

RESEARCH

Open Access



# The chromatin landscape of the euryarchaeon *Haloferax volcanii*

Georgi K. Marinov<sup>1\*</sup> , S. Tansu Bagdatli<sup>1</sup>, Tong Wu<sup>2</sup>, Chuan He<sup>2,3,4</sup>, Anshul Kundaje<sup>1,5</sup> and William J. Greenleaf<sup>1,6,7,8</sup>

\*Correspondence:  
GKM359@gmail.com

<sup>1</sup> Department of Genetics, Stanford University, Stanford, CA 94305, USA

<sup>2</sup> Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA

<sup>3</sup> Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA

<sup>4</sup> Howard Hughes Medical Institute, The University of Chicago, Chicago, IL 60637, USA

<sup>5</sup> Department of Computer Science, Stanford University, Stanford, CA 94305, USA

<sup>6</sup> Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA

<sup>7</sup> Department of Applied Physics, Stanford University, Stanford, CA 94305, USA

<sup>8</sup> Chan Zuckerberg Biohub, San Francisco, CA, USA

## Abstract

**Background:** Archaea, together with Bacteria, represent the two main divisions of life on Earth, with many of the defining characteristics of the more complex eukaryotes tracing their origin to evolutionary innovations first made in their archaeal ancestors. One of the most notable such features is nucleosomal chromatin, although archaeal histones and chromatin differ significantly from those of eukaryotes, not all archaea possess histones and it is not clear if histones are a main packaging component for all that do. Despite increased interest in archaeal chromatin in recent years, its properties have been little studied using genomic tools.

**Results:** Here, we adapt the ATAC-seq assay to archaea and use it to map the accessible landscape of the genome of the euryarchaeote *Haloferax volcanii*. We integrate the resulting datasets with genome-wide maps of active transcription and single-stranded DNA (ssDNA) and find that while *H. volcanii* promoters exist in a preferentially accessible state, unlike most eukaryotes, modulation of transcriptional activity is not associated with changes in promoter accessibility. Applying orthogonal single-molecule footprinting methods, we quantify the absolute levels of physical protection of *H. volcanii* and find that *Haloferax* chromatin is similarly or only slightly more accessible, in aggregate, than that of eukaryotes. We also evaluate the degree of coordination of transcription within archaeal operons and make the unexpected observation that some CRISPR arrays are associated with highly prevalent ssDNA structures.

**Conclusions:** Our results provide the first comprehensive maps of chromatin accessibility and active transcription in *Haloferax* across conditions and thus a foundation for future functional studies of archaeal chromatin.

## Introduction

Life on earth is now understood to be divided into two deep fundamental clades — Archaea and Bacteria. Archaea were only discovered as a separate branch of the tree of life in the 1970s [1], yet it was noticed very early on that they share a number of common features with the more organizationally complex eukaryotes, especially in the organization of their information processing cellular machinery. Based on these



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

similarities it was suggested that eukaryotes evolved from archaea [2, 3], a view strengthened in the phylogenomic era [4], and eventually solidified with the discovery of archaeal lineages such as the Lokiarchaeota [5]. Thus, we now know that many of the complex cellular features that characterize eukaryotes trace their origins to their archaeal ancestry [6, 7].

One of the most notable such features is nucleosomal chromatin. Nearly all eukaryotic genomes are packaged by nucleosomes, consisting of two tetramers of the four core histones H2A, H2B, H3, and H4, wrapping around ~147 bp of DNA. These proteins are, with very rare exceptions [8, 9], the most evolutionarily conserved among eukaryotes [10], in large part because aside from their packaging function they are also subject to a large number of precisely regulated posttranslational modifications (PTMs) at key residues [11], through which they play a pivotal role in all aspects of chromatin biology (transcription and its regulation, DNA replication, DNA repairs, mitosis, and others).

As early as the 1980s, it was noticed that some archaea possess proteins and structures similar to eukaryotic histones and nucleosomes [12, 13]. We now know that most archaea have histones genes [14–16] and that these histones are ancestral to the eukaryotic histones. Archaeal histones differ substantially from those in eukaryotes — while they share the core histone fold domain, they usually do not have the unstructured tails of H2A/H2B/H3/H4 that are the main sites of key PTMs. Archaeal histones also do not form octameric nucleosomes; instead, only one or a very small number of histone genes are found in archaeal genomes, and the structures they form are very different from those of eukaryotes. The diversity of histone sequences across the whole archaeal phylogeny is very large and still largely unexplored experimentally, but the available structural [17], biochemical, and modeling work suggests that in at least some species histones can form so-called hypernucleosomes or archaeosomes, consisting of a protein core of individual histones stacked next to each other, around which DNA is wrapped [15, 18], ranging from 60 to 500 bp [19]. It has also been proposed that archaeal histones exhibit an inherently dynamic association with DNA [19], in contrast to the much more stable association of nucleosomes with DNA in eukaryotes.

However, chromatin is not nucleosomal in all archaeal lineages — some lack histone genes and use other proteins such as Alba/Sac10b, Sul7d, Cren7, and CC1 [20, 21] to package their genomes, and even for the ones that do have histone genes it is not always clear that histone proteins are in fact the main packaging component of chromatin (see the “Discussion” section further below). The latter situation is particularly notable in haloarchaea [22–27] — the tentative conclusion that their histones do not play a dominant role in DNA packaging is based on the low abundance of the HstA protein in *Haloferox* proteomics datasets [26, 27] and on data suggesting that its functions are more analogous to those of a transcription factor [24].

Despite the relevance of archaeal chromatin to understanding the deep evolution of chromatin organization, up until now, the structure of archaeal chromatin has received little direct experimental investigation using modern genomic tools, with the exception of early MNase-seq studies nearly a decade ago that mapped nucleosomal positioning in the euryarchaeotes *Haloferox volcanii* [28] and *Methanothermobacter thermautotrophicus* and *Thermococcus kodakarensis* [29], and more recent MNase-seq studies in *Methanothermobacter fervidus* [30], *Thermoplasma acidophilum* [31], and *Methanocaldococcus*

*jannaschii* [32]. Furthermore, very little is known about the relationship between chromatin structure and the regulation of gene expression in these organisms.

In order to begin to fill these gaps in our understanding of the organization of archaeal chromatin, we mapped chromatin accessibility and active transcription in *Haloferax volcanii* using a combination of bulk and single-molecule techniques such as ATAC-seq [33] (Assay for Transposase-Accessible Chromatin using sequencing), NOMe-seq/dSMF [34] (Nucleosome Occupancy and Methylome sequencing/dual Single-Molecule footprinting) and KAS-seq [35] (kethoxal-assisted single-stranded DNA sequencing). We find that chromatin in *H. volcanii* exhibits similar features to that of eukaryotes on a broad level, with preferentially accessible promoter regions. However, unlike in eukaryotes, chromatin accessibility at promoters does not relate to transcriptional activity. Using single-molecule footprinting we estimate absolute protein occupancy levels over the *H. volcanii* TSSs to be comparable to, or possibly slightly lower than those in eukaryotes. However, unlike what is seen in most eukaryotes, we do not observe stably positioned nucleosome protection footprints, but rather only statistically elevated accessibility around promoters. We also revisit the question about the degree of 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.

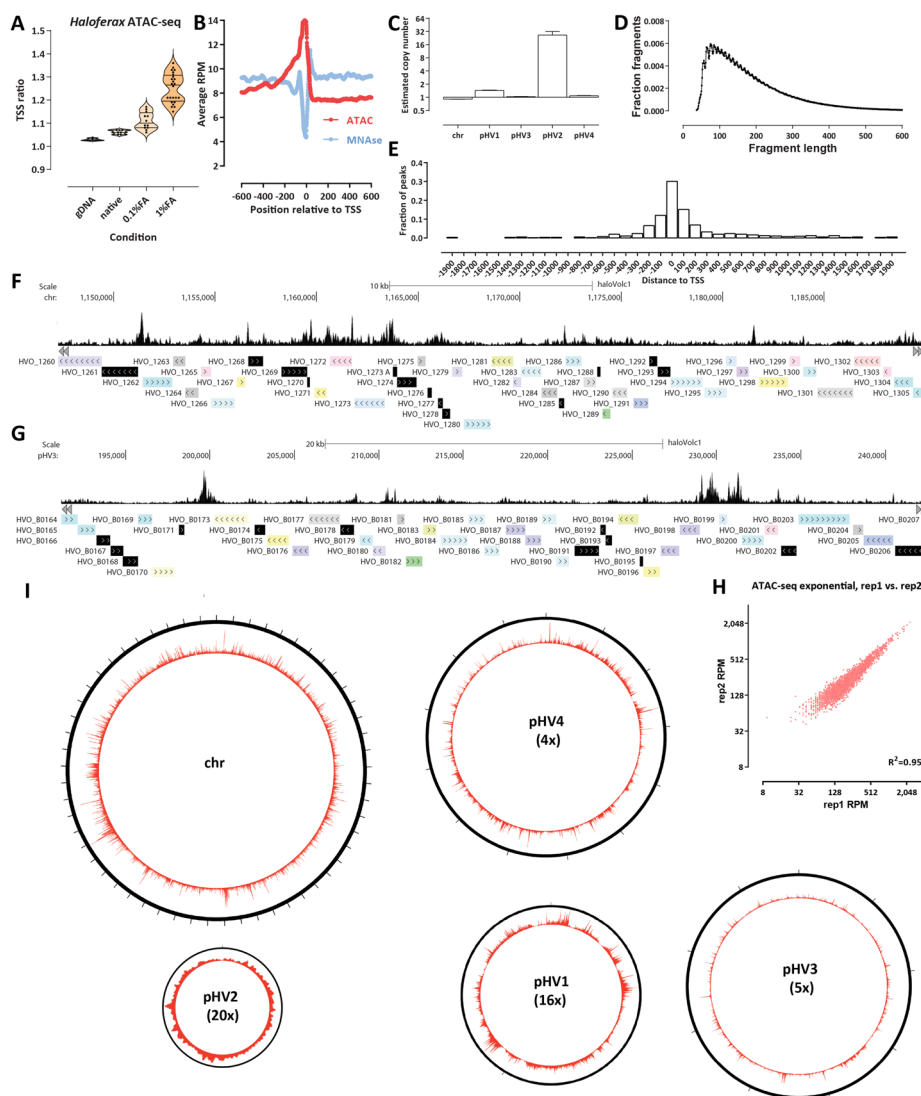
## Results

### ATAC-seq reveals the open chromatin landscape of *H. volcanii*

In order to study chromatin accessibility in archaea, we adapted the ATAC-seq assay [33] to the *Haloferax volcanii* archaeon. *H. volcanii* is a halophile with a strong preference for very high salt concentrations in the growth medium (see the “Methods” section), which grows optimally at 42°C [36, 37], and is a widely used archaeal model system.

