Chromatin is an ancient innovation conserved between Archaea and Eukarya

Ron Ammar1,2, Dax Torti2, Kyle Tsui1,3, Marinella Gebbia1,2, Tanja Durbic2, Gary D Bader1,2, Guri Giaever1,3, Corey Nislow1,*

1Department of Molecular Genetics, University of Toronto, Toronto, Canada; 2Donnelly Centre, University of Toronto, Toronto, Canada; 3Department of Pharmaceutical Sciences, University of Toronto, Toronto, Canada

Abstract The eukaryotic nucleosome is the fundamental unit of chromatin, comprising a protein octamer that wraps ~147 bp of DNA and has essential roles in DNA compaction, replication and gene expression. Nucleosomes and chromatin have historically been considered to be unique to eukaryotes, yet studies of select archaea have identified homologs of histone proteins that assemble into tetrameric nucleosomes. Here we report the first archaeal genome-wide nucleosome occupancy map, as observed in the halophile Haloferax volcanii. Nucleosome occupancy was compared with gene expression by compiling a comprehensive transcriptome of Hfx. volcanii. We found that archaeal transcripts possess hallmarks of eukaryotic chromatin structure: nucleosome-depleted regions at transcriptional start sites and conserved −1 and +1 promoter nucleosomes. Our observations demonstrate that histones and chromatin architecture evolved before the divergence of Archaea and Eukarya, suggesting that the fundamental role of chromatin in the regulation of gene expression is ancient.

DOI: 10.7554/eLife.00078.001

Introduction

Archaeal nucleosome core particles protect ~60 bp of DNA, approximately half that of eukaryotic nucleosomes, as demonstrated by the landmark work of Pereira et al. (1997). Comparing both eukaryotic and archaeal nucleosomes, the former is an octamer composed of heterodimers of histones H2A, H2B, H3 and H4 whereas the latter histones assemble from homologs of H3 and H4 proteins (Pereira and Reeve, 1998; Talbert and Henikoff, 2010). Archaeal histones can form both homodimers and heterodimers, as well as homotetramers, whereas eukaryotic histones contain hydrophobic dimerization surfaces that restrict assembly of the octamer from H2A-H2B and H3-H4 heterodimers (Sandman and Reeve, 2006; Talbert and Henikoff, 2010).

Using single-nucleotide resolution maps of archaeal nucleosome occupancy and gene expression, we demonstrate that the architecture of archaeal chromatin and the occupancy of its nucleosomes along transcription units are conserved. We constructed a nucleosome occupancy map of the halophilic archaeon Haloferax volcanii, a member of the phylum euryarchaeota, originally discovered in the highly saline sediment of the Dead Sea (Mullakhanbhai and Larsen, 1975). The genome of Hfx. volcanii has an average GC content of 65% and a total genome length of 4 Mb (Hartman et al., 2010) composed of five circular genetic elements: a 2.8 Mb main chromosome, three smaller chromosomes pHV1, pHV3 and pHV4 and the plasmid pHV2. It is highly polyploid with ~15 genome copies during exponential growth and ~10 during stationary phase (Breuert et al., 2006). The histone protein of Hfx. volcanii, hstA (HVO_0520), has a domain architecture containing two distinct histone fold domains within the same peptide that heterodimerize similar to the Methanopyrus kandleri histone (HMk) (Geer et al., 2002; Talbert and Henikoff, 2010; Marchler-Bauer et al., 2011).
Results

