to have arisen from paused RNA polymerase and not from other factors.

We did not make the claim that we are confident that the Cas6 peak is the result of paused polymerase. This was discussed as one of a set of possibilities, which also includes other ssDNA structures.

9. The authors studied Haloferax volcanii cultures in exponential phase, in stationary phase, and that were standing at room temperature for 1 week and for 3 months. They also studied H. volcanii cultures growing at 42 oC, 37 oC, and 23 oC, and on cultures that were provided with a cold-shock at 4 oC. The authors do not describe the motivation for studying these conditions. Also, they do not properly describe their observations. Even basic comments such as general trends in the KAS-Seq, ATAC-Seq, and NOMe-Seq/dSMF profiles in terms of the effect of temperature, and growth phase are missing.

As explained above, the 3-month condition was a serendipitous experiment that resulted in an unexpected but important discovery, and thus had to be honestly discussed.

The other conditions were picked with the aim of profiling chromatin and transcription under differential conditions, with what these conditions are precisely not being important.

10. Figure 3E shows a KAS-seq metaprofile of the H. volcanii genes. This is not an appropriate representation since genetic elements differ in their organisation. A profile that does not distinguish between the TSS and TTS of convergent, divergent, and directional genes may supress the characteristic features associated with these organisations. Note that directional genes can also be segregated on the basis of whether they are expressed in the direction of replication or against the direction of replication.

This is the standard and most compact way of presenting metaprofiles in the field. Also, "convergent" genes in the context of *Haloferax* are a mix of actually convergent genes and operons, which confounds the interpretation of that category. The point of the figure was that KAS-seq signal is enriched in TSSs rather than peak inside gene bodies.

11. The 42 oC and exponential phase samples are biological replicates. The profiles of these replicates do not always match (See figure 5, and supplementary figures 2A, B, D, E, and H).

These experiments were carried out several months apart, not side by side as proper biological replicates (those do agree with each other). This is simply the variation that is observed between different cultures.

12. The authors correlate chromatin accessibility as detected with ATAC-seq with ssDNA prevalence detected with KAS-seq. They find the second CRISPR array to be interesting and present the corresponding data in figure 4C. The authors say, in relation to this, "Curiously, only the second CRISPR array in H. volcanii displays this strong ssDNA structure, while the other two do not, but all three arrays show elevated chromatin accessibility in ATAC-seq datasets, which is not focused on the beginning of the array but covers its whole length (Figure 4C)". I disagree with this conclusion. The authors compare the KAS-seq profile of a 3-month standing culture and the ATAC-seq profiles of exponential phase, stationary phase, and a 1-week standing culture. Given the differences the authors observe between these four cultures (compare the profiles in figure 4B, 5, 7, and supplementary figure 2), the authors should correlate the KAS-seq and ATAC-seq profiles of the 3-month standing culture to draw appropriate conclusions.

We do not compare the KAS-seq profile of the 3-month standing culture with ATAC-seq from exponential and stationary conditions. The 3-month standing culture dataset is where the discovery was made, but ATAC-seq was also generated (from the very same cultures) for all other conditions, and we see the same pattern there – a KAS-seq peak on CRISPR array 2 and no KAS-seq peak on the other two CRISPR arrays. We have added a clarification on that point in the text.

13. How do the KAS-seq/ATAC-seq profiles of the 3 month and 1 week standing cultures compare to the exponentially growing culture of Haloferax volcanii at 23 oC? Conceptually, the 3-month and 1-week cultures standing at room temperature are late/early stationary phase cultures of the 23 oC cultures in exponential phase.

We did not generate ATAC-seq for the 3-month culture and the 23C one.

Minor comments and Text corrections:

1. One of the lines in the abstract reads, "... modulation of transcriptional activity is not associated with changes in promoter accessibility...". The immediate question I have – having read only the abstract – is what is it associated with, then? That gives an impression of the abstract being incomplete. This sentence should be re-written.

This is what we observe. The question what is associated with is an open one, but its answer will have to be given in the future.

2. The abstract has the phrase, "We also evaluate the degree of coordination of transcription..." What does this phrase mean? What exactly was evaluated? What is 'degree of coordination'?