The principle behind ATAC-seq is the very strong preference of the Tn5 transposase [33] for insertion into accessible DNA as opposed to tagmentation of protected (by nucleosomes, transcription factors, or other proteins) DNA. Tn5 insertion then tags accessible sites with landing sites for PCR primers, allowing for highly efficient amplification of open chromatin regions in the genome.

After extensive testing of a variety of different experimental protocols (fixation conditions and input cell numbers), we arrived at the following modifications of the standard ATAC protocol. First, because archaea are not eukaryotes and do not have a nucleus, we omitted the cell lysis and nuclei isolation step that is a standard feature of eukaryotic ATAC-seq protocols, such as the now standard omniATAC [38]. Second, and most importantly, we reasoned that if the previously reported dynamic association with DNA of archaeal nucleosomes (or of other proteins) occurs in *H. volcanii*, optimal results might be obtained by introducing a crosslinking step into the standard ATAC protocol, which would “freeze” any histones and other structural proteins in place and not allow transposition into DNA that might change from protected to accessible during the duration of the transposition reaction. Indeed, comparing the TSS (transcription start site) enrichment generated without fixation and with light (0.1% formaldehyde) and strong (1% formaldehyde) fixation showed that stronger fixation produces higher TSS enrichment (Fig. 1A). We then compared the *H. volcanii* ATAC-seq TSS metaprofile with that



**Fig. 1** Archaeal ATAC-seq and the open chromatin landscape of *H. volcanii*. **A, B** Adaptation and optimization of the ATAC-seq assay to the archaeal context. **A** Distribution of TSS ratio scores (see the “Methods” section for details) for native, 0.1%- and 1%-formaldehyde ATAC-seq libraries. **C** Fragment length distribution in *H. volcanii* ATAC-seq datasets. **D** Estimated relative copy number of *H. volcanii* chromosomes. Genomic DNA was tagmented and amplified ( $n=4$ ) and normalized read coverage was estimated for each chromosome/plasmid. The average ratios are shown. **E** Distribution of MACS2 ATAC-seq peaks relative to TSSs. **F, G** Representative browser snapshots of ATAC-seq profiles along the *H. volcanii* genome. **F** High reproducibility of *H. volcanii* chromatin accessibility measurements using ATAC-seq. Shown is the between-replicate correlation over TSSs in RPM (reads per million) units. **I** Global ATAC-seq profile over each of the five *H. volcanii* chromosomes. The number in brackets corresponds to the magnification of the true proportional size of plasmids relative to the main chromosome

from the previously published MNase-seq dataset and observed the expected inverse relationship (Fig. 1B). *H. volcanii* TSSs exhibit elevated accessibility in the 0 to -400 bp upstream region, decreasing away from the TSS.

The *H. volcanii* genome consists of multiple replicons [39], with a main chromosome (“chr”) and four plasmids of very different sizes — pHV4, pHV3, pHV1, and pHV2 (in order of decreasing size), which together comprise ~30% of the total genome. To

properly interpret sequencing data (which is typically normalized to total read coverage), we determined the relative copy number distribution of these replicons using tagged naked genomic DNA control samples (Fig. 1C). The smallest plasmid — pHV2 — appears to exist in  $\sim 26$  copies for each main chromosome, pHV1 is found at  $\sim 1.4$  copies for each main chromosome, while the two large plasmids — pHV4 and pHV3 — exist in a 1:1 ratio to the main chromosome.

The *H. volcanii* ATAC-seq fragment length distribution is unimodal, peaking at 90–100 bp, and does not show an analog to the eukaryotic mono-, di- and tri-nucleosomal signature (Fig. 1D), even scaled down to the smaller protection footprint of archaeal histones. This suggests that *Haloferox* chromatin may not be packaged into abundant nucleosomal structures consisting of strings of closely positioned individual nucleosomes.

Peak calling using MACS2 [40] (Fig. 1E) and genome browser visualization of ATAC-seq profiles (Fig. 1F–G) revealed an accessibility landscape largely reminiscent of that in eukaryotes with compact genomes such as the budding yeast *Saccharomyces cerevisiae* [41] — nearly all ATAC-seq peaks are located within 200 bp of an annotated TSS, and these peaks are very strong and localized.

ATAC-seq measurements in *H. volcanii* are also highly reproducible between experimental replicates (Fig. 1F).

As previous studies of chromatin openness in bacteria have reported the existence of very large domains of lower and higher accessibility [42], we wondered whether the same is observed in *Haloferox*. We do not observe such domains in our datasets (Fig. 1I). Recently, an ATAC-seq dataset was reported from the crenarchaeote *Sulfolobus islandicus* [43], which lacks histones, and instead packages its genome mainly through Alba/Sac10b proteins [44, 45]. In that species, large domains similar to those in bacteria were reported. This observation may suggest that such large-scale domains of elevated chromatin accessibility are a feature associated with the lack of nucleosomal chromatin in prokaryotes, while archaea that contain histone genes, such as *H. volcanii*, exhibit eukaryote-like organization, but such a generalization is contingent on *Haloferox* chromatin being in fact nucleosomal. We also reexamined the *Sulfolobus islandicus* dataset and found that it displays a much more modest TSS enrichment than that seen in *H. volcanii*, which is also more narrowly concentrated around the TSS position (Additional file 1: Fig. S1).

Finally, we observed an anti-correlation between ATAC-seq signal and genomic GC content (Additional file 1: Fig. S2). *Haloferox volcanii* exhibits a rather high average GC content of 65% [39], decreasing to 58% in intergenic regions, and areas with even higher GC content ( $> 65\%$ ) show markedly lower ATAC-seq signal. This observation is corroborated by the available external MNase-seq dataset, which shows a positive correlation with GC content (i.e., the inverse of ATAC-seq, as expected) and naked DNA controls (which show no correlation with GC content, indicating that PCR biases during sequencing library preparation are not the reason for the observed patterns).

#### **Absolute DNA occupancy/protection levels in *H. volcanii***

While ATAC-seq is immensely helpful for identifying the location of accessible regions in the genomes and measuring their relative accessibility, it is a bulk method that does not provide information about the absolute levels of protection/accessibility

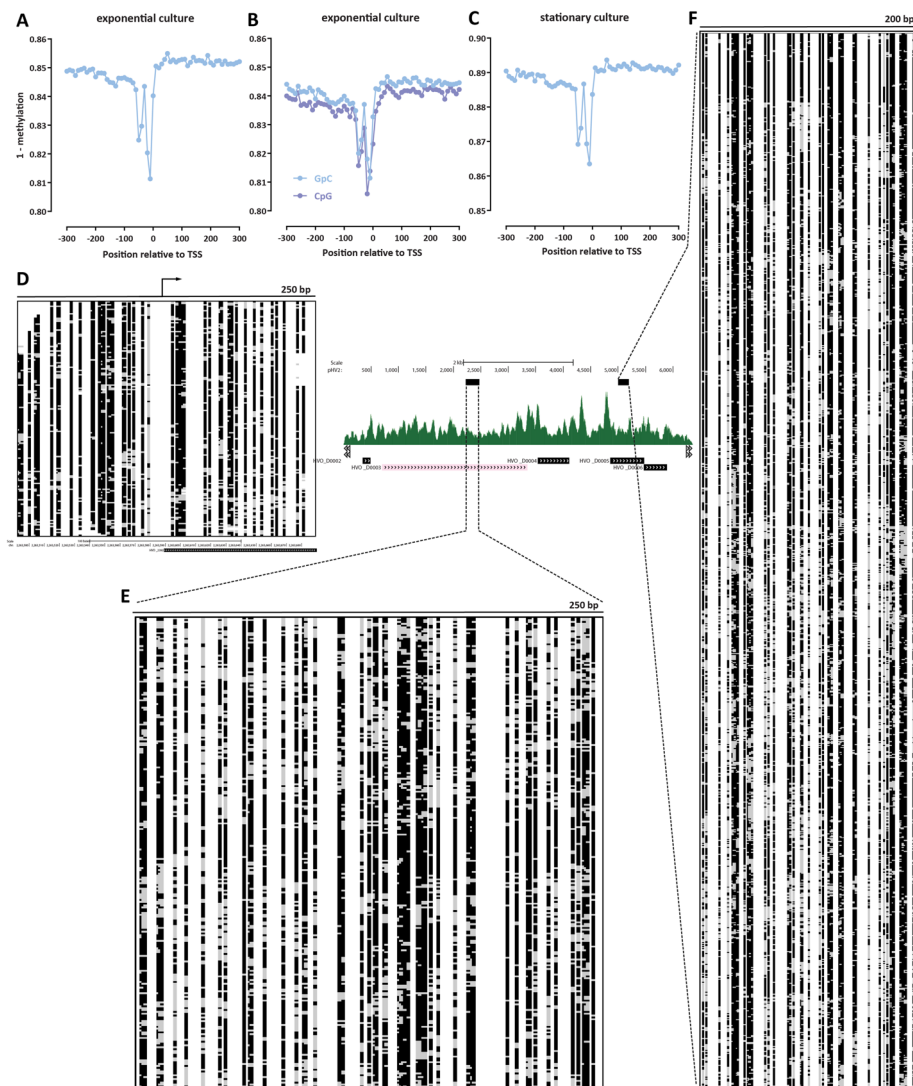
in the genome. Instead, absolute accessibility must be measured by either restriction digestion-based or enzymatic labeling single-molecule methods. To quantify absolute occupancy/protection levels in the *H. volcanii* genome, we applied NOME-seq [46] and dSMF [34] to *Haloferax* chromatin. These methods rely on the preferential methylation of accessible cytosine nucleotides (5mC) by a recombinant methyltransferase that modifies specifically in GpC contexts (NOME-seq) or a combination of methyltransferases that label both GpC and CpG (dSMF).

