Old cogs, new tricks: the evolution of gene expression in a chromatin context

Paul B. Talbert, Michael P. Meers and Steven Henikoff *

Abstract | Sophisticated gene-regulatory mechanisms probably evolved in prokaryotes billions of years before the emergence of modern eukaryotes, which inherited the same basic enzymatic machineries. However, the epigenomic landscapes of eukaryotes are dominated by nucleosomes, which have acquired roles in genome packaging, mitotic condensation and silencing parasitic genomic elements. Although the molecular mechanisms by which nucleosomes are displaced and modified have been described, just how transcription factors, histone variants and modifications and chromatin regulators act on nucleosomes to regulate transcription is the subject of considerable ongoing study. We explore the extent to which these transcriptional regulatory components function in the context of the evolutionarily ancient role of chromatin as a barrier to processes acting on DNA and how chromatin proteins have diversified to carry out evolutionarily recent functions that accompanied the emergence of differentiation and development in multicellular eukaryotes.

Histone fold domain
(HFD). A protein dimerization domain of three helices separated by two loops that is characteristic of archaeal and core eukaryotic histones, TATA-binding protein-associated factors and some other proteins.

Four-helix bundles
Structures formed by two helices of each of two histones that enable dimers to assemble into more complex structures.

Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

**e-mail: steveh@fhcrc.org*

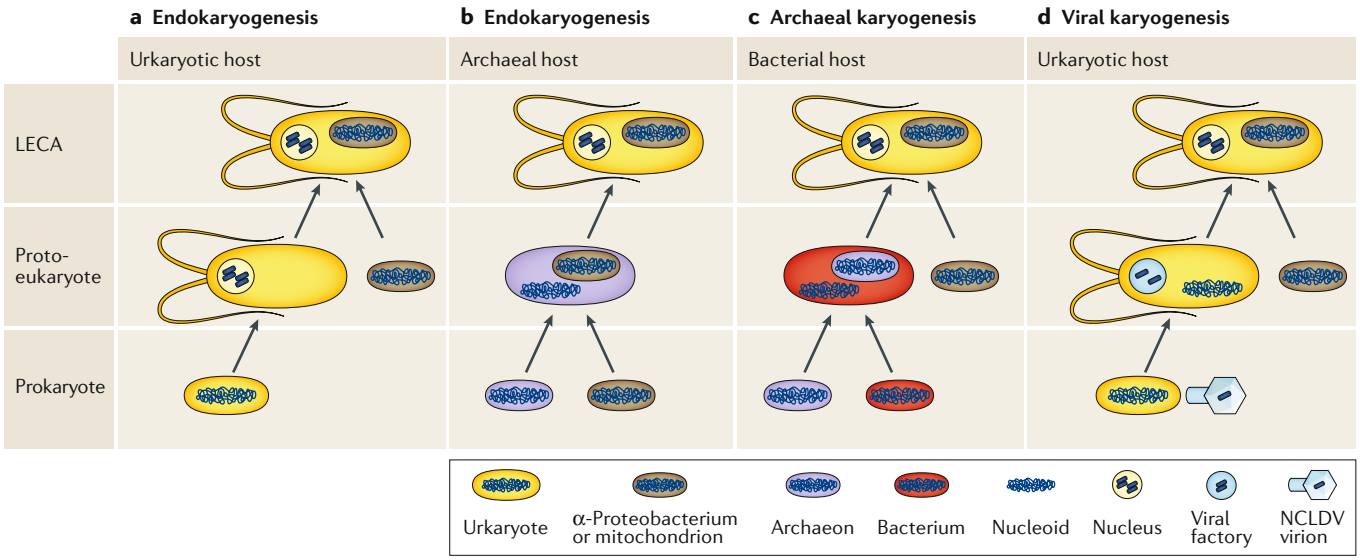
https://doi.org/10.1038/s41576-019-0105-7