We cultured *Hfx. volcanii* in rich media containing 2 M NaCl (*Mullakhanbhai and Larsen, 1975*). Genomic DNA was cross-linked and digested with micrococcal nuclease (MNase), with cell disruption accomplished by bead-beating (*Tsui et al., 2012*). Nucleosome-bound cross-linked genomic regions are protected from MNase digestion, in contrast to the linker DNA between nucleosomes. Mononucleosome-sized (50–60 bp) DNA fragments were gel purified and libraries were sequenced on an Illumina HiSeq2000 (*Figure 1A*). Sequence reads were aligned to the published *Hfx. volcanii* genome (*Hartman et al., 2010*) to generate a genome-wide nucleosome occupancy map. Controls included crosslinked DNA without MNase digestion as well as MNase treated nucleosome-free genomic DNA. The nucleosome occupancy data was significantly different than the control MNase digest of deproteinized ‘naked’ genomic DNA (*r = 0.071*), indicating that the nucleosome map is unaffected by any potential MNase sequence bias (*Chung et al., 2010*).
To determine nucleosome midpoints, we smoothed the occupancy data using a symmetrical convolution sum with a Gaussian filter (Smith, 1997). Extrema were detected in the smoothed signal, and maxima were defined as nucleosome midpoints. In the smoothed signal, the mean peak-to-peak distance for the main chromosome was 68.5 bp in genic regions and 76.1 bp in non-genic regions. Genic regions were defined as the transcribed region plus 40 bp (the average promoter length based on Palmer and Daniels, 1995) upstream of the 5′ end. We observed a greater nucleosome density in *Hfx. volcanii* vs all eukaryotes likely due to the shorter length of DNA wrapped around the archaeal histone tetramer (Pereira et al., 1997). Based on our data, the *Hfx. volcanii* genome has 14.2 nucleosomes per kilobase compared to 5.2 nucleosomes per kilobase in *Saccharomyces cerevisiae*. The resulting map reveals a periodic pattern similar to that seen in all eukaryotes examined to date; with protected regions appearing as peaks and linker regions as troughs. Sequence analysis of the entire nucleosome map showed that nucleosome midpoints were enriched with G/C nucleotides from 61.4% GC at the edge of the protected fragment to 74.6% GC at the midpoint (dyad). We found an increase of G/C nucleotides and a decrease in A/T nucleotides at the midpoint, as described recently for human cell lines (Valouev et al., 2011). In contrast to previous studies in eukaryotes, we did not observe a periodicity in dinucleotide frequency relative to the nucleosome midpoint (Satchwell et al., 1986; Bailey et al., 2000; Albert et al., 2007).

We next investigated the relationship between nucleosome occupancy and gene expression. The existing genome annotation for *Haloferax* is derived almost exclusively from ORF predictions (Hartman et al., 2010). To augment these predictions, we used deep sequencing to create a high confidence transcriptome of the main chromosome of *Hfx. volcanii*. This map allowed us to define both 5′UTR...
lengths, transcriptional start sites (TSSs) and transcriptional termination sites (TTSs). Total RNA was extracted from *Hfx. volcanii* cells, repetitive RNA was partially depleted via duplex-specific nuclease (DSN) normalization followed by RNA-seq (see ‘Materials and methods’) (Zhulidov et al., 2004). Transcript sequences were aligned, assembled and quantified using TopHat and the Genome Analysis Toolkit (Trapnell et al., 2009; McKenna et al., 2010) and transcript boundaries were further trimmed based on RNA-seq coverage information, as described previously (Wurtzel et al., 2010). The final set of transcripts were manually curated yielding 3059 transcriptional units in *Hfx. volcanii*, a number that is greater than observed previously in the comparable transcriptome of the sulfur-metabolizing archaeon *Sulfolobus solfataricus* (Wurtzel et al., 2010) but fewer than the 4073 predicted *Hfx. volcanii* genes. It is likely that in the rich media conditions used in this study, not all genes are expressed. Specifically 75% of the predicted transcripts were detectably expressed, and this fraction is consistent with observations obtained for yeast gene expression in rich media (David et al., 2006). Thirty-two novel transcripts (absent from the predicted sequence annotation) were identified in the RNA-seq data. Most of these 32 transcripts lack significant sequence homologs, and several were classified as transposases with paralogs in *Hfx. volcanii* (Supplementary file 1). Notably, the gene that was most highly expressed in the transcriptome (NTRANS_0004) was not previously annotated and contains a putative N-Acyltransferase (NAT) superfamily domain. Homology searches revealed that this transcript appears to be restricted to the genomes of other halophilic archaea (Altschul et al., 1990). The architecture of this domain is homologous to chain A of the well-characterized histone acetyltransferases Gcn5, Gna1, Hpa2 in *S. cerevisiae*, suggesting a possible role for this transcript in regulating transcription via histone acetylation (Marchler-Bauer et al., 2011). Additional acetyltransferases with a similar architecture have been implicated in bacteriophage-encoded DNA modifiers as well as in cold and ethanol tolerance in yeast (Du and Takagi, 2007; Kaminska and Bujnicki, 2008). Thus, while histone post-translational modifications have not been observed in archaeal histones (Forbes et al., 2004), our observation suggests that some rudimentary control over chromatin accessibility may occur via the action of ancient NAT family members. Furthermore acetyltransferase and deacetylase orthologs, which appear to have enzymatic activity based on their sensitivity to the histone deacetylase (HDAC) inhibitor trichostatin A have been identified in *Hfx. volcanii* (Altman-Price and Mевrечч, 2009). In our subsequent analysis, we focused on all genes we empirically determined to be expressed.