However, these methods are potentially confounded by the presence of endogenous methylation in either context. Fortunately, in the case of *H. volcanii* endogenous DNA modifications have been previously studied using PacBio single molecule sequencing, and no CpG and GpC modifications were found. Instead, only two restriction modification system-associated modifications in different contexts, specifically, 4-methylcytosine in a C(m4)TAG context and N6-methyladenine in a GCA(m6)BN6VTGC context were identified [47].

Figure 2A–C show the metaprofiles of average methylation around *H. volcanii* TSSs for NOME-seq and dSMF datasets generated from exponentially growing and stationary cultures (post log-phase in the growth curve). We observe baseline absolute protection levels around 84–85% in the exponentially growing cells and ~89% in stationary cells. For comparison, analogous studies in eukaryotes, such as the budding yeast *S. cerevisiae* [41, 48, 49], have shown absolute protection levels around 90% ( $\pm 5\%$ ). Thus, *Haloferax* chromatin exhibits broadly similar, though perhaps somewhat lower levels of protection than what is observed in conventional eukaryotes.

The base-pair resolved nature of these single-molecule methods enabled an observation of a feature not readily apparent in ATAC-seq and MNase-seq datasets — a protection footprint immediately upstream of the TSS. We also observe this feature as a protection footprint in a few percent of single molecules at individual promoters (Fig. 2D). At present we are not able to confidently identify its functional association — its width is likely too small for it to be a positioned -1 nucleosome, and we hypothesize it may correspond to one of the complexes involved in the archaeal transcriptional cycle, analogous to how similar protection footprints associated with the RNA polymerase and the preinitiation complex (PIC) in eukaryotes have been observed in dSMF datasets [34].

On the other hand, unlike this unique protection footprint, we do not observe strongly positioned individual nucleosome-like features along the *Haloferax* genome, such as those seen in conventional eukaryotes (Fig. 2D). However, our NOMEseq and dSMF datasets were sequenced at an effective depth of  $\sim 100\times$  for fragments of width 200 bp. To determine if higher sequencing depth would clearly reveal such putative positioned nucleosomes (or other proteins), we turned to the pHV2 plasmid, which, as previously discussed, exists in high copy numbers in *H. volcanii* cells. Over the pHV2 plasmid, we obtained  $\sim 200\times$  coverage for fragments of length 250 bp (Fig. 2E) and  $\sim 1200\times$  coverage for fragments of length 200 bp (Fig. 2F). These high-depth maps also reveal considerable heterogeneity of footprints and accessible sites. These observations are not consistent with the existence of abundant strongly positioned nucleosomes (or other structural proteins) associated with the *Haloferax* genome; they can be explained by dynamic and/or positionally unstable association of chromatin proteins with DNA.



**Fig. 2** Absolute DNA occupancy/protection levels in *H. volcanii*. **A–C** TSS methylation profiles in different conditions (two replicates of an exponentially dividing culture, and a stationary culture). **D** Single-molecule map (250-bp) around a main-chromosome TSS. Black indicates unmethylated and therefore protected sites, and gray indicates methylated and thus accessible sites. **E, F** Single-molecule maps over the pHV2 plasmid: 250-bp window map (**E**) and very high coverage ( $\geq 1200$  single molecules) 200-bp window map (**F**)

Finally, we also note that the absence of distinct states in the *Haloferax* NOME-seq/dSMF data points to all copies of its various replicons existing in mostly the same state. The two small *Haloferax* plasmids are certainly polyploid, and highly so in the case of pHV2, but quite likely even the main chromosome exists in multiple copies in individual cells. We see no evidence that some of these copies may exist in a different chromatin state (e.g., silent versus active) than the others.

### The ssDNA and active transcription landscape in the *H. volcanii* genome

We then turned our attention to the landscape of active transcription in *H. volcanii*. To this end, we used KAS-seq [35], which measures with high specificity the presence of single-stranded DNA in the genome, by labeling unpaired guanine

nucleotides with N<sub>3</sub>-kethoxal, using click-chemistry to add a biotin moiety, then enriching these fragments via a streptavidin pulldown. Most ssDNA is usually found within the transcriptional bubbles associated with RNA polymerase molecules engaged with DNA [35]. KAS-seq provides several advantages in the *H. volcanii* context. First, due to the absence of straightforward methods to deplete *H. volcanii* ribosomal RNA (rRNA) from RNA sequencing libraries analogous to polyA-selection in eukaryotes [50, 51], it enables measurement of transcriptional activity at a much lower cost than deep RNAseq experiments. Second, and most importantly, it measures actively engaged polymerase molecules, and thus unlike the steady-state transcript levels that conventional RNA-seq quantifies, it provides a means to quantify active transcriptional activity. Third, it also identifies other ssDNA structures, such as those resulting from paused polymerase molecules, G-quadruplexes, and others.

We carried out a time course analysis of *Haloferax* growth and applied both KAS-seq and ATAC-seq during the “exponential” log-phase of growth, and on the “stationary” post-log phase stage, as well as on “standing” cultures, which had been left at room temperature for ~1 week. We also carried out KAS-seq on exponentially growing cells that were then incubated at different temperatures — the typical growing temperature of 42 °C, 37 °C, 23 °C and a cold shock at 4 °C for 4 h.

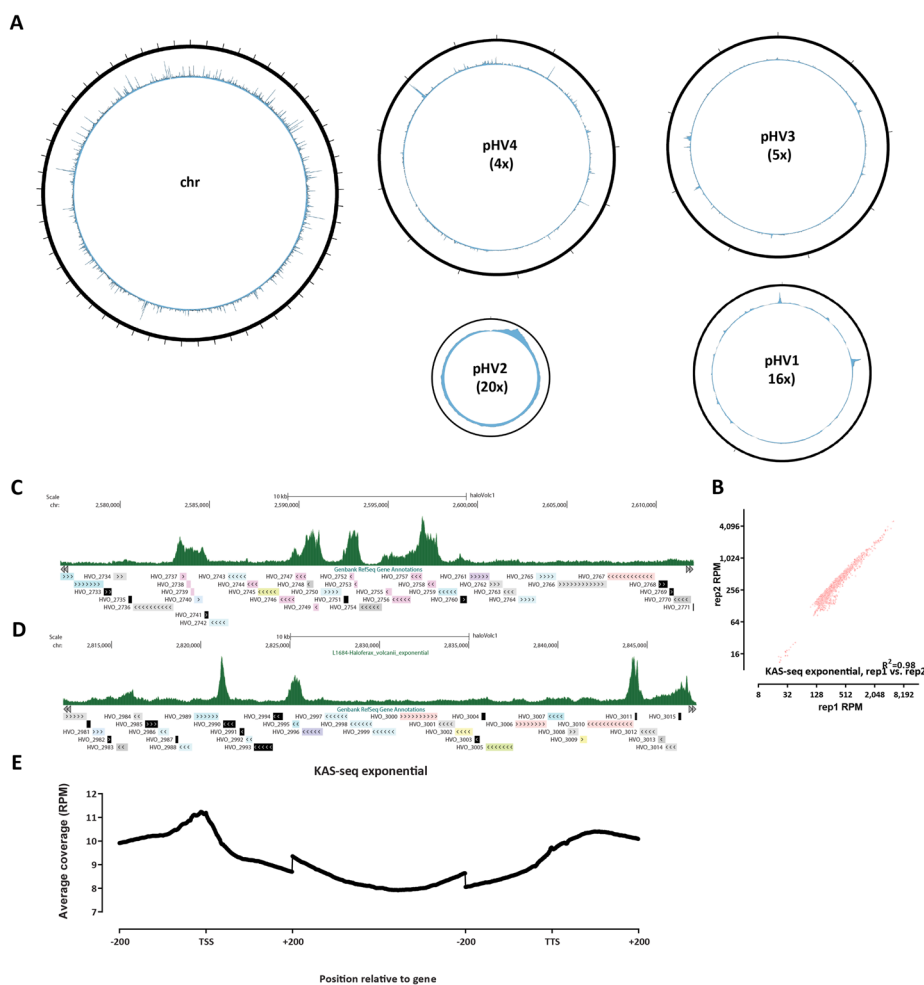
At a global level, we observe uniform levels of KAS signal along the length of *H. volcanii* chromosomes (Fig. 3A), with sharp localized peaks. KAS-seq measurements in *H. volcanii* are highly reproducible between experimental replicates (Fig. 3B). Locally, at the level of individual genes, we observe a combination of high-signal peaks at the promoters of some genes and elevated KAS-seq signal along gene bodies (Fig. 3C–D). This feature is also in KAS-seq metaprofiles over all genes (Fig. 3E), suggesting that in *H. volcanii* RNA polymerases spend substantial amount of time associated with the TSS, perhaps in a paused state analogous to that observed in metazoans [52] or the archaeon *Sulfolobus solfataricus* [53].

#### **Strong, culture condition-dependent ssDNA signals are associated with some *H. volcanii* CRISPR arrays**

In the process of optimization of the KAS-seq assay in *Haloferax*, we carried out KAS-seq on a *H. volcanii* culture that had been left standing at room temperature for ~3 months. These data revealed that CRISPR arrays can become highly enriched for KAS-seq signals in these viable, yet dormant, cultures. CRISPR (clustered regularly interspaced short palindromic repeats) arrays are a key element in the defense systems against foreign genetic material of many prokaryotes and consist of multiple identical repeats interspersed with non-repetitive sequences that target foreign plasmids and phages, together with a set of Cas genes. *H. volcanii* is one of the prokaryotic systems where these elements were first originally observed [54–56].