An ancestral feature of eukaryotic genomes is the presence of nucleosomes that wrap DNA around an octamer of two copies of each of the four core histones H2A, H2B, H3 and H4. Nucleosomes constrain negative DNA supercoils and limit DNA accessibility, necessitating nucleosome mobilization to carry out gene regulation[1]. The core histones are each characterized by a histone fold domain (HFD) that enables them to dimerize in specific antiparallel pairs, H3 with H4 and H2A with H2B, that can further assemble by forming four-helix bundles between dimers, leading to a central H3–H4 tetramer flanked by two H2A–H2B dimers[2]. In addition to HFDs, core histones have unstructured tails that are subject to numerous post-translational modifications (PTMs) with important roles in gene regulation. Histone variants, especially of H2A and H3, may replace the corresponding core histone to form nucleosomes with distinct properties[3].

Fossil stromatolites, interpreted as microbial mats and fossil fibres from hydrothermal vents, suggest that bacterial cells with early forms of gene expression were extant ~3.7 billion years ago[4–6]. By contrast, steranes — probably of eukaryotic (or proto-eukaryotic) origin — date from 2.5–2.7 billion years ago[7], and large ornamented fossil cells confidently related to an extant eukaryote group (red algae) date from 1.2–1.6 billion years ago[8,9]. Although putatively early-diverging eukaryotes such as metamonads and kinetoplastids may have separated from other eukaryotes before the appearance of algae[10], the fossil record suggests a 1–2 billion year gap between the early origin of prokaryotic transcription factor (TF)-based gene expression systems and the nucleosome-based regulation of modern eukaryotic genes.

This gap raises the question of how the acquisition of eukaryotic nucleosomes fundamentally altered gene-regulatory processes. Except for octameric nucleosomes, nearly all enzymatic components of the eukaryotic chromatin landscape, including TFs, polymerases, topoisomerases, acetyltransferases, deacetylases, SET domain methyltransferases and even homologues of ATP-dependent chromatin remodellers, are present in prokaryotes. The traditional view is that nucleosomes act as repressors of gene expression[1] and that certain activating PTMs of histones, histone variants, and chromatin remodellers promote gene expression. To the extent that they do this, what are their mechanisms of action? How did they acquire these roles in an emerging nucleosome landscape? The last eukaryotic common ancestor (LECA) was a complex nucleated cell with an endomembrane system, cytoskeleton, mitochondrion, and linear chromosomes that underwent mitosis and meiosis[11]. The origin of eukaryotic cells is controversial, and the first eukaryotic common ancestor (FECA) has been variously proposed to be a hypothetical cell equally ancient as, but independent of, bacteria and archaea (urkaryote), a bacterium with an archaeal endosymbiont, or an archaeon with a bacterial endosymbiont (BOX 1); however, in any scenario,

Box 1 | **Models of eukaryogenesis**



The origin of the eukaryotic cell (eukaryogenesis) is controversial, particularly the origin of the nucleus (karyogenesis). The chimeric nature of eukaryotes became clear when the endosymbiotic origin of mitochondria and chloroplasts from an α-proteobacterium and cyanobacterium, respectively, became generally accepted[158,159]. Around the same time, Woese and colleagues[160,161] discovered that archaea were a distinct domain of life separate from bacteria and eukaryotes, with particular signatures in tRNA and ribosomal RNA (rRNA) molecules and lipid membranes distinct from those of bacteria and eukaryotes. Woese argued that these three domains were anciently diverged and referred to the ancestral lineage of modern eukaryotes as urkaryotes (three-domains hypothesis; see the figure, part **a**). Directional models of evolution applied to protein domain data sets of all three cellular domains support the primary divergence of the eukaryotic lineage[162].