Operon genes are expected to be transcribed at the same levels, if they are a true operon.

3. One of the statements in the abstract reads, "... archaeal nucleosomal chromatin is similarly or slightly more accessible, in aggregate, than that of eukaryotes." What is the significance of this finding?

The a priori expectation was for a much looser structure than in eukaryotes.

4. In the introduction: "Based on these and other similarities...", the phrase 'other similarities' is vague. The statement would benefit from mentioning some of these similarities.

We have removed the "and other" phrase.

5. A statement in the introduction says, "We also examine the coordination of transcriptional activity and chromatin accessibility within Haloferax operons, and make the unexpected discovery that some CRISPR arrays are associated with very strong ssDNA signatures." At this stage of the manuscript, without an explanation that ssDNA signatures are used as a proxy for active transcription in the study, this statement misses an important link. The word 'correlation' fits the statement better than 'coordination' given the results that have been presented in the manuscript. Also, sufficient background has not been provided in the introduction to explain or even suggest why this finding is unexpected.

We have added a clarification on what the expectation was.

6. In the results: "... nearly all ATAC-seq peaks are located within 200 bp of an annotated TSS." 'Nearly all' would be better replaced with a percentage.

We disagree. The *Haloferax* annotation is not sufficiently precise in terms of its identification of TSSs, thus a precise percentage is meaningless. The global trend is what is important.

7. The figure reference at the end of the sentence that reads, "To properly interpret sequencing data..." should be Figure 1H instead of 1G.

We than the reviewer for spotting this discrepancy and have corrected it.

8. The Y-axis of Figure 2A is labelled '1-methylation'. The label appears to be colloquial and does not represent what exactly is shown on the Y-axis. The axis should be relabelled.

This is the standard way of presenting NOMe/SMF data.

9. 'TSS metaprofiles' in the figure legend of figures 2A-C is also colloquial. The text should be clearer.

This is also standard.

10. For Figures 2E and 2F, displaying the relevant section of the ensemble DNA protection profile (in green) directly on top of the single molecule DNA protection maps will make it easier to compare the two. Also, label the Y-axis of the (green) ensemble graph.

CAN'T REALLY DO FOR TECHNICAL REASONS

11. In supplementary figure 1, what do the SRR codes mean? What is the experimental difference between 'gDNA SRR18275628' and 'ATAC on gDNA SRR18275639'? If these are both ATAC-seq metaprofiles around TSS, why are the metaprofiles of SRR18275628 and SRR18275639 so different? Why are the ATAC-Seq metaprofiles for fixed and unfixed cells a combination of two datasets, while the gDNA datasets are represented separately?

These are external datasets for *Sulfolobus*, not datasets we generated *Haloferax*. We don't have the precise details how they were generated beyond what is provided in the published description. The SRR codes refer to the SRA (Short Read Archive) accession numbers.

12. A statement in the results reads, "... exponentially growing and stationary cultures (post log-phase in the growth curve)." The phase of growth at which these samples were collected would be better represented by marking the phase on a Haloferax volcanii growth curve. The growth curve can be a supplementary figure.

XXX SKIPPED XXX

13. The authors regularly refer to 'core eukaryotes'. This is not standard terminology.

Under the current understanding of the origin of eukaryotes, they branch from within archaea. By "core eukaryotes" we mean the archaea/eukaryotes distinction, as well as the exclusion of "weird" eukaryotes that don't have proper nucleosomal chromatin (e.g. dinoflagellates). We have replaced "core" with "conventional".

14. In the results, the authors write, "We also observe this feature as a protection footprint in a few small percent of single molecules at individual promoters (Figure 2D)." This feature should be marked in the figure. It is not clear what the authors are referring to.

It refers to the second methylation site to the left of the TSS.

15. The authors have neglected several interesting features in their KAS-seq maps. For instance, the ssDNA peak in pHV2 at the 5-10 minute region that is elevated in both, the exponential phase, and 3 month standing cultures. The 46-50 minutes region on the main chromosome of Haloferax volcanii is also interesting, since it shows a region of ssDNA signal in a 3-month standing culture but not during exponential phase.