In standing *Haloferax* cultures, transcriptional activity is likely suppressed, as the cells enter a dormant state. Consistent with a transcriptionally inactive state, we observe largely flat, low levels of KAS-seq signal from this standing culture KAS-seq dataset (Fig. 4A). However, we also observed a single large, sharp peak of KAS-seq signal on the pHV4 plasmid. This peak resides between the second (as numbered

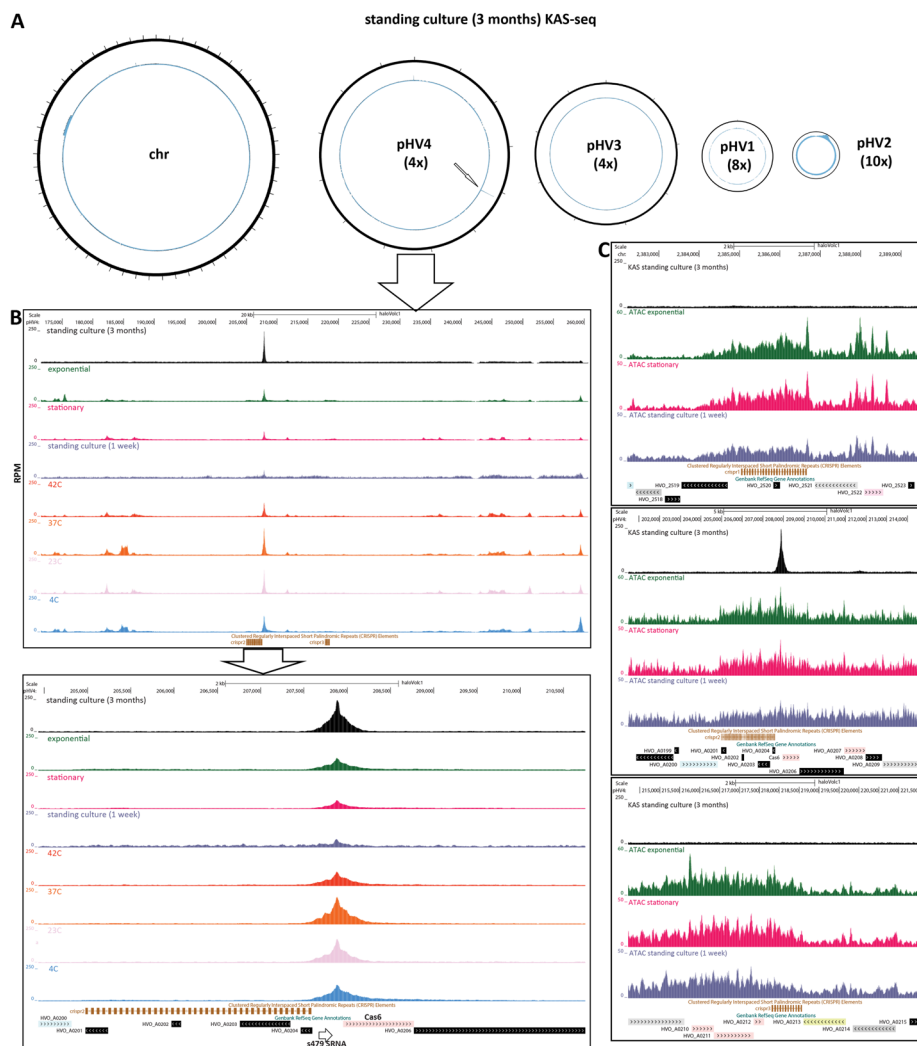




**Fig. 3** The ssDNA and active transcription landscape in the *H. volcanii* genome as measured by KAS-seq. **A** Global KAS-seq profiles over each of the five *H. volcanii* chromosomes in an exponential culture. **B** High reproducibility of active transcription measurements using KAS-seq. **C–D** Representative browser snapshots of KAS-seq profiles along the *H. volcanii* genome. **E** KAS-seq metaprofile along *H. volcanii* gene bodies

in the available genome annotation) CRISPR array in the *H. volcanii* genome [57] and its associated Cas6 gene (Fig. 4B); a previously reported small RNA (sRNA) — s479 [58] — is also located in between Cas6 and the CRISPR array, but our observed KAS-seq peak is not associated with this putative promoter, but is instead situated downstream of the array. While this KAS-seq signal peak is also found in all other conditions we assessed, it stands out in the long-term standing culture due to the absence of other peaks that result from the active transcription of other genes.

Curiously, only the second CRISPR array in *H. volcanii* displays this strong ssDNA structure, while the other two do not (this is true across all assayed conditions). However, all three arrays show elevated chromatin accessibility in ATAC-seq datasets — an accessibility signal that is not focused on the beginning of the array but covers its whole length (Fig. 4C). Possible interpretations of these observations are discussed below.



**Fig. 4** Abundant ssDNA structures associated with some *H. volcanii* CRISPR arrays in specific conditions. **A** Global KAS-seq profiles over each of the five *H. volcanii* chromosomes in a long-standing culture (~3 months) reveal an extremely strong ssDNA peak associated with one of the CRISPR arrays on the pHV4 plasmid. **B** KAS-seq signal levels around pHV4 plasmid CRISPR arrays in different conditions. **C** KAS-seq and ATAC-seq levels around all three *H. volcanii* CRISPR arrays in different conditions

### Coordination between chromatin accessibility and transcriptional activity within *H. volcanii* operons

Being prokaryotes, archaea often have genes organized into operons [59], with multiple genes transcribed as a single unit, which are therefore expected to share a common promoter and to exhibit similar levels of active transcription. However, transcription of these operons is still little studied using modern genomic tools. To address this gap, we used our KAS-seq and ATAC-seq data, which provide information about the chromatin accessibility and active transcription in different conditions, to investigate the extent of coordination between the transcriptional activity of different units in operons in the *H. volcanii* genome.

We first inspected the two *H. volcanii* operons that include rRNA genes [60]. Figure 5A shows KAS-seq and ATAC-seq profiles along rRNA genes in the different conditions we



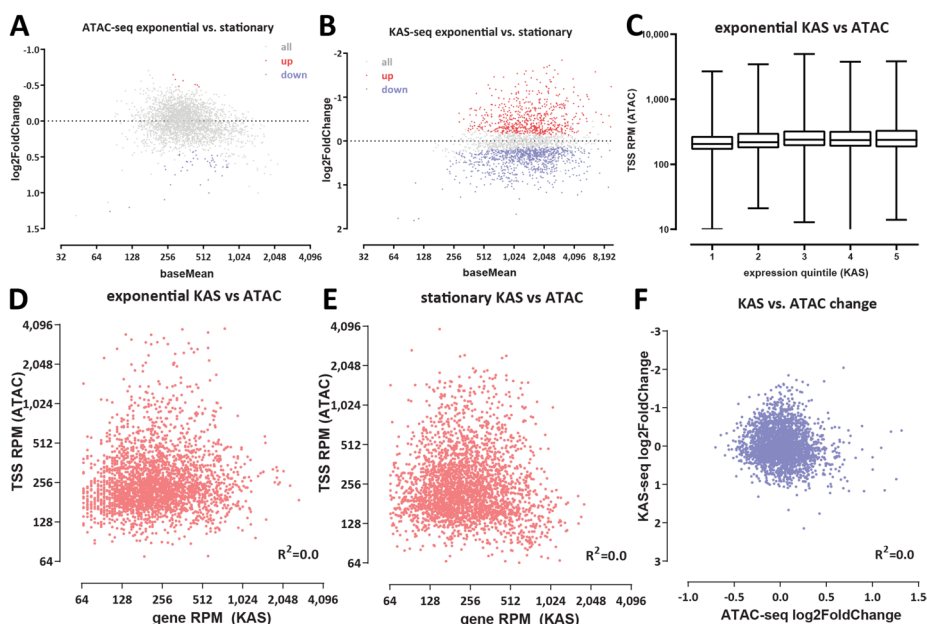
**Chromatin accessibility does not correlate with transcriptional activity in *H. volcanii***

In eukaryotes, the regulation of chromatin accessibility at regulatory elements (promoters and enhancers) is key to gene regulation, as nucleosomal chromatin is generally refractive to occupancy by regulatory proteins and to active transcription [64], and while perfect correlation between accessibility levels at promoters and gene expression is rare, open chromatin states are generally associated with increased transcriptional activity.

In contrast, the relationship between chromatin accessibility and transcriptional activity in archaea has not been systematically studied as chromatin accessibility has not been mapped globally, across conditions, and in conjunction with global measurements of active transcription.

We first identified differentially accessible promoter regions between the different conditions we studied (Fig. 6A). In contrast to an a priori expectation that changes in gene expression would be associated with shifts in chromatin accessibility levels around promoter regions, we did not find strong changes between exponentially growing and stationary cells (Fig. 6A). We observed large apparent differences in ATAC-seq signal in each of those two conditions and standing cultures (Additional file 1: Fig. S4), but in those comparisons, the profiles are highly skewed towards increased higher accessibility in the actively growing and stationary cells instead of showing the typical more symmetric changes (i.e., accessibility increasing and decreasing in both directions) between two conditions.

We believe this pattern is due to the dormant state of standing cultures in which we do not observe as strong ATAC-seq peaks as are observed in actively growing cells.