The close relationship between informational molecules of archaea and eukaryotes has led to a popular class of models in which the ancestral host cell for the mitochondrion was not an urkaryote but an archaeal cell (two-domains hypothesis; see the figure, part **b**), most recently proposed to be among the Asgardarcheota[163], although this remains disputed[162,164]. This scenario requires an unprecedented replacement of the cell membrane and virome of the archaeal host[13]. Alternative models propose that the mitochondrial host was a chimeric cell resulting from an endosymbiotic archaeon in a bacterial cell[165,166] (see the figure, part **c**). Regardless of the nature of the host cell, the eukaryotic genome encodes as many protein fold superfamilies specifically shared with viral genomes as with archaea[13]. Models of viral eukaryogenesis have attributed the origin of the nucleus to virus factories of nucleocytoplasmic large DNA viruses (NCLDVs) (see the figure, part **d**) or proposed its evolution as a protection from viral infections[60]. Large viruses have contributed numerous genes to eukaryotes[167], possibly including the mRNA-capping enzyme, certain DNA and RNA polymerases, topoisomerase IIA[58,60], and even core histones[31].

LECA, last eukaryotic common ancestor.

the complexity of LECA suggests dramatic evolution in a proto-eukaryote stage between FECA and LECA, which already possessed most features of modern chromatin regulation[12]. Thus, there are few clues to the intermediate steps in the remarkable transition to nucleosome-based gene regulation. The chimeric genome of LECA had already assembled the component genes of chromatin regulation from homologues in bacteria, archaea and viruses and invented many new genes[13]. Despite the dearth of information on this transition, we find it of interest to place the available data in an evolutionary context.

A fundamental role of nucleosomes in the evolution of eukaryotic genomes has been as a nonspecific, passive barrier to DNA-templated processes. We propose that the evolution of chromatin proteins from prokaryotic homologues has been shaped by their increasing participation in the dynamic processes whereby nucleosome assembly competes with TF binding and transcription by RNA polymerase II (RNAPII) or other RNAPs, both of which require transient unwrapping of nucleosomes. It is likely that the acquisition of features that enable or reduce nucleosome mobilization by some histone modifiers facilitated the evolution of genome complexity in multicellular eukaryotes and established a generalized modular framework for complex cell and tissue differentiation regimes that continues to evolve today.

We focus here on histones and non-histone chromatin proteins that mobilize nucleosomes for transcription initiation by RNAPII, including topoisomerases, TFs, remodellers, histone acetyltransferases (HATs), histone deacetylases (HDACs), the histone variant H2A.Z, and PTMs associated with initiation. We also discuss Polycomb group negative regulators of transcription. We do not discuss in detail PTMs of transcriptional elongation or constitutive heterochromatin, as excellent reviews exist on the descriptive and mechanistic aspects of these topics[14–16]. Also outside of our scope are the regulatory roles of long non-coding RNAs, as our limited understanding of the evolutionary history of these RNAs is discussed elsewhere[17].

Transcriptional machinery

In the bacterium *Escherichia coli*, regulation of ~4,500 genes is achieved by seven different sigma factors that recruit RNAP to different sets of genes. Transcription is further modulated by 300 TFs that bind between

Metamonads
Anaerobic cells typically with two pairs of basal bodies with one posterior and three anterior flagella. Metamonads include the diplomonad *Giardia intestinalis* and the parabasalid *Trichomonas vaginalis*, among others, and may represent one of the earliest branches of the eukaryotic phylogenetic tree.

Kinetoplastids
Flagellates that have a dense mass of DNA called a kinetoplast, which contains many copies of the mitochondrial genome. Kinetoplastids include Bodonids and trypanosomes and are thought to represent an early branch of the eukaryotic phylogenetic tree.