We haven't commented on it because pHV2 contains only hypothetical proteins. That peak is the promoter of the largest of them, but what it is and what function it plays we don't know. That plasmid is also the one with the highest copy number.

16. The authors refer to the rRNA operons as being the most conspicuous operons in the H. volcanii genome. Taking into account the context in which these operons were mentioned, they should be marked in figures 3 and 4.

SKIPPED

17. In figure 4, the authors mention an "... extremely strong ssDNA peak associated with one of the CRISPR arrays on the pHV4 plasmid." For clarity, this peak should be marked on Figure 4A with an arrow or square. Also, the region that has been zoomed into in the lower panel of figure 4B should be marked on the top panel of figure 4A.

We've added an arrow as requested.

18. The ssDNA signal at Cas6 from decreases from exponential phase at 42, 37, and 23 oC to the 1-week standing culture, and then increases towards 3 months. Are there any (environmental) factors/signals that would contribute to such a pattern of expression? Do the ATAC-seq or NOMe-seq/dSMF datasets provide a clue?

We cannot make such claims without having generated multiple replicates of long-term standing cultures (and various time points too).

19. The ssDNA signal in the regions flanking the Cas6 ssDNA peak is increased in the Haloferax volcanii 1-week standing culture (see figure 4B). Is a similar pattern observed all over the genome? Do the ATAC-seq or NOMe-seq/dSMF datasets provide an explanation for this pattern?

A global increase of ssDNA signal cannot be truly measured, currently only relative profiles are obtained. ATAC profiles do not change much (Figure 4C).

20. In 'Coordination between chromatin accessibility and transcriptional activity within H. volcanii operons', paragraph 1 mentions the use of both KAS-seq and ATAC-seq data to address the title. This comparison is only vaguely described in the last paragraph. Correlation studies for this section are weak and incomplete.

OK

21. In 'Coordination between chromatin accessibility and transcriptional activity within H. volcanii operons', the conclusions and models have been inappropriately described. The conclusion is that the internal KAS-seq peaks may arise from polymerase pausing, and the model is that this is a result of transcription-translation coupling. The text in the manuscript presents the model as the conclusion.

We disagree that we made such a "conclusion", we described several possibilities very clearly, and did not pick one as a definitive conclusion.

22. In supplementary figure 2 panel E, the KAS-seq profiles of H. volcanii at 23 oC and 4 oC differ from the KAS-seq profiles at higher temperatures. What is the cluster of genes upstream of the putative ABC transporter? Why would the profiles seem so different? How does it compare to the ATAC-seq profiles? What do the KAS-seq and ATAC-seq profiles of the gene cluster look like in standing cultures. In supplementary figure 2 panel G, the shape of the KAS-seq peak at the beginning of the gene cluster coding for RNA polymerase subunits looks unique compared to the other profiles. Are there any elements in the underlying genomic sequence at this site to indicate why the profile looks as such?

Unfortunately we do not have ATAC-seq profiles for the 4C and 23C cultures.

23. The authors mention "... very strong and immediate coupling..." while describing transcription-translation coupling. Insufficient evidence/explanation has been provided to frame it in this manner. Why is the coupling 'very strong' instead of 'strong' or 'weak'? Why is it necessary to mention 'immediate' while discussing transcription-translation coupling in this context?

SKIPPED

24. The authors say, "In some cases (e.g. Figure 5C), these internal KAS-seq peaks are also associated with matched ATAC-seq peaks." For clarity, these cases should be marked in figure 5.

The operons are clearly marked in those figures. Adding arrows would only create more clutter.

25. What does the phrase "... large apparent differences in promoter accessibility magnitude..." (see results section) mean?

Promoters have much higher ATAC-seq signal. We have replaced this phrase.

26. What does "... the typical more symmetric changes between two conditions..." mean?

See above. This has been corrected/clarified.

27. In the results, the statement, "Strikingly, we did not find strong changes... ATAC-seq peaks as in actively growing cells." is unclear in terms of what question the authors were attempting to answer and why, what exactly was done to address the question and why it was approached in that manner, what the observations were, and what the observations mean.

We have rephrased.