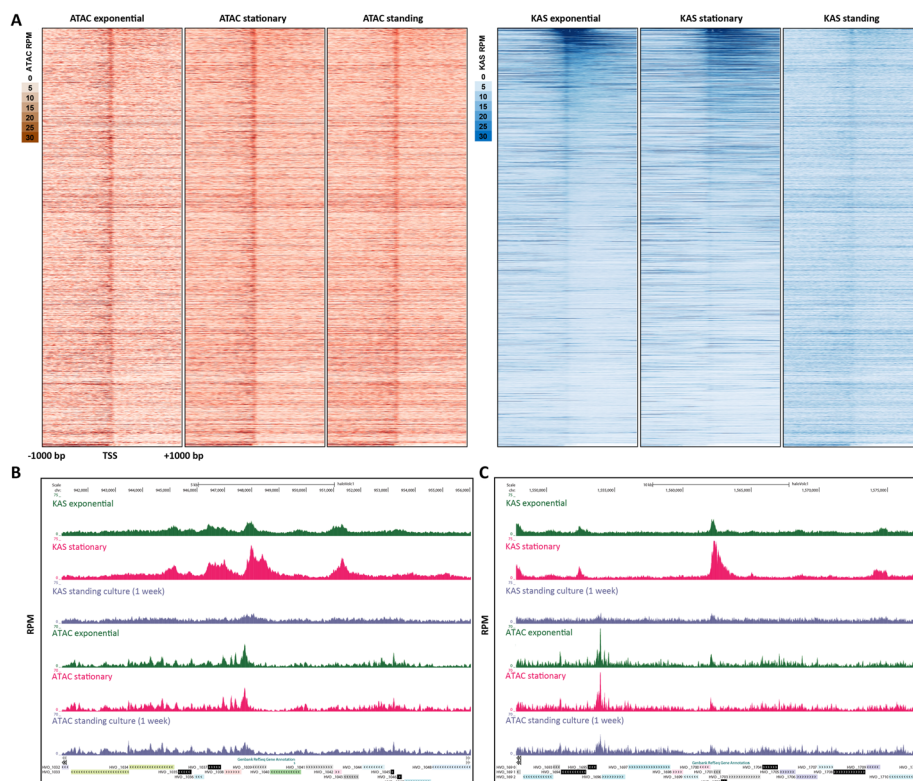
**Fig. 6** Chromatin accessibility does not correlate with transcriptional activity in *H. volcanii*. **A** Differential chromatin accessibility between exponential and stationary conditions. **B** Differential KAS-seq levels between exponential and stationary conditions. **C–E** Lack of correlation between KAS and ATAC signals in exponential and stationary conditions. **F** Lack of correlation between changes in chromatin accessibility and changes in transcriptional activity

In stark contrast to the lack of accessibility changes between non-dormant conditions, we find a large number of genes that display strong differential KAS-seq signals over their gene bodies (Fig. 6B). Thus while we observe no major changes in chromatin accessibility across these two conditions, we do note large-scale changes in RNA polymerase occupancy (as inferred by KAS-seq).

We then quantified the degree of correlation between KAS and ATAC signals (Fig. 6C–E) and found no significant correlation. We also found no correlation between the level of changes in chromatin accessibility and changes in RNA polymerase association with DNA (as inferred by KAS-seq) between exponential and stationary cells (Fig. 6F).

These global observations are supported by the study of individual loci. Figure 7A shows ATAC-seq levels and KAS-seq levels over all genes across different conditions, clearly indicating that these two signals are decoupled from one another. Figure 7B and C depict other examples of genes for which transcriptional activity shifts between conditions yet ATAC-seq profiles are largely identical.

We thus conclude that based on the currently available data, the modulation of chromatin accessibility does not appear to be a major determinant/correlate of transcriptional activity in *Haloferax* archaea.



**Fig. 7** Chromatin accessibility does not correlate with transcriptional activity in *H. volcanii*. **A** Genome-wide heatmaps of ATAC-seq and KAS-seq signals around *H. volcanii* TSSs, sorted by KAS-seq levels in the exponential condition. **B, C** Representative snapshots of genes with significantly altered transcriptional activity between the exponential and standing conditions, but no corresponding changes in chromatin accessibility

## Discussion

In this study, we adapted and applied methods for global profiling of chromatin accessibility and ssDNA in the euryarchaeote *Haloferax volcanii*, revealing the chromatin architecture of this representative of the haloarchaea. We identified several convergent and divergent characteristics with respect to those of conventional eukaryote properties.

The *H. volcanii* genome displays an accessibility landscape very similar to that of eukaryotes with compact genomes such as budding yeast — accessibility peaks are almost exclusively found very close to promoters. Absolute accessibility/protection levels are similar, perhaps slightly lower than those in budding yeast, with a baseline protection level of 85–90%, although these numbers need to be interpreted with some caveats. In yeast and other eukaryotes, applying NOME-seq/dSMF to fixed and native chromatin returns comparable absolute protection values (unpublished data), but it is possible that in *Haloferax* chromatin is nevertheless a bit more “open” than in eukaryotes, as the fixation step provides a “frozen” in time snapshot of protein occupancy on DNA while association of proteins with DNA can still be more dynamic than it is for eukaryotic nucleosomes. We do not observe the strongly positioned nucleosome-like features such as those typical in eukaryotes in *Haloferax*, but instead observe a heterogeneous picture of footprints, consistent with such a dynamic association of chromatin proteins with DNA.

These observations do make a certain amount of sense in the light of recent reports rejecting the nucleosomal packaging of haloarchaeal genomes [22–24] but are also puzzling on their own. Once again, the picture of chromatin accessibility revealed by ATAC-seq in *Haloferax* is nearly identical to that of conventional eukaryotes with nucleosomal chromatin and densely packed genomes. At an arbitrary locus, one would note very few general differences between a *Haloferax* ATAC-seq signal track and a yeast ATAC-seq signal track. Both exhibit strong localized peaks at promoters, and similar levels of absolute protection/occupancy. What protein can provide such high levels of physical protection is at present unknown. Experiments involving in vitro reconstitution of the association of archaeal histones and other putative chromatin proteins with DNA as well as their heterologous expression in yeast might shed light onto this question in the future.

We also note that, unlike what has been reported about bacteria and archaea without histones, the *H. volcanii* genome does not exhibit large-scale domains of diminished and elevated accessibility.

In contrast to the norm in eukaryotes, accessibility at *Haloferax* promoters does not correlate with transcriptional activity. This intriguing observation will require further functional dissection in future work. The idea that chromatin accessibility is not necessary for transcription in archaea is supported by previous observations that transcription by the archaeal RNA Polymerase is slowed, but not blocked by archaeal nucleosomes [65]. However, the molecular determinants of the observed accessibility remain unclear. Nearly all promoters in *H. volcanii* show some level of accessibility (Fig. 7A), but their levels differ greatly between individual genes. How these differential states are specified, and whether they might in fact change in conditions that we have not assayed remains to be determined. MNase-seq studies *Methanothermobacter thermautotrophicus* and *Thermococcus kodakarensis* [29] have suggested

that nucleosome positioning in those organisms is significantly influenced by DNA sequence, but no such strong association was reported for *Haloferox volcanii* [28]. In the currently available datasets we observe anti-correlation between chromatinization and high genomic GC content, but whether this is the primary determinant of nucleosomal or other protein occupancy, and whether this correlation can account for the large differences in promoter accessibility combined with a general absence of such strong peaks elsewhere in the genome observed in *Haloferox* remains uncertain. Generating chromatin accessibility and nucleosome positioning data across other archaeal clades and also in multiple closely related species will allow us to generalize these observations and to train fully powered models that relate sequence to chromatin accessibility, potentially identifying such determinants.

We also made the observation that operons in *Haloferox* display non-uniform levels of single-stranded DNA signal consistent with transcriptional activity, and may in fact consist of multiple distinct transcriptional units. The phenomenon of independent transcription of operon genes has been suggested by some previous studies in *Haloferox* and *Sulfolobus* [66–68], but this is again an observation and a hypothesis that needs to be generalized to and tested in not only more archaeal species, but also in bacteria, where the application of KAS-seq to the study of transcriptional activity may also result in unanticipated findings. In *Haloferox* we were only able to examine several dozen of unambiguous operons (e.g., unidirectional arrays of functionally related genes).

Bacteria are also highly relevant to the other surprising observation we made — the strong ssDNA structure present at the second *Haloferox* CRISPR array, especially in dormant cells that are otherwise mostly transcriptionally silent, but not at the other two CRISPR arrays. The second CRISPR array is uniquely associated with the Cas6 gene. Cas6 is the endoribonuclease that generates guide RNAs [57, 69], thus one possible explanation for the strong ssDNA peak between the CRISPR array and Cas6 is that it represents paused RNA polymerase at the Cas6 promoter. If this explanation is correct, the fact that Cas6 is the only gene in the genome with this property in dormant cells is remarkable and perhaps points to the importance of retaining the ability to process CRISPR transcripts even in a dormant cellular state. Alternatively, the ssDNA structure might be related to the transcription of the CRISPR array itself; however, such an explanation does not speak to the uniqueness of the KAS-seq signal at the second CRISPR array and the absence of the same strong KAS peaks at the other two CRISPR arrays. The second CRISPR array is also separated from both the Cas6 gene and the KAS-seq peak by the *s479* sRNA, the reported function of which is involved in the regulation of zinc transport [58]; how this second array might fit into the overall picture is not clear. The functional significance of elevated chromatin accessibility over CRISPR arrays is also currently unknown. As prokaryotes exhibit an immense variety of CRISPR systems and number and organization of CRISPR arrays, mapping these properties in multiple other prokaryotes would be highly informative.

## Methods

Except where explicitly indicated otherwise, data was processed using custom-written Python scripts (<https://github.com/georgimarinov/GeorgiScripts>).

### ***Haloferax volcanii* cell culture**

*H. volcanii* cells were obtained from the DSMZ German Collection of Microorganisms and Cell Cultures GmbH (Cat # 3757) and cultured in *Halobacterium* media [70, 71], prepared as follows: 7.50 g casamino acids, 10.00 g yeast extract, 3.00 g sodium citrate, 2.00 g KCl, 20.00 g  $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$ , 0.05 g  $\text{FeSO}_4 \times 7 \text{H}_2\text{O}$ , 0.20 mg  $\text{MnSO}_4 \times \text{H}_2\text{O}$ , and 250.00 g NaCl were mixed with distilled water in a total volume of 1 L. The media were then autoclaved and allowed to cool. *H. volcanii* was typically grown at 42 °C, except where otherwise indicated. Cultures were stored at room temperature when not actively growing.