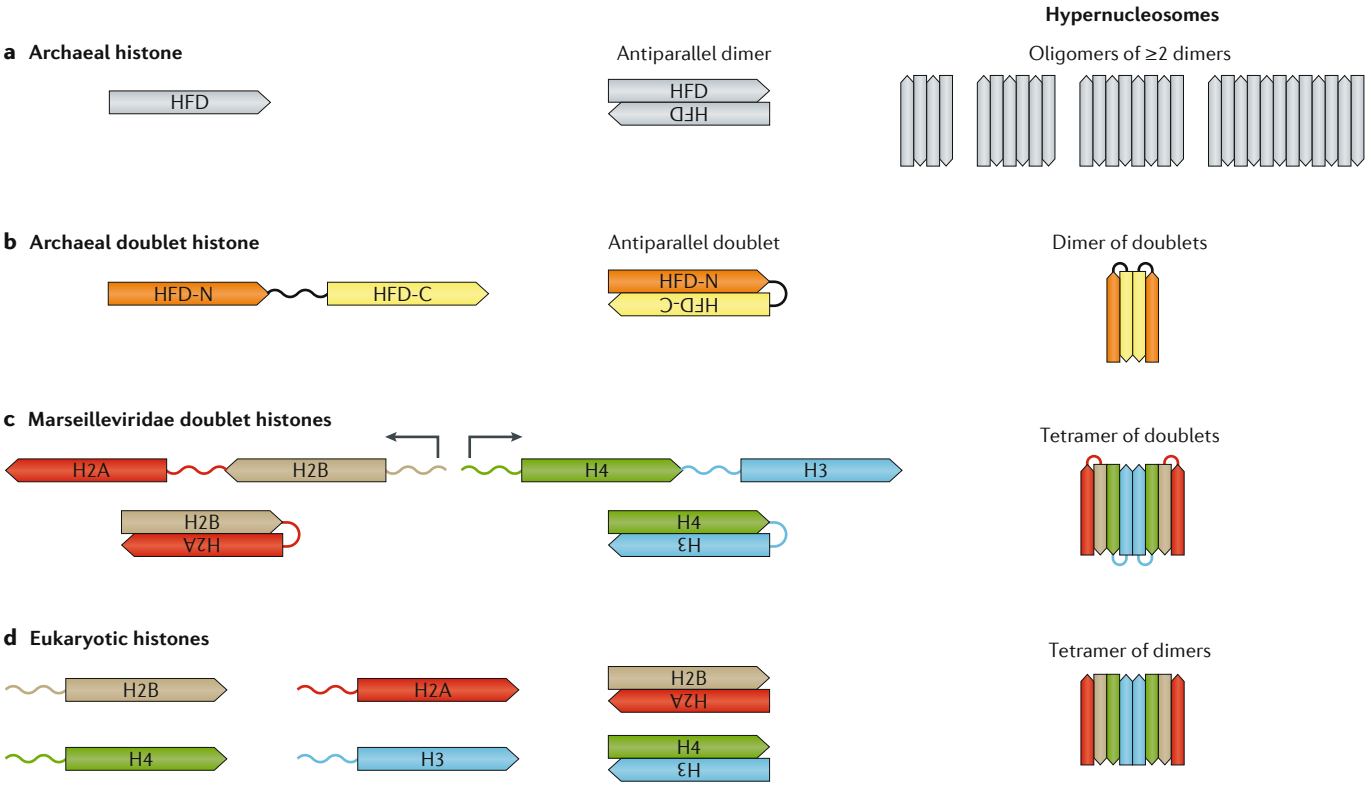


Fig. 1 | **Histones and nucleosomes. a** | Archaeal histones mostly lack tails and form dimers that can assemble into hypernucleosomes with four or more histone fold domains (HFDs). **b** | Archaeal histone doublets force divergent amino-terminal and carboxy-terminal HFDs to pair and then dimerize to form a four-HFD particle. **c** | Marseilleviridae histones are divergently transcribed as doublets that probably assemble into an eight-HFD tetrameric nucleosome. **d** | Eukaryotic histones form octamers with two copies each of H2A, H2B, H3 and H4.