In eukaryotes, the TSS of the majority of expressed genes is characterized by a nucleosome-depleted region (NDR) (Jiang and Pugh, 2009). This NDR is flanked by the well-positioned −1 and the +1 nucleosomes. These regions direct RNA polymerase II to initiate transcription and influence the binding of promoter regulatory elements (Jiang and Pugh, 2009). This stereotypical pattern of nucleosome depletion at promoters and well-ordered nucleosomes in gene bodies is found in all eukaryotes, including yeast, *Drosophila*, *Arabidopsis* and humans. Using the RNA-seq-derived transcripts for *Hfx. volcanii*, we computed the degree of aggregate nucleosome occupancy for the 2343 transcripts on the main chromosome, and found that the NDR and −1 and +1 nucleosomes are conserved in *Hfx. volcanii* (Figure 2) suggesting that the interplay between chromatin and transcription is conserved in archaeal promoters. We generated nucleosome occupancy profiles for each transcript and clustered them hierarchically. Differential nucleosome density was observed with profiles encompassing between four to six nucleosomes in a 400-bp DNA segment spanning 200 bp on each side of the TSS (Figure 2C). NDRs at TTSs are also observed, and similar to those found in eukaryotes (Lee et al., 2007) they are less prominent than promoter NDRs in *Hfx. volcanii* (Figure 3).

**Discussion**

Our study establishes that genome-wide nucleosome occupancy is conserved between archaea and eukaryotes (Figure 4). We further show that the nucleosomal protected fragments and NDRs are shorter in archaea than in eukaryotes. Our findings are particularly noteworthy because *Hfx. volcanii* likely resembles a deeply rooted ancestor that possessed eukaryotic genome architecture hallmarks such as histones, as well as bacterial hallmarks such as the Shine-Dalgarno sequence (Sartorius-Neef and Pfeifer, 2004). Archaeal histone tetramers likely resemble an ancestral state of chromatin, as it has been observed that functional (H3-H4), tetramers can be formed in vitro from eukaryotic histones, and these tetramers are functional; they facilitate more rapid transcription in vitro compared to native histone octamers (Puerta et al., 1993). The observation that archaea contain (H3-H4), tetramers is consistent with the proposal that formation of the canonical eukaryotic nucleosome octamer begins with (H3-H4) tetramer assembly (Talbert and Henikoff, 2010).
Our study demonstrates that both histones and chromatin architecture arose before the divergence of Archaea and Eukarya, suggesting that the fundamental role of chromatin in the regulation of gene expression is ancient. As well, owing to the small bacterial-sized archaean genome, we suggest that archaean chromatin is not required for genome compaction. This leads us to postulate that higher-order chromatin (Sajan and Hawkins, 2012) is a eukaryotic invention and that archaean chromatin is necessary but not sufficient for genome compaction. Additionally our observations provide a rich dataset that addresses the evolution of chromatin and its fundamental role in the regulation of gene expression.