### ***Haloferax volcanii* genome assembly and annotations**

For all analyses, the genome assembly and annotation for the *Haloferax volcanii* DS2 strain, downloaded from the NCBI database, and also matching the haloVolc1 version on the UCSC Microbial Genome Browser [72] (<http://microbes.ucsc.edu/>), was used. The UCSC Microbial Genome Browser was used for visualization of genome browser tracks.

### **ATAC-seq experiments**

Several variations of the ATAC-seq assays were tested. As *H. volcanii* is an archaeon, i.e., a prokaryote without a nucleus, and as it does not have a cell wall (as many other prokaryotes do), the nuclei isolation step typical for ATAC-seq protocols used in eukaryotes was omitted.

For native ATAC-seq, cells ( $\sim 0.1$ ,  $\sim 1$ , or  $\sim 10 \times 10^6$  cells as measured by  $\text{OD}_{600}$ ) were pelleted at 10,000 g for 2 min, then resuspended in 50  $\mu\text{L}$  transposition mix (25  $\mu\text{L}$   $2 \times \text{TD}$  buffer, 2.5  $\mu\text{L}$  Tn5, 22.5  $\mu\text{L}$  ultrapure  $\text{H}_2\text{O}$ ), and incubated at 37 °C for 15 min. The reaction was stopped by adding 250  $\mu\text{L}$  PB Buffer (Qiagen, Cat # 28,006) and purified using the MinElute PCR Purification Kit (Qiagen, Cat # 28006), eluting in 10  $\mu\text{L}$  EB buffer. PCR was carried out by mixing the 10  $\mu\text{L}$  eluate, 10  $\mu\text{L}$   $\text{H}_2\text{O}$ , 2.5  $\mu\text{L}$  i5 primer, 2.5  $\mu\text{L}$  i7 primer, and 25  $\mu\text{L}$  NEBNext High-Fidelity  $2 \times \text{PCR}$  Master Mix, using the following thermocycler program: 3 min at 72 °C, 30 s at 98 °C, 10 cycles of: 98 °C for 10 s, 63 °C for 30 s, 72 °C for 30 s. Final libraries were purified using the MinElute PCR Purification Kit.

For crosslinked ATAC-seq, cells were fixed by adding 37% formaldehyde (Sigma) at a final concentration of either 0.1% or 1% and incubating for 15 min at room temperature. Formaldehyde was then quenched using 2.5 M glycine at a final concentration of 0.25 M. Cells were subsequently centrifuged at 10,000 g for 2 min, washed once in  $1 \times \text{PBS}$ , and centrifuged again at 10,000 g for 2 min. Transposition was carried out as above for 15 min. The reaction was stopped with the addition of 150  $\mu\text{L}$  IP Elution Buffer (1% SDS, 0.1 M  $\text{NaHCO}_3$ ) and 2  $\mu\text{L}$  Proteinase K (Promega, Cat # MC5005), then incubated at 65 °C overnight to reverse crosslinks. DNA was isolated by adding an equal volume of 25:24:1 phenol to chloroform to isoamyl solution, vortexing and centrifuging for 3 min at 14,000 rpm, then purifying the top aqueous phase using the MinElute PCR Purification Kit, eluting in 10  $\mu\text{L}$  EB buffer. Libraries were generated as described above.

Sequencing was carried out on a NextSeq 550 in a  $2 \times 38\text{mer}$  format, to a depth of  $\sim 1 \text{ M}$  read pairs.



### ATAC-seq data processing

Demultiplexed FASTQ files were mapped to the *H. volcanii* genome as  $2 \times 36$ mers using Bowtie [73] (version 1.0.1) with the following settings: `-v 2-k 2-m 1-best-strata`. Duplicate reads were removed using picard-tools (version 1.99).

TSS scores were calculated as the ratio of ATAC signal in the region  $\pm 100$  bp around TSSs versus the ATAC signal of the 100-bp regions centered at the two points  $\pm 2$  kbp of the TSS as previously described [74].

Peak calling was carried out using MACS2 [40] with the following settings: `-g4000000-fBAM-to-large-keep-dup all -nomodel`.

### DNA isolation and naked DNA sequencing

Genomic DNA was isolated by centrifuging cells at 10,000 g and resuspending the pellet in 200  $\mu$ L  $1 \times$  PBS, then using the MagAttract HMW DNA Kit (Qiagen, Cat # 67,563), following the manufacturer's instructions.

Genomic DNA libraries were prepared using 5 ng of DNA in a 50- $\mu$ L transposition reaction ( $x$   $\mu$ L DNA,  $22.5-x$   $\mu$ L  $H_2O$ , 25  $\mu$ L  $2 \times$  TD buffer, 2.5  $\mu$ L Tn5). The reaction was carried out for 5 min at 55  $^{\circ}C$ , then stopped with 250  $\mu$ L PB buffer. DNA was isolated using the MinElute PCR Purification Kit and amplified as described above for ATAC-seq.

### NOMe-seq and dSMF experiments

NOMe-seq/dSMF experiments were carried out as previously described [41], with some modifications. Cells were pelleted at 10,000 g, then crosslinked as described for ATAC-seq at 1% formaldehyde concentration.

Fixed cells were resuspended in 100  $\mu$ L M.CviPI Reaction Buffer (50 mM Tris-HCl pH 8.5, 50 mM NaCl, 10 mM DTT), then treated with M.CviPI by adding 200 U of M.CviPI (NEB), SAM at 0.6 mM and sucrose at 300 mM, and incubating at 30  $^{\circ}C$  for 20 min. After this incubation, 128 pmol SAM and another 100 U of enzyme were added, and a further incubation at 30  $^{\circ}C$  for 20 min was carried out. For dSMF experiments, M.SssI treatment followed immediately, by adding 60 U of M.SssI (NEB), 128 pmol SAM, and  $MgCl_2$  at 10 mM and incubation at 30  $^{\circ}C$  for 20 min. The reaction was stopped by adding an equal volume of Stop Buffer (20 mM Tris-HCl pH 8.5, 600 mM NaCl, 1% SDS, 10 mM EDTA).

Crosslinks were reversed overnight at 65  $^{\circ}C$ , and DNA was isolated using the MinElute PCR Purification Kit (Qiagen, Cat # 28,006).

Enzymatically labeled DNA was then sheared on a Covaris E220, and converted into sequencing libraries following the EM-seq protocol, using the NEBNext Enzymatic Methyl-seq Kit (NEB, Cat # E7120L).

SMF/NOMe-seq libraries were sequenced as  $2 \times 150$ mers on a NovaSeq S4 through Novogene to a depth of  $\sim 100 \times$  for 200-bp fragments.

### NOMe-seq data processing

Adapters were trimmed from reads using Trimmomatic [75] (version 0.36). Trimmed reads were aligned against the *H. volcanii* genome using bwa-meth with default settings.

Duplicate reads were removed using picard-tools (version 1.99). Methylation calls were extracted using MethylDackel (<https://github.com/dpryan79/MethylDackel>). Additional analyses were carried out using custom-written Python scripts (<https://github.com/georgimarinov/GeorgiScripts>).

### KAS-seq experiments

KAS-seq experiments were carried out following the previously published protocol [35] with some modifications in the sequencing library generation part.

Briefly, a 500-mM  $N_3$ -kethoxal solution was brought to 37 °C and then added to 2 mL of culture at a final concentration of 5  $\mu$ M. Cells were then incubated for 5 min at 37 °C in a ThermoMixer at 1000 rpm.

Cells were then pelleted by centrifugation at 10,000 g for 1 min, resuspended in 200  $\mu$ L 1  $\times$  PBS buffer, and DNA was immediately isolated using the Monarch Genomic DNA Purification Kit (NEB, Cat # T3010S), with the modification that elution was carried out with 50  $\mu$ L 25 mM  $K_3BO_3$  solution (pH 7.0).

Biotin was clicked onto kethoxal-modified guanines by mixing 50  $\mu$ L DNA, 2.5  $\mu$ L 20 mM DBCO-PEG4-biotin (Sigma, Cat # 760749; DMSO solution), 10  $\mu$ L 10  $\times$  PBS, and 22.5  $\mu$ L 25 mM  $K_3BO_3$  and incubating at 37 °C for 90 min.

DNA was isolated using AMPure XP beads and eluted in 130  $\mu$ L 25 mM  $K_3BO_3$  (pH 7.0), then sheared on a Covaris E220 for 120 s down to  $\sim$ 150–200 bp.

Libraries were built on beads using the NEBNext Ultra II DNA Library Prep kit (NEB, Cat # E7645L). Biotin pull-down was initiated by pipetting 20  $\mu$ L Dynabeads MyOne Streptavidin T1 beads (ThermoFisher Scientific, Cat # 65306) into DNA lo-bind tubes. Beads were separated on magnet, resuspended in 200  $\mu$ L of 1  $\times$  TWB buffer (Tween Washing Buffer; 5 mM Tris–HCl pH 7.5; 0.5 mM EDTA; 1 M NaCl; 0.05% Tween 20), then separated on magnet again and resuspended in 300  $\mu$ L of 2  $\times$  BB (Binding Buffer; 10 mM Tris–HCl pH 7.5, 1 mM EDTA; 2 M NaCl). The DNA (130  $\mu$ L) was added together with 170  $\mu$ L 0.1  $\times$  TE buffer, and incubated at RT on a rotator for  $\geq$  15 min. Beads were separated on a magnet, resuspended in 200  $\mu$ L of 1  $\times$  TWB, and incubated at 55 °C in a Termomixer for 2 min with shaking at 1000 rpm. Beads were again separated on magnet and the 200- $\mu$ L 55 °C TWB wash step was repeated. Beads were separated on a magnet and resuspended in 50  $\mu$ L 0.1  $\times$  TE.

End repair was carried out by adding 7  $\mu$ L NEB End Repair Buffer and 3  $\mu$ L NEB End Repair Enzyme, incubating at 20 °C for 30 min, then at 65 °C for 30 min.