one and hundreds of genes each, together with RNAP components, and stabilize or occlude RNAP binding (reviewed elsewhere[18]). By contrast, eukaryotes are often considered to be the sisters or descendants of archaea, because eukaryotic RNAPII and its associated general transcription factors TATA-binding protein (TBP), TFIIB and TFIIE — as well as RNAPI, RNAPIII and their own paralogous general transcription factors — all have homologues in archaea. Both archaea and eukaryotes assemble a pre-initiation complex (PIC) in which TBP binds to a TATA box, bending DNA, and is stabilized by TFIIB or its homologues, which together recruit the RNAP[19]. TFIIE and its homologues may open or stabilize the transcription bubble[20]. TBP is highly conserved in eukaryotes, but archaea may have several paralogues of TBP and the TFIIB homologue TFB, although to what extent these bind to different sequences or TFs is not clear[19]. Archaea lack homologues of the eukaryotic factors TFIIA, TFIIF, TFIIH, the TBP-associated factor (TAF) proteins that together with TBP form TFIID, and the mediator complex, as well as polymerase-specific factors for RNAPI and RNAPIII[19,21]. Some TAF proteins make additional DNA contacts that facilitate TBP binding to promoters that lack a TATA box[19], and TAF1 inhibits the TBP-specific remodeller MOT1 from removing TBP from these lower-affinity sites[22]. A recent analysis proposes that eukaryotic RNAPIII is the direct homologue of archaeal RNAP and was transferred from an urkaryote or proto-eukaryote to the ancestor of nucleocytoplasmic large DNA viruses (NCLDVs), where it diversified along with them. Subsequently, RNAPII and the large subunit of RNAPI were acquired from different NCLDV lineages by a proto-eukaryote before LECA[23].

Origin of eukaryotic nucleosomes

LECA packaged its genome in nucleosomes similar to those of modern eukaryotes, with PTMs of histone tails and variant histones such as cenH3 in centromeres[3]. HFD proteins are found in all three cellular domains, although the functions and DNA-binding abilities of some HFD families in bacteria and archaea are unknown[24]. Many archaea have one or more histones of the HMfB family that bind to DNA and are more similar to eukaryotic histones, although none is specifically related in sequence to eukaryotic histones, despite close structural alignment of the HFDs. This suggests either a sister or ancestral relationship between eukaryotic and archaeal histones, depending on the model of eukaryogenesis (BOX 1). Archaeal histones mostly lack the unstructured tails of eukaryotic histones, but recently some have been found to have tails containing lysine, raising the possibility that they may be modified[25]. Archaeal histones fold together in dimers, usually homomeric, which can be further polymerized through four-helix bundles to form tetrameric HFD particles that wrap 60 bp of DNA or more extended DNA-wrapping polymers[26] termed hypernucleosomes[25]

SET domain
A protein domain generally associated with protein lysine methyltransferase activity.

Last eukaryotic common ancestor
(LECA). The most recent cell that was ancestral to all extant eukaryotes, inferred to be a complex nucleated cell with an endomembrane system, a cytoskeleton, a mitochondrion, and linear chromosomes that were packaged in nucleosomes and underwent mitosis and meiosis.

First eukaryotic common ancestor
(FECA). A hypothetical first cell that was ancestral to all eukaryotes and distinct from bacteria and archaea, variously conceived as an urkaryote or a chimeric organism composed of host and endosymbiont.

Urkaryote
A hypothetical cell ancestral to the eukaryotic cytoplasm with distinct features of the translational apparatus that distinguish it from archaea and bacteria and that arose independently from the last universal cellular ancestor.

Endosymbiont
A cell or organism living inside another cell or organism.

Proto-eukaryote
A cell in any stage of eukaryotic evolution between the first eukaryotic common ancestor to the last eukaryotic common ancestor.

General transcription factors
The factors that together with RNA polymerase II make up the pre-initiation complex for transcription.

Nucleocytoplasmic large DNA viruses
(NCLDVs). A diverse monophyletic group of fairly large, complex DNA viruses, including giant viruses such as the Marseilleviridae as well as other more modestly sized viruses.

CenH3
Histone H3 variant specific to centromeric nucleosomes, examples of which include mammalian CENP-A and budding yeast Cse4.

Eukaryogenesis
The process of evolving from the first eukaryotic common ancestor to the last eukaryotic common ancestor.

Hypernucleosomes
Archaeal histone polymers of variable length that wrap DNA.

Monophyletic group
A group of genes, proteins or organisms that includes all the descendants of a single common ancestor and excludes others.

Capsid
A proteinaceous structure that encloses a viral genome for protection and dispersion.

Winged helix domain
A protein domain with combined specific and nonspecific DNA-binding affinity characterized by a helix–turn–helix motif flanked by β-sheets on one or both sides.