28. The authors say in the results (with marked corrections), "...the explanation for this pattern is that it is due This is due to the dormant state in which of standing cultures are in, and in which we do not observe as strong ATAC-seq peaks as in actively growing cells." This simply says that stationary phase cells lack the ATAC-seq peaks of actively-growing cells because they are dormant.

Correct.

29. The results presented in supplementary figure 3 are significant enough to compare with figure 6A, hence, the results should be part of figure 6.

Figure S3 is a whole-page figure. There isn't the space for it in Figure 6A unless we just promote it to a main figure.

30. Which standing cultures are presented in supplementary figure 3 and figure 5A? Why are the KAS-seq profiles of standing cultures not presented in figures 5B-D, and supplementary figure 2?

It is the 1-week standing cultures. The 3-month one is only relevant to CRISPR arrays.
We omitted the standing profiles from the snapshots because they are mostly flat (aside from the CRISPR array) and there was already a lot of clutter in the figures.

31. Why are the results sections titled 'Coordination between chromatin accessibility and transcriptional activity within H. volcanii operons' and 'Chromatin accessibility does not correlate with transcriptional activity in H. volcanii' separated?

Different questions were addressed – correlation between transcription and accessibility on one hand, transcription coordination within an operon on the other.

32. References in the following phrases are missing: "We then compared the H. volcanii ATAC-seq TSS metaprofile with that from the previously published MNase-seq dataset (REF)...", "...baseline protection level of 85-90%.(REF)" and "...dynamic association of archaeosomes with DNA.(REF)"

We did reference the relevant studies upstream in the text.

33. The statement, "Nearly all promoters in H. volcanii... not assayed remains to be determined." does not really provide any additional information or new discussion points. It can be deleted.

SKIPPED

34. The paragraph "We also made the surprising observation that ... of functionally related genes)." does not add much in terms of content to the article. Observations should be discussed beyond 'a hypothesis that needs to be generalised to and tested'.

SKIPPED

35. The definition of an archaeon in the Methods section is incorrect. The implication that eukaryotes typically have a cell wall is also incorrect.

That wasn't the implication, the cell wall comment referred to many other prokaryotes. We have added a clarification.
Of course "a prokaryote without a nucleus" isn't the definition of an archaeon, but it is the description of *Haloferax*'s properties relevant to doing ATAC-seq.

36. The composition of non-standard buffers such as PB buffer and 1X TWB should be described in Methods.

The PB buffer is the one from the Qiagen MinElute kit. We have specifically mentioned that in the revised version.
The composition of TWB was already specified in the original version of the manuscript.

37. In the last paragraph of the results section titled 'The ssDNA and active transcription landscape in the H. volcanii genome', the references to Figure 2B, 2C-D, and 2E should be Figure 3B, 3C-D, and 3E.

We thank the reviewer for noticing this discrepancy. It has now been corrected.

Reviewer 3:

> Marinov et. al. investigate chromatin accessibility in Haloferax volcanii using ATAC-seq, NOMe-seq, and KAS-seq. NOMe-seq and ATAC-seq results are consistent with each other and suggest that promoter regions are generally more accessible, though position-specific accessibility is dynamic. KAS-seq, which preferentially identifies ssDNA, is presented as an alternative to RNA-seq and used to identify active sites of transcription. Interestingly, promoter accessibility does not appear to correlate with ssDNA complexes identified by KAS-seq. This is the first time that each method has been reported in Halobacteria, and more generally Euryarchaea, which comprises a notable advance in available methodologies for this clade, and would enable the research community at-large to apply these methods to their relevant research questions.
>
> Unfortunately, the impact of this study in its current form is limited by a lack of relevant citations about H. volcanii throughout the paper, inclusion of unoriginal results, and insufficient reporting of experimental controls, all of which hinder compelling interpretation of the results. I have summarized the evidence for each of these claims below and expanded on these topics in the "Major Comments" section.