End repair was followed by adaptor ligation by adding 2.5  $\mu$ L NEB Adaptor, 1  $\mu$ L NEB Ligation Enhancer, and 30  $\mu$ L NEB Ligation Mix, incubating at 20 °C for 20 min, then adding 3  $\mu$ L USER Enzyme and incubating at 37 °C for 15 min. Beads were separated on a magnet, resuspended in 200  $\mu$ L of 1  $\times$  TWB, then incubated at 55 °C in a Thermomixer for 2 min with shaking at 1000 rpm. Subsequently, beads were separated on a magnet and resuspended in 100  $\mu$ L of 0.1  $\times$  TE, separated on a magnet again, resuspended in 15  $\mu$ L of 0.1  $\times$  TE Buffer, and transferred to PCR tubes.

Beads were then incubated at 98 °C for 10 min, and libraries were amplified by adding 5  $\mu$ L of i5 primer, 5  $\mu$ L of i7 primer and 25  $\mu$ L of 2  $\times$  Q5 Hot Start Polymerase Mix, using

the following PCR program: 30 s at 98 °C; 15 cycles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 30 s; and a final extension at 72 °C for 5 min.

Beads were separated on a magnet and the final libraries were purified from the supernatant using 50 µL AMPure XP beads, eluting in 0.1 × TE buffer.

Sequencing was carried out on a NextSeq 550 in a 2 × 38mer format, to a depth of 1–10 M read pairs.

### KAS-seq data processing

Demultiplexed FASTQ files were mapped to the *H. volcanii* genome as 2 × 36mers using Bowtie [73] with the following settings: -v 2-k 2-m 1–best–strata. Duplicate reads were removed using picard-tools (version 1.99).

### Multimapping reads analysis

For the purpose of examining repetitive regions in the genome, such as the rRNA operons, which exist in two identical copies in the genome, and are thus not uniquely mappable, reads were mapped with the -a option instead of -k 2-m 1. Normalization was carried out as previously described and discussed [76].

### Differential accessibility/KAS-seq analysis

The analysis of differential chromatin accessibility as measured using ATAC-seq or enriched for KAS-seq signal was carried out using DESeq2 [77]. Read counts were calculated over promoters or gene bodies and used as input into DESeq2.

### External sequencing datasets

MNase-seq datasets for *H. volcanii* were downloaded from NCBI accession PRJNA174818 [28] and processed as described above for ATAC-seq and KAS-seq.

ATAC-seq for *Suflolobus islandicus* [43] was downloaded through the Short Read Archive (SRA) from BioProject accession 814106.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-023-03095-5>.

**Additional file 1: Supplementary Figure 1.** TSS enrichment levels in ATAC-seq data for *Suflolobus islandicus*. **Supplementary Figure 2.** Correlation between the extent of chromatinization and GC content in *Haloflexax*. **Supplementary Figure 3.** Coordination between chromatin accessibility and transcriptional activity within *H. volcanii* operons. **Supplementary Figure 4.** Differential ATAC-seq in KAS-seq analysis for standing *H. volcanii* cells. **Additional file 2.** Peer review history.

### Acknowledgements

The authors would like to thank Samuel Kim for input on the KAS-seq protocol, Matthew P. Swaffner for technical assistance, Zohar Shipony and members of the Greenleaf and Kundaje labs, and Tobias Warnecke, Fabian Blombach, and Anita Marchfelder for helpful discussions and suggestions for improvements to the text and analysis. This work was supported by NIH grants (P50HG007735, RO1 HG008140, U19AI057266, and UM1HG009442 to W.J.G.; 1UM1HG009436 to W.J.G. and A.K.; 1DP2OD022870-01 and 1U01HG009431 to A.K.), the Rita Allen Foundation (to W.J.G.), the Baxter Foundation Faculty Scholar Grant, and the Human Frontiers Science Program grant RGY006S (to W.J.G.). W.J.G. is a Chan Zuckerberg Biohub investigator and acknowledges grants 2017-174468 and 2018-182817 from the Chan Zuckerberg Initiative.

### Review history

The review history is available as Additional file 2.

### Peer review information

Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Authors' contributions**

G.K.M. conceptualized the study; carried out the cell culture; performed the ATAC-seq, KAS-seq, and NOME-seq/dSMF experiments; analyzed the data; and wrote the manuscript, with input from W.J.G. S.T.B. performed the ATAC-seq, KAS-seq, and NOME-seq/dSMF experiments. T.W. and C.H. provided key reagents. A.K. and W.J.G. supervised the study.

**Availability of data and materials**

The sequencing datasets generated for this study can be accessed from GEO accession GSE207470 [78].

Previously published [28] MNase-seq and control datasets for *Haloflexax* were obtained from SRA accessions SRX188663 [79], SRX188665 [80], and SRX185902 [81].

Previously published [43] *Sulfolobus* ATAC-seq data was downloaded from SRA under project number PRJNA814106 [82].

The data processing and visualization code used can be found on GitHub [83] and Zenodo [84].

**Declarations****Ethics approval and consent to participate**

Not applicable to this study.

**Competing interests**

W.J.G. is a consultant and equity holder for 10x Genomics, Guardant Health, Quantapore, and Ultima Genomics and cofounder of Protillion Biosciences and is named on patents describing ATAC-seq.

A.K. is a consulting Fellow with Illumina; a member of the SAB of OpenTargets (GSK), PatchBio, and SerImmune; and a scientific co-founder of RavelBio.

Received: 14 February 2023 Accepted: 24 October 2023

Published online: 06 November 2023

**References**

1. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*. 1977;74(11):5088–90.
2. Lake JA, Henderson E, Oakes M, Clark MW. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc Natl Acad Sci U S A*. 1984;81(12):3786–90.
3. Lake JA. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature*. 1988;331(6152):184–6.
4. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. The archaeobacterial origin of eukaryotes. *Proc Natl Acad Sci U S A*. 2008;105(51):20356–61.
5. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*. 2015;521(7551):173–9.
6. Koonin EV. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol*. 2010;11(5):209.
7. Koonin EV, Yutin N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb Perspect Biol*. 2014;6(4):a016188.
8. Marinov GK, Lynch M. Diversity and divergence of dinoflagellate histone proteins. *G3 (Bethesda)*. 2015; 6(2):397–422.
9. Marinov GK, Lynch M. Conservation and divergence of the histone code in nucleomorphs. *Biol Direct*. 2016;11(1):18.
10. Postberg J, Forcob S, Chang WJ, Lipps HJ. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol*. 2010;10:259.
11. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001;293(5532):1074–80.
12. Searcy DG, Delange RJ. *Thermoplasma acidophilum* histone-like protein. Partial amino acid sequence suggestive of homology to eukaryotic histones. *Biochim Biophys Acta*. 1980; 609(1):197–200.
13. Searcy DG, Stein DB. Nucleoprotein subunit structure in an unusual prokaryotic organism: *Thermoplasma acidophilum*. *Biochim Biophys Acta*. 1980;609(1):180–95.
14. Sandman K, Reeve JN. Archaeal histones and the origin of the histone fold. *Curr Opin Microbiol*. 2006;9:520–5.
15. Henneman B, van Emmerik C, van Ingen H, Dame RT. Structure and function of archaeal histones. *PLoS Genet*. 2018;14(9): e1007582.
16. Stevens KM, Swadling JB, Hoher A, Bang C, Gribaldo S, Schmitz RA, Warnecke T. Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proc Natl Acad Sci U S A*. 2020;117(52):33384–95.
17. Mattioli F, Bhattacharyya S, Dyer PN, White AE, Sandman K, Burkhart BW, Byrne KR, Lee T, Ahn NG, Santangelo TJ, Reeve JN, Luger K. Structure of histone-based chromatin in Archaea. *Science*. 2017;357(6351):609–12.
18. Henneman B, Brouwer TB, Erkelens AM, Kuijntjes GJ, van Emmerik C, van der Valk RA, Timmer M, Kirolos NCS, van Ingen H, van Noort J, Dame RT. Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res*. 2021;49(8):4338–49.
19. Bowerman S, Wereszczynski J, Luger K. Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. *eLife*. 2021;10:e65587.
20. Zhang Z, Guo L, Huang L. Archaeal chromatin proteins. *Sci China Life Sci*. 2012;55(5):377–85.
21. Sandman K, Reeve JN. Archaeal chromatin proteins: different structures but common function? *Curr Opin Microbiol*. 2005;8(6):656–61.
22. Dulmage KA, Todor H, Schmid AK. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. *mBio*. 2015; 6(5):e00649–15