(FIG. 1a). In vitro, archaeal histones can repress transcription initiation and slow elongation, but promoter occupancy of histones can be outcompeted by a TF[27,28]. A few archaeal histones have tandemly linked HFD doublets that are constrained to fold together, enabling divergence between the two HFDs in the same protein (FIG. 1b). In *Haloferax volcanii*, such doublets can then dimerize to form a structure with four HFDs that shows regular spacing on the chromosome, including depletion of such particles at promoters, similar to eukaryotic nucleosomes[29]. Such histone doublets were proposed to be intermediates in the evolution of nucleosomes[30], facilitating the diversification of the four families of core histones in specific pairs: H2A with H2B and H3 with H4.

The giant viruses of the family Marseilleviridae within the NCLDVs encode HFD doublets that form obligate heterodimers, and these HFDs have unstructured tails and are orthologous to the eukaryotic core histone pairs, with an H2B–H2A-like doublet and a H4–H3-like doublet[31] (FIG. 1c). Each class of Marseilleviridae HFD orthologues of the four core histone classes forms a monophyletic group that is a sister to all eukaryotic histones of the corresponding class, branching at the base of the eukaryotic tree before variants within a class, such as H2A and its variant H2A.Z or H3 and its variant cenH3. The doublet structures make unlikely the scenario that all four core histone classes were separately acquired from modern eukaryotes (FIG. 1d) and then rapidly diverged and reorganized into doublets in Marseilleviridae. It is more parsimonious to suggest that a common ancestor of viral and eukaryotic histones existed before the diversification of histone variants. Marseilleviridae histones may have been acquired from a proto-eukaryote to protect the viral genome from host nucleases and may preserve the predicted doublet intermediate stage of nucleosome evolution. Alternatively, doublet histones may have been acquired from an archaeon and subsequently specialized to package large viral genomes into the constrained space of a capsid. The viral histones could then have been acquired by a proto-eukaryote. Acquisition of a major chromatin-packaging protein family from NCLDVs has precedent in dinoflagellates, which do not package the bulk of their large condensed chromosomes in nucleosomes but nevertheless encode many divergent histones that may be present at transcribed genes[32]. DNA-packaging proteins donated by the NCLDV family Phycodnaviridae to the dinoflagellate ancestor[33] possibly drove histone depletion[34].

All eukaryotes encode homologues of the core histone families H2A, H2B, H3 and H4; however, the linker histone H1, which does not have an HFD, has an independent origin and may not have been present in LECA. Homologues of the carboxy-terminal portion of animal H1 are found in bacteria, but not archaea, and are present in most protists except the early diverging metamonads[35]. The winged helix domain of H1 found in animals and plants, however, is absent in several protist groups, suggesting that this domain may have been independently acquired in the animal, fungal and plant lineages[36] (TABLE 1).

Whether eukaryotic nucleosomes derive from an urkaryote, archaeal symbiont, or a giant virus, their packaging function is required for mitotic chromosome condensation[37,38] and may have facilitated the expanded size of eukaryotic genomes. However, they may have been initially selected to restrict access to the genome by viruses and transposons[39]. Nucleosomes are barriers to access of genomic DNA by DNA transposons, which prefer to insert into exposed DNA[40]. Although DNA methylation was absent in LECA and is not found in early-diverging eukaryotes, later-diverging eukaryotes repeatedly adapted bacterial DNA methylation enzymes (TABLE 1), used in prokaryotes to discriminate host DNA from foreign DNA, to silence transposons[41,42]. However, because nucleosomes are also barriers to DNA methylation, chromatin remodellers are necessary to methylate DNA in nucleosomes, and histone modifications help regulate where methylation occurs (BOX 2).