Figure 2. Nucleosome occupancy in Haloferax volcanii. (A) Degree of normalized nucleosome occupancy in aggregate for the main chromosome. As observed in eukaryotes, there is a prominent nucleosome-depleted region (NDR) at the transcriptional start site (TSS) preceded by a −1 nucleosome and followed by a +1 nucleosome, demonstrating that promoter genome architecture is conserved between archaea and eukaryotes. (B) Hierarchical clustergram for the 2343 expressed transcripts on the main Haloferax chromosome. Green represents nucleosome-depleted regions and red represents occupied regions. (C) The clustered heatmap was subdivided into the largest six subclades, and differential density of nucleosomes can be observed with occupancy profile clusters containing between four to six nucleosomes. DOI: 10.7554/eLife.00078.004
**Materials and methods**

**Sample preparation**

*Halofex volcanii* DS2 cells (obtained from the ATCC) were grown to mid-log phase at 42°C in ATCC 974 Halobacterium medium containing 2 M NaCl. Cells were fixed with 2% formaldehyde for 30 min then quenched with 125 mM of glycine for 5 min. An unfixed control sample was also prepared to serve as a deproteinized, ‘naked’ DNA control, as described previously ([Chung et al., 2010](#)). Cells were pelleted and snap frozen prior to MNase digestion and DNA extraction. Frozen cells were processed according to a modified protocol from [Rizzo et al. (2011)](#) and [Tsui et al. (2012)](#). Samples were digested with increasing concentrations of MNase and a no MNase control. After digestion, fragments 50–60 bp in length were size-selected using an Agilent Bioanalyzer High Sensitivity chip (Agilent, Santa Clara, CA, part# 5067-4626) and further processed for Illumina deep sequencing. Nucleosomal and genomic libraries were pooled equally according to qPCR quantitation, and sequenced using v3 chemistry on one single-read HiSeq2000 lane (50 x 8). Samples were demultiplexed using an 8-bp index read at the end of read 1.

**Sequence read filtering and alignment**

Illumina sequencers require the ligation of an adapter oligonucleotide to facilitate cluster formation on the flow cell. Because the library inserts were short (∼60 bp), many sequence reads extended into the Illumina adapter sequences. The adapter subsequences were computationally trimmed to ensure maximal read mapping. Then, using a sequence quality cutoff of Phred20, reads were trimmed from both 5’ and 3’ ends to ensure accurate mapping. These trimmed reads from control and MNase-treated genomic DNA were aligned to the *Hfx. volcanii* DS2 genome using the Bowtie 2 gapped short read aligner ([Langmead and Salzberg, 2012](#)). Sequence coverage was computed using the Genome Analysis Toolkit (GATK) depth of coverage walker, which revealed the periodicity in the occupancy data ([DePristo et al., 2011](#)).

**Nucleosome identification**

To detect nucleosome midpoint positions, sequence data were Gaussian-smoothed as described previously by [Shivaswamy et al. (2008)](#) and [Kaplan et al. (2009)](#). This is appropriate because signals generated by processes that are random, such as sequence coverage noise, usually have a probability density function defined by a Gaussian distribution ([Smith, 1997](#)).
The Gaussian filter was defined as:

\[ G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \]

where \( \mu \) is the mean of the distribution and \( \sigma \) is the standard deviation.

A symmetrical convolution sum was applied with the following format:

\[ y[i] = \sum_{j=-M}^{M} h[j] \times x[i-j], \]

where \( M \) is an integer bandwidth, \( y[j] \) is the output, \( x[j] \) is the input and \( h[j] \) is an \( M \)-point function.

So, to smooth the coverage data, we applied the following convolution sum:

\[ y[i] = \sum_{j=-M/2}^{M/2} G[j] \times x[i-j], \]

where \( \sigma = \frac{M}{6} \). The interval length \( M \) is constrained to \( 6\sigma \) because this encompasses 99.75% of the Gaussian (Smith, 1997).