23. Sakrikar S, Hackley RK, Pastor MM, Darnell CL, Vreugdenhil A, Schmid AK. Comparative analysis of genomewide protein-DNA interactions across domains of life reveals unique binding patterns for hypersaline archaeal histones. *bioRxiv*. 2022; 2022.03.22.485428.
24. Sakrikar S, Schmid AK. An archaeal histone-like protein regulates gene expression in response to salt stress. *Nucleic Acids Res*. 2021;49(22):12732–43.
25. Hocher A, Borrel G, Fadhlouli K, Brugère JF, Gribaldo S, Warnecke T. Growth temperature and chromatinization in archaea. *Nat Microbiol*. 2022; 7(11):1932–1942.
26. Knüppel R, Trahan C, Kern M, Wagner A, Grünberger F, Hausner W, Quax TEF, Albers SV, Oeffinger M, Ferreira-Cerca S. Insights into synthesis and function of KsgA/Dim1-dependent rRNA modifications in archaea. *Nucleic Acids Res*. 2021; 49(3):1662–1687.
27. Jevtic Z, Stoll B, Pfeiffer F, Sharma K, Urlaub H, Marchfelder A, Lenz C. The response of *Haloflex volcanii* to salt and temperature stress: a proteome study by label-free mass spectrometry. *Proteomics*. 2019;19(20): e1800491.
28. Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader GD, Giaever G, Nislow C. Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife*. 2012;1: e00078.
29. Nalabothula N, Xi L, Bhattacharyya S, Widom J, Wang JP, Reeve JN, Santangelo TJ, Fondufe-Mittendorf YN. Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics*. 2013;14:391.
30. Rojec M, Hocher A, Stevens KM, Merckenschlager M, Warnecke T. Chromatinization of *Escherichia coli* with archaeal histones. *Elife*. 2019;8: e49038. <https://doi.org/10.7554/eLife.49038>.
31. Hocher A, Rojec M, Swadling JB, Esin A, Warnecke T. The DNA-binding protein HTa from *Thermoplasma acidophilum* is an archaeal histone analog. *Elife*. 2019;8: e52542.
32. Ofer S, Blombach F, Erkelens A, Barker D, Schwab S, Smollett K, Matelska D, Fouqueau T, Nvd V, Kent N, Dame R, Werner F. DNA-bridging by an archaeal histone variant via a unique tetramerisation interface. *Research Square*. 2023. <https://doi.org/10.21203/rs.3.rs-2183355/v1>.
33. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8.
34. Krebs AR, Imanici D, Hoerner L, Gaidatzis D, Burger L, Schübeler D. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol Cell*. 2017;67(3):411–422.e4.
35. Wu T, Lyu R, You Q, He C. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity *in situ*. *Nat Methods*. 2020;17(5):515–23.
36. Mullakhanbhai MF, Larsen H. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch Microbiol*. 1975; 104(3):207–214.
37. Pohlschroder M, Schulze S. *Haloflex volcanii*. *Trends Microbiol*. 2019;27(1):86–7.
38. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14(10):959–62.
39. Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, Madupu R, Robinson J, Khouri H, Ren Q, Lowe TM, Maupin-Furlow J, Pohlschroder M, Daniels C, Pfeiffer F, Allers T, Eisen JA. The complete genome sequence of *Haloflex volcanii* DS2, a model archaeon. *PLoS ONE*. 2010;5(3): e9605.
40. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012;7(9):1728–40.
41. Shipony Z, Marinov GK, Swaffer MP, Sinnott-Armstrong NA, Skotheim JM, Kundaje A, Greenleaf WJ. Longrange single-molecule mapping of chromatin accessibility in eukaryotes. *Nat Methods*. 2020;17(3):319–27.
42. Melfi MD, Lasker K, Zhou X, Shapiro K. ATAC-seq reveals megabase-scale domains of a bacterial nucleoid. *bioRxiv*. 2021; 2021.01.09.426053.
43. Badel C, Samson RY, Bell SD. Chromosome organization affects genome evolution in *Sulfolobus* archaea. *Nat Microbiol*. 2022. <https://doi.org/10.1038/s41564-022-01127-7>.
44. Cajili MKM, Prieto EI. Interplay between Alba and Cren7 Regulates Chromatin Compaction in *Sulfolobus solfataricus*. *Biomolecules*. 2022;12(4):481.
45. Bell SD, Botting CH, Wardleworth BN, Jackson SP, White MF. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science*. 2002;296(5565):148–51.
46. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res*. 2012;22(12):2497–506.
47. Ouellette M, Jackson L, Chimileski S, Papke RT. Genome-wide DNA methylation analysis of *Haloflex volcanii* H26 and identification of DNA methyltransferase related PD-(D/E)XK nuclease family protein HVO A0006. *Front Microbiol*. 2015;6:251.
48. Oberbeckmann E, Wolff M, Krietenstein N, Heron M, Ellins JL, Schmid A, Krebs S, Blum H, Gerland U, Korber P. Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res*. 2019;29(12):1996–2009.
49. Chereji RV, Eriksson PR, Ocampo J, Prajapati HK, Clark DJ. Accessibility of promoter DNA is not the primary determinant of chromatin-mediated gene regulation. *Genome Res*. 2019;29(12):1985–95.
50. Babski J, Haas KA, Nather-Schindler D, Pfeiffer F, Forstner KU, Hammelmann M, Hilker R, Becker A, Sharma CM, Marchfelder A, Soppa J. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloflex volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics*. 2016;17(1):629.
51. Pastor MM, Sakrikar S, Rodriguez DN, Schmid AK. Comparative analysis of rRNA removal methods for RNA-Seq differential expression in halophilic archaea. *Biomolecules*. 2022;12(5):682.
52. Core L, Adelman K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev*. 2019;33(15–16):960–82.
53. Blombach F, Fouqueau T, Matelska D, Smollett K, Werner F. Promoter-proximal elongation regulates transcription in archaea. *Nat Commun*. 2021;12(1):5524.
54. Mojica FJ, Rodriguez-Valera F. The discovery of CRISPR in archaea and bacteria. *FEBS J*. 2016;283(17):3162–9.

55. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol.* 2005; 60(2):174–182.
56. Mojica FJ, Ferrer C, Juez G, Rodríguez-Valera F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol Microbiol.* 1995; 17(1):85–93.
57. Maier LK, Stachler AE, Brendel J, Stoll B, Fischer S, Haas KA, Schwarz TS, Alkhnbashi OS, Sharma K, Urlaub H, Backofen R, Gophna U, Marchfelder A. The nuts and bolts of the Haloferax CRISPR-Cas system I-B. *RNA Biol.* 2019;16(4):469–80.
58. Märkle P, Maier LK, Maaß, Hirschfeld C, Bartel J, Becher D, VoßB, Marchfelder A. A small RNA is linking CRISPR-Cas and zinc transport. *Front Mol Biosci.* 2021; 8:640440.
59. Price MN, Arkin AP, Alm EJ. The life-cycle of operons. *PLoS Genet.* 2006;2(6): e96.
60. Garrett RA, Dalgaard J, Larsen N, Kjems J, Mankin AS. Archaeal rRNA operons. *Trends Biochem Sci.* 1991;16(1):22–6.
61. Koide T, Reiss DJ, Bare JC, Pang WL, Facciotti MT, Schmid AK, Pan M, Marzolf B, Van PT, Lo FY, Pratap A, Deutsch EW, Peterson A, Martin D, Baliga NS. Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Syst Biol.* 2009;5:285.
62. Laass S, Monzon VA, Kliemt J, Hammelmann M, Pfeiffer F, Förstner KU, Soppa J. Characterization of the transcriptome of *Haloferax volcanii*, grown under four different conditions, with mixed RNA-Seq. *PLoS ONE.* 2019;14(4): e0215986.
63. Grünberger F, Knüppel R, Jüttner M, Fenk M, Borst A, Reichelt R, Hausner W, Soppa J, Ferreira-Cerca S, Grohmann D. Exploring prokaryotic transcription, operon structures, rRNA maturation and modifications using Nanoporebased native RNA sequencing. *bioRxiv.* 2019; 2019.12.18.880849.
64. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet.* 2019;20(4):207–22.
65. Xie Y, Reeve JN. Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J Bacteriol.* 2004;186:3492–8.
66. León-Sobrino C, Kot WP, Garrett RA. Transcriptome changes in STSV2-infected *Sulfolobus islandicus* REY15A undergoing continuous CRISPR spacer acquisition. *Mol Microbiol.* 2016; 99(4):719–728
67. Maier LK, Benz J, Fischer S, Alstetter M, Jaschinski K, Hilker R, Becker A, Allers T, Soppa J, Marchfelder A. Deletion of the Sm1 encoding motif in the lsm gene results in distinct changes in the transcriptome and enhanced swarming activity of *Haloferax* cells. *Biochimie.* 2015;117:129–137.
68. Berkemer SJ, Maier LK, Amman F, Bernhart SH, Wörtz J, Märkle P, Pfeiffer F, Stadler PF, Marchfelder A. Identification of RNA 3' ends and termination sites in *Haloferax volcanii*. *RNA Biol.* 2020; 17(5):663–676.
69. Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 2008;22(24):3489–96.
70. Torreblanca M, Rodríguez-Valera F, Juez G, Ventosa A, Kamekura M, Kates M. Classification of non-alkaliphilic halobacteria based on numerical taxonomy and polar lipid composition, and description of *Haloarcula* gen. nov. and *Haloferax* gen. nov. *Syst Appl Microbiol.* 1986;8:89–99.
71. Rodríguez-Valera F, Juez G, Kushner DJ. *Halobacterium mediterranei* spec. nov., a new carbohydrate-utilizing extreme halophile. *System Appl Microbiol.* 1983;4:369–81.
72. Chan PP, Holmes AD, Smith AM, Tran D, Lowe TM. The UCSC Archaeal Genome Browser: 2012 update. *Nucleic Acids Res.* 2012; 40(Database issue):D646–52.
73. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
74. Marinov GK, Shipony Z. Interrogating the accessible chromatin landscape of eukaryote genomes using ATAC-seq. *Methods Mol Biol.* 2021;2243:183–226.
75. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
76. Marinov GK, Wang J, Handler D, Wold BJ, Weng Z, Hannon GJ, Aravin AA, Zamore PD, Brennecke J, Toth KF. Pitfalls of mapping high-throughput sequencing data to repetitive sequences: Pivi's genomic targets still not identified. *Dev Cell.* 2015;32(6):765–71.
77. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
78. Marinov GK, Bagdatli ST, Wu T, He C, Kundaje A, Greenleaf WJ. The chromatin landscape of the euryarchaeon *Haloferax volcanii*. Datasets. *Gene Expression Omnibus.* 2023. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207470>.
79. Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader G, Giaever G, Nislow C. Nucleosome data (MNase-seq) ID SRX188663. Publicly available at SRA. 2012. (<http://www.ncbi.nlm.nih.gov/sra/SRX188663>).
80. Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader G, Giaever G, Nislow C. undigested DNA control ID SRX188665. Publicly available at SRA. 2012. (<http://www.ncbi.nlm.nih.gov/sra/SRX188665>).
81. Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader G, Giaever G, Nislow C. naked DNA control (MNaseq) ID SRX185902. Publicly available at SRA. 2012. (<http://www.ncbi.nlm.nih.gov/sra/SRX185902>).
82. The influence of chromosome organization on genome evolution in *Sulfolobus* archaea. ID PRJNA814106. Publicly available at SRA. 2022. (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA814106>).
83. Marinov GK. GeorgiScripts. GitHub. 2023. (<https://github.com/georgimarinov/GeorgiScripts>).
84. Marinov GK. GeorgiScripts. Zenodo. 2023. (<https://doi.org/10.5281/zenodo.5877027>).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.