Supercoiling and nucleosomes

Perhaps the most basic consequence of wrapping DNA around nucleosomes is that it constrains negative DNA supercoils throughout the eukaryotic genome, altering the roles of DNA superhelicity or torsion in gene regulation and genome organization. In *E. coli*, TFs are unknown for the majority of genes, and, instead, superhelicity has a major role in transcriptional regulation[43]. Topoisomerases regulate superhelicity by relaxing or adding supercoils that can form when DNA is constrained in a circle or by anchoring proteins or cellular structures (reviewed elsewhere[44]). Type I topoisomerases make transient single-stranded breaks in DNA, and type II topoisomerases make transient double-stranded breaks, passing DNA through the break before resealing it with an altered superhelicity. Supercoils are generated by polymerases as they unwind helical DNA for transcription or replication (FIG. 2), and topoisomerases relax the (overtwisted) positive supercoils generated ahead of RNAP during transcription and the (undertwisted) negative supercoils formed in its wake[44]. Although the positive and negative supercoils would be expected to cancel out any net change to the superhelicity of the DNA, nucleoid-associated DNA-binding proteins can constrain the negative supercoils and direct their energy towards promoters, whereas no abundant DNA-binding proteins constrain positive supercoils, resulting in a net negative superhelicity that helps to unwind DNA for transcription or replication[43]. In highly transcribed operons loaded with multiple polymerases, the negative supercoils behind a leading polymerase cancel the positive supercoils ahead of a following polymerase, facilitating multiple rounds of transcription. Transcription-coupled supercoils can quickly propagate to nearby genes to create local supercoiling domains that coordinate expression from neighbouring genes in temporal expression patterns[45]. In *Caulobacter crescentus* swarmer cells, interactions between nearby loci form chromosomal interaction domains that are dependent on transcriptional elongation and on DNA gyrase (topoisomerase type IIA family), which can introduce negative supercoils. The interaction domains probably arise from supercoiled plectonemes (twisted loops) separated by more open, highly expressed genes[46].

Whereas fluid superhelicity is a major factor in bacterial and archaeal transcription regulation, eukaryotic

Table 1 | **Phylogenomic distribution of chromatin proteins**

Protein families or subfamilies	Bacteria	Archaea	NCLDVs	Metamonads	Kinetoplastids	Plants	Fungi	Animals
Histones H2A, H2B, H3, H4	–	(+)	+	+	+	+	+	+
H2A.Z	–	–	–	–	+	Multiple	+	1 or 2
H2A.Z-specific H2B	–	–	–	–	H2B.V	–	–	–
H2A.W	–	–	–	–	–	Seed plants	–	–
Short H2As	–	–	–	–	–	–	–	Placental mammals
H1	H1-cterm	–	–	–	H1-cterm	H1 (WH)	H1 (WH)	H1 (WH)
RNA pol	+	+	+	RNAPI, RNAPII*, RNAPIII	RNAPI, RNAPII, RNAPIII	RNAPI, RNAPII, RNAPIII	RNAPI, RNAPII, RNAPIII	RNAPI, RNAPII, RNAPIII
TBP	–	+	Some	+	+	+	+	+
TFIIA	–	–	–	–	+	+	+	+
Mediator	–	–	–	+	+	+	+	+
Capping enzyme	–	–	+	+	+	+	+	+
Topo type IA	Topo IA, Topo III, reverse gyrase	Topo III, reverse gyrase	–	Topo III	Topo III	Topo III	Topo III	Topo III
Topo type IB	+	Thaum	Pox	Topo I	Topo I	Topo I	Topo I	Topo I
Topo type IIA	Gyrase & Topo IV	Gyrase (some)	Topo II	Topo II	Topo II	Topo II	Topo II	Topo II
Topo type IIB	Some	Topo VI	–	Spo11	Spo11	Spo11	Spo11	Spo11
SWI/SNF ATPases	(+)	(+)	+	6	13+	19+	19+	19+
HAT families	(+)	(+)	–	4	6+	6+	6+	6+
HDAC families	(+)	(+)	–	2	7+	7+	7+	7+
SET domain methylases	(+)	(+)	–	5	8+	9+	9+	9+
Histone demethylases	(+)	–	–	–	8	12+	12+	12+
DNA methylases	(+)	(+)	–	–	Dnmt6	Dnmt1, Dnmt3, Dnmt5, CMT	Dnmt1, Dnmt4, Dnmt5, Dim2	Dnmt1, Dnmt3