OK

> First, the manuscript introduction focuses archaeal nucleosomal chromatin and the open question of nucleosomal chromatin in Euryarchaea. But, H. volcanii is unlikely to be good model organism to study Euryarchaeal histones or chromatin for several well-described reasons. First, haloarchaea have a unique fused-dimer histone which prevents polymerization to form hypernucleosome structures. In fact, it has not been established that histones are responsible for DNA compaction in haloarchaea, which the authors take for granted. Generally, the manuscript does not address aspects of H. volcanii biology that are already known and necessary for meaningful interpretation of their results (polyploidy, small regulatory RNAs, experimentally determined TSS and TTS sites).

We have based our writing on the available literature on Haloferax chromatin (i.e. the previous MNase study). In general it is thought that where there are histones, they are the main packaging component, even in archaea. But we are not wedded to the view in any way, and have revised the manuscript accordingly.

> Second, several of the main findings have been published previously. Correct scholarly practices require that this context is provided. The result that promoter accessibility does not correlate with transcriptional activity is novel and notable. However, it is not convincingly established that KAS-seq is a good indicator of transcriptional activity in this model system, complicating the finding. Furthermore, chromosome structure and its relation to transcriptional activity has been investigated using Hi-C [1, uncited]. This knowledge should be included in the introduction to aid understanding of H. volcanii genome organization. NoME-seq is an interesting and useful method, and its application here is novel. However, it is impossible to interpret the results without addressing the polyploidy of H. volcanii. Other main findings in the current study, such as internal TSS in operons and the importance of s497 (the small regulatory RNA upstream of cas6 on pHV4) in stationary growth have also been reported previously [2-4, all uncited]. I strongly recommend updating the KAS-seq analysis to include the experimentally reported TSS as well small and anti-sense RNAs to help distinguish between novel and previously reported results.

Whether KAS-seq measures active transcription is really not dependent on the model system, this is just a physical property of RNA polymerases (and the archaeal one is in fact homologous to the eukaryotic one).

No results in our study concern 3D genome organization (though in fact we did do Hi-C too, and have not published it because we cannot reproduce the results from the published Hi-C analysis neither with our own, nor with the data from that study itself).

Polyploidy is completely irrelevant to NOMe-seq data.

We thank the reviewer for the suggested citations about operons and have added them to the text.

> Finally, some of the data lack appropriate controls, benchmarks, and exploration expected when executing sequencing methods in a novel system. The authors should provide controls demonstrating

that Tn5 transposition is either un-biased with respect to genomic context, or that their normalization appropriately accounts for insertion bias. KAS-seq should be compared with previously published RNA-seq transcriptomes for this experimental system [2, 3, 4] in order for it to be proposed as a valid alternative, and to show whether peak enrichment correlates with magnitude of expression. Lacking transcriptome-level correlations, known highly expressed genes should be used as internal controls and validated in-house with RT-qPCR. The methods should be updated to include information about sequencing depth, coverage, and power for each experiment and include a more detailed account of what, if any, significance or LFC thresholds were applied. If results from replicates are merged, details about the process should be included in the methods as well.

Tn5 insertion bias is relevant to local footprinting, not to large-scale accessibility profiles.

We used KAS-seq because we wanted to map **active** transcription, not steady-state transcription as all the previous studies relying on RNA-seq did. KAS-seq is very well established as a good proxy from active transcription from work in mammals.

Standard differential analysis was carried out using DESeq2.

In short, the methodologies reported here are of major interest to haloarchaeal researchers. This report would have greater impact to both the archaea research community and the scientific community generally if substantial additional biological context, interpretations relevant to the model system, and data validation were provided. Aside from reading and citing papers relevant to H. volcanii, one possible way to address these deficiencies and improve the manuscript would be to collaborate or consult with haloarchaea experts.

We have now updated the manuscript based on multiple exchanges with such researchers.

Major Comments:

• "Nucleosomal chromatin" is not defined in the text, but it appears to refer specifically to histone-mediated, higher order DNA structure. If this is true, the authors take for granted the presence of nucleosomal chromatin in Haloferax volcanii. Multiple publications have begun to cast doubt on the assumption that histones are the major chromatin-associated protein in H. volcanii [5-7]. The authors do not acknowledge recent work demonstrating that histones in haloarchaea are not essential (a clear deviation from eukaryotes) and instead regulate expression of specific genes like a transcription factor [8-10]. Furthermore, Sandman and Reeve suggest that the fused dimer structure of haloarchaeal histones (unique among archaea) would prevent polymerization resulting in "slinkie" or hypernucleosome structures [11]. In order to conclude that the ATAC-seq profiles shown here are caused by histone-based nucleosomes, the authors need to show that the ATAC-seq accessibility changes significantly when the sole histone encoded in H. volcanii is knocked out.