We also optimized nucleosome midpoint detection by convoluting a two-pass simple moving average (SMA) filter, but the Gaussian filter detected midpoints with greater resolution. Optimal interval size for the Gaussian convolution sum, as determined by Pearson’s correlation coefficient with the raw data, was 27 bp. For the two-pass SMA it was 40 bp for first-pass and 15 bp for second-pass.

Nucleosome occupancy was normalized genome-wide by transforming sequence coverage data into binary-like data that existed in states of ‘occupied’, ‘depleted’ or transitioning between those two states. This final occupancy map was used to define nucleosome positions. Nucleosome occupancy profiles were clustered hierarchically by average linkage using Pearson’s correlation coefficient as the similarity metric in the Cluster 3.0 software package. Clusters were visualized with Java Treeview (Figure 2B).

**Transcript identification and genome annotation**

RNA was extracted with Trizol reagent (Invitrogen, Carlsbad, CA, 15596-026), and DNase treated (Invitrogen, Carlsbad, CA, AM1907) according to manufacturer specifications. A cDNA library was generated using 100 ng of total RNA according to Illumina TruSeq RNA Sample Prep protocol (Illumina, RS-122-2001) prior to duplex-specific nuclease (DSN) treatment. 100 ng of cDNA library was incubated in hybridization buffer (50 mM HEPES, 500 mM NaCl) for 2 min at 98°C, followed by 1 hr at 68°C. Ribosomal RNA (rRNA) was not specifically depleted (He et al., 2010). Instead, we used duplex-specific nuclease (DSN) normalization to remove abundant...
RNA (rRNA, tRNA) from the total RNA sample, thereby enriching mRNA (Zhulidov et al., 2004). Samples were immediately treated with 4 units of DSN enzyme (Evrogen, Moscow, Russia EA001) in 1× DSN buffer and incubated for an additional 25 min at 68°C, prior to addition of stop solution, and purification with Ampure XP beads (Beckman Coulter, Brea, CA, A63881). RNA libraries were pooled equally according to qPCR quantitation, and sequenced using v3 chemistry on a paired-end single HiSeq2000 lane (100 × 8 × 100). Samples were demultiplexed using an 8-bp index read at the end of read 1. Total RNA was sequenced at extremely high coverage (2587 × mean coverage) so that rRNA sequences (~77% of all sequence reads) could be computationally excluded, as described by Wurtzel et al. (2010).

After quality score trimming (described earlier), sequence reads were aligned using TopHat (Trapnell et al., 2009). The RNA-seq data displayed a great deal of overlap with the predicted annotations (Hartman et al., 2010), with 92.1% of the existing annotations being confirmed. Of the 4073 predicted annotations, 3751 were confirmed, and, of these, 744 were merged with other transcripts to form longer transcripts. A heuristic approach was applied to adjust the transcript 5′ and 3′ positions of the Hartman et al. (2010) predicted annotations based on the boundaries of high RNA-seq coverage regions. This was vital as TSS accuracy is of great importance for NDR identification (Figure 5).

Because 85% of the Haloferax genome is predicted to be coding (Hartman et al., 2010), transcript detection is complicated by transcript overlap. To overcome this, computationally identified transcripts were manually curated yielding a total of 3059 expressed transcripts in Hfx. volcanii. Of these, 32 transcripts are novel (Supplementary file 1). Of these transcripts, NTRANS_0004 was the most abundant transcript in the transcriptome, excluding the six rRNA genes. Homology data was obtained using BLASTX with a BLOSUM45 matrix against the non-redundant protein sequence database (Altschul et al., 1990). Conserved domains were identified using the Conserved Domain Database (Marchler-Bauer et al., 2011). Sequence data, nucleosome and transcriptome maps and supplemental tables have been deposited to the Short Read Archive and Dryad, as indicated in the datasets statement. Additionally this data is available at http://chemogenomics.med.utoronto.ca/supplemental/chromatin/.