We have updated the manuscript with the suggested citations and revised the Results and Discussion sections accordingly.

We have no viable way to knock out the histone at the moment.

• There is no discussion of ATAC-seq coverage or Tn5 insertion bias. Though H. volcanii has high GC content (68%), there are regions with higher AT, such as pHV4 [1]. While the authors account for copy number variation of replisomes relative to the main chromosome, they do not state whether the relative abundance of replisomes is constant across growth phases, which is unlikely given that ploidy is growth-phase dependent in H. volcanii. Details regarding copy number calculations are absent from the methods.

No analysis presented in the paper is contingent on relative ploidy being constant across phases.

We have added a basepair composition analysis.

• H. volcanii is polyploid, with 20 copies of the genome during exponential growth. Having multiple copies with differing accessibility could contribute to the dynamic protection patterns observed. NOMe-Seq single molecule maps (Figure 2) should be clustered, which would aid interpretation and potentially reveal distinct populations. The authors should address how varying polyploidy could confound interpretation of accessibility results. It is possible to reduce the ploidy of H. volcanii under phosphate limitation [12]. This could be a potential ploidy control.

Single-molecule maps are already clustered.
And they feature single molecules, so ploidy does not really play a role in that analysis.

- KAS-seq data should be compared to previously published RNA-seq datasets [2-4].

See above.

- There are several reports describing regulatory RNAs and their likely TSS [2-4, 13]. As the observed KAS peak in P2 CRISPR array is likely do to sRNA binding, are there other KAS peaks correlating with the location of sRNAs?

The observed peak is actually not associated with the s479 RNA, and the s479 RNA is not reported to act in *cis*. Most sRNAs in *Haloferax* are also trans-acting on RNA, not on chromatin.
Distinguishing the peak of the promoter of a sRNA from that of the nearby gene is not always straightforward (it is possible in the case of the CRISPR array because of its particular configuration).
We have revised the manuscript accordingly to include a discussion of s479.

Minor comments:

- The detection of a protection footprint upstream of TSS in NOMe-Seq data is particularly interesting – how does the location of this protection footprint relate to BRE and TATA sites? Is there a particular motif comprising this protected region? Does the degree of protection (frequency protected vs unprotected) correlate with level of expression?

SKIPPED

- Violin plots with data points overlaid should be used in Figure 1A rather than a boxplot.

Violin plots have been made.

- Figure S1: separate lines for the A and B compartment genes should be included from the S. islandicus data. It is possible that the A compartment, which contains most essential and highly expressed genes, has similar TSS enrichment to Hfx. volcanii.

Strong localized eukaryote-like peaks such as those in *Haloferax* are not visible anywhere along the *Sulfolobus* genome so the lack of peaks there isn't about the A/B compartment distinction.

- Methods should include more information about culture handling – what OD or cell density were considered exponential? What was the growth rate of cultures when harvested for ATAC- and KAS-seq? How old were the plates when cultures were inoculated? How and why was $1X10^6$ cells determined to be optimal for ATAC-seq?

SKIPPED

- Numerous IS elements have been identified across the genome, and were implicated in chromosome organization. What is chromatin accessibility like at some of the boundaries reported in [1]?

The boundaries reported in [1] are not reproducible both in our own Hi-c data and in the Hi-C data from [1] itself when reprocessed through the standard Juicer pipeline. This is why we have avoided all discussion related to Hi-C and "boundaries" in our own manuscript as this would require us to start a fight we want to avoid for the time being.

- Several rRNA removal methods are available for archaea and have techniques have been recently been compared in H. volcanii [14]. The text should be updated to reflect this.

There is no [14] in the provided citations list. Presumably PMID 35625610 is referred to. That appeared 2.5 years after we did our KAS-seq experiments. We have updated the text to acknowledge more recent developments.