Figure 5. Sample screenshot of all data tracks loaded into the Savant genome browser (Fiume et al., 2010). The nucleosome sequence data is displayed, and the periodicity reflects protected and unprotected fragments after MNase digestion (magnitude of peak is not considered). Peaks represent nucleosome midpoints, which were detected and marked. Below are the corresponding RNA-seq and curated gene tracks. In this screenshot, one can observe seven entire ORFs in line with their NDRs and −1 and +1 nucleosomes.

DOI: 10.7554/eLife.00078.007
Acknowledgements
We thank H van Bakel for advice with nucleosome map and transcriptome construction.

Additional information

Funding

| Funder                          | Grant reference number | Author                  |
|---------------------------------|------------------------|-------------------------|
| Canadian Institutes for Health Research | MOP-86705              | Corey Nislow            |
| Canadian Cancer Society         | 20380                  | Guri Giaever, Corey Nislow |
| Ontario Graduate Scholarship   |                        | Ron Ammar               |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions
RA, Conception and design, acquisition of data, analysis and interpretation of data, drafting and revising the article; DT, Acquisition of data analysis and interpretation of data; KT, Prepared samples for sequencing; MG, Prepared samples for sequencing, drafting and revising the article; Conception and design, acquisition of data analysis and interpretation of data, drafting and revising the article; TD, Prepared samples for sequencing; performed the sequencing; GB, Conception and design, analysis and interpretation of data, drafting and revising the article; GG, Interpretation of data, drafting and revising the article; CN, Directed the research, prepared biological material, conception and design, acquisition of data, analysis and interpretation of data, drafting and revising the article.

Additional files

Supplementary files
• Supplementary file 1. A table describing the 32 novel transcripts identified in *Hfx. volcanii*. DOI: 10.7554/eLife.00078.008

Major datasets
The following datasets were generated

| Author(s)                       | Year | Dataset title                        | Dataset ID and/or URL                                      | Database, license, and accessibility information |
|---------------------------------|------|--------------------------------------|-----------------------------------------------------------|---------------------------------------------------|
| Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader G, Giaever G, Nislow C | 2012 | Nucleosome data (MNase-seq)          | SRX188663; http://www.ncbi.nlm.nih.gov/sra/SRX188663     | Publicly available at SRA (http://www.ncbi.nlm.nih.gov/sra). Reporting Standards: N/A |
| Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader G, Giaever G, Nislow C | 2012 | undigested DNA control               | SRX188665; http://www.ncbi.nlm.nih.gov/sra/SRX188665     | Publicly available at SRA (http://www.ncbi.nlm.nih.gov/sra). Reporting Standards: N/A |
| Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader G, Giaever G, Nislow C | 2012 | naked DNA control (MNase-seq)        | SRX185902; http://www.ncbi.nlm.nih.gov/sra/SRX185902     | Publicly available at SRA (http://www.ncbi.nlm.nih.gov/sra). Reporting Standards: N/A |
RNA-seq

Chromatin is an ancient innovation conserved between Archaea and Eukarya

References

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, et al. 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature 446:572–6.

Altman-Price N, Mewarech M. 2009. Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon Haloferax volcanii. J Bacteriol 191:1610–7.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–10.

Bailey KA, Pereira SL, Widom J, Reeve JN. 2000. Archaeal histone selection of nucleosome positioning sequences and the procarcyotic origin of histone-dependent genome evolution. J Mol Biol 303:25–34.

Breuer S, Allers T, Spohn G, Soppa J. 2006. Regulated polyploidy in halophilic archaea. PloS One 1:e92.

Chang GS, Noegel AA, Mavrich TN, Muller R, Tomsho LP, Ward E, et al. 2012. Unusual combinatorial involvement of poly-A/T tracts in organizing genes and chromatin in Dictyostelium. Genome Res 22:1098–106.

Chung HR, Dunkel I, Heise F, Linke C, Krobitsch S, Ehrenhofer-Murray AE, et al. 2010. The effect of micrococcal nuclease digestion on nucleosome positioning data. PloS One 5:e15754.

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, et al. 2006. A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA 103:5320–5.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–8.

Du X, Takagi H. 2007. N-Acetyltransferase Mpr1 confers ethanol tolerance on Saccharomyces cerevisiae by reducing reactive oxygen species. Appl Microbiol Biotechnol 75:1343–51.

Fiume M, Williams V, Brook A, Bruudno M. 2010. Savant: genome browser for high-throughput sequencing data. Bioinformatics 26:1938–44.

Forbes AJ, Patrie SM, Taylor GK, Kim YB, Jiang L, Kelleher NL. 2004. Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. Proc Natl Acad Sci USA 101:2678–83.

Geer LY, Domrachev M, Lipman DJ, Bryant SH. 2002. CDART: protein homology by domain architecture. Genome Res 12:1619–23.

Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, Madupu R, et al. 2010. The complete genome sequence of Haloferax volcanii DS2, a model archaeon. PloS One 5:e9605.

He S, Wurtzel O, Singh K, Froula JL, Yilmaz S, Tringe SG, et al. 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat Methods 7:807–12.

Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10:161–72.

Kaminska KH, Bujnicki JM. 2008. Bacteriophage Mu Mom protein responsible for DNA modification is a new member of the acetyltransferase superfamily. Cell Cycle 7:120–1.

Kaplan N, Moore IK, Fondeuf-Mittendorf Y, Gossett AJ, Tillo D, Field Y, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458:362–6.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–9.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR & Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet 39:1235–44.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Deweese-Scott C, et al. 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res 39:D225–9.

Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–303.

Mullakhpanbhai MF, Larsen H. 1975. Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. Arch Microbiol 104:207–14.

Palmer JR, Daniels CJ. 1995. In vivo definition of an archaeal promoter. J Bacteriol 177:1844–9.

Pereira SL, Reeve JN. 1998. Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. Extremophiles 2:141–8.

Pereira SL, Grayling RA, Lux R, Reeve JN. 1997. Archaeal nucleosomes. Proc Natl Acad Sci USA 94:12633–7.
Puerta C, Hernandez F, Gutierrez C, Pineiro M, Lopez-Alarcon L, Palacian E. 1993. Efficient transcription of a DNA template associated with histone (H3.H4)2 tetramers. J Biol Chem 268:26663–7.
Rizzo JM, Mieczkowski PA, Buck MJ. 2011. Tup1 stabilizes promoter nucleosome positioning and occupancy at transcriptionally plastic genes. Nucleic Acids Res 39:8803–19.
Sajan SA, Hawkins RD. 2012. Methods for identifying higher-order chromatin structure. Annu Rev Genomics Hum Genet 13:59–82.
Sandman K, Reeve JN. 2006. Archaeal histones and the origin of the histone fold. Curr Opin Microbiol 9:520–5.
Sartorius-Neef S, Pfeifer F. 2004. In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon Halobacterium salinarum. Mol Microbiol 51:579–88.
Satchwell SC, Drew HR, Travers AA. 1986. Sequence periodicities in chicken nucleosome core DNA. J Mol Biol 191:659–75.
Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol 6:e65.
Smith SW. 1997. The scientist and engineer’s guide to digital signal processing. San Diego: California Technical Pub.
Talbert PB, Henikoff S. 2010. Histone variants—ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 11:264–75.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–11.
Tsui K, Durbic T, Gebbia M, Nislow C. 2012. Genomic approaches for determining nucleosome occupancy in yeast. Methods Mol Biol 833:389–411.
Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. Nature 474:516–20.
Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. 2010. A single-base resolution map of an archaeal transcriptome. Genome Res 20:133–41.
Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner GL, Khaspekov GL, Kozhemyako VB, et al. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res 32:e37.