Presence of proteins is indicated with + or a number representing the minimum number of protein families or subfamilies. (+) indicates homologous proteins in archaea and bacteria that do not form or act on octameric nucleosomes. H1 proteins may be homologous only to the carboxyl terminus of human H1 proteins (H1-cterm) or additionally have a winged helix (WH) domain. RNAPII* indicates that, in metamonads, RNAPII lacks the characteristic heptad repeat at the carboxyl terminus. Spo11 is the recombination double-strand break enzyme of eukaryotes that lacks topoisomerase activity. CMT, chromomethylase; Dnmt, DNA methyltransferase; HAT, histone acetyltransferase; HDAC, histone deacetylase; NCLDVs, nucleocytoplasmic large DNA viruses; Pox, poxviruses; RNA Pol, RNA polymerase; TBP, TATA-binding protein; Thaum, Thaumarchaeota; Topo, topoisomerase. Eukaryotic tree simplified from REF.[12]. Data compiled from REFS[12,21,31,35,36,58,59,115,168,185,196,200].

nucleosomes restrict the free propagation of supercoils. Changes in superhelicity are largely accommodated by topoisomerases acting at the linker regions between nucleosomes and at nucleosome-free promoters[47]. The linker histone H1 can bind to positively or negatively crossed entry or exit DNA of nucleosomes[48] and may help to control the torsional state of the chromatin. Positive torsional stress induced by RNAPII progression promotes H2A–H2B dimer loss[49,50] and can displace nucleosomes downstream of elongating RNAPII in *Drosophila melanogaster* S2 cells[51], releasing the constrained negative supercoiling energy that was stored in the evicted nucleosomes to counteract positive torsion while eliminating the nucleosome barrier that otherwise induces polymerase backtracking and arrest[52]. In *D. melanogaster*, topoisomerase I is primarily responsible for relaxing positive supercoils for efficient RNAPII progression[51]. In human HCT116 cells, topoisomerase I associates with paused RNAPII at the promoter and throughout the gene body during transcriptional elongation[53]. Its activity is enhanced by phosphorylation of the conserved heptapeptide repeat YSPTSPS of the RNAPII carboxy-terminal domain (CTD) on Ser2, which also regulates release from the pausing of RNAPII.

Superhelicity
The degree of torsion or supercoiling in a DNA molecule.

Plectonemes
Writhed loops resulting from supercoiling.

Box 2 | **DNA methylation and nucleosomes**

In bacteria, cytosine methylation protects DNA from host-encoded restriction endonucleases. In eukaryotes, methylation has multiple functions, and there are several families of DNA methyltransferases independently acquired from bacteria[41], which may be preferentially specialized for de novo (Dnmt3/DRM and CMT/Dim-2 families) or for maintenance methylation during replication (Dnmt1/MET1 and Dnmt5 families)[42,168,169].

Nucleosomes are barriers to methylation, and in three kingdoms of marine algae, CG methylation by Dnmt5 forms a periodic pattern in linker regions that disfavours nucleosomes, suggesting that mutual antagonisms between methylation and nucleosomes may reinforce nucleosome positions[168]. Dnmt1/MET1 methylates linkers around well-positioned nucleosomes in the absence of linker histone H1 and of the chromatin remodellers DDM1 in *Arabidopsis thaliana* and LSH in mouse, which enable methylation of nucleosomes[170]. Loss of DDM1 or methyltransferases results in activation of transposons[171]. Transcriptional silencing of transposons is a key function of DNA methylation in plants, fungi and vertebrates[42], and methylation also silences gene promoters during mammalian development and differentiation and establishes genomic imprinting (see recent reviews[169,172]). Unusually, DNA methylation of a promoter element regulating the *A. thaliana* REPRESSOR OF SILENCING 1 (*ROS1*; also known as *DML1*) DNA demethylase gene upregulates the promoter, resulting in reduced global DNA methylation, forming a genomic rheostat that maintains steady-state levels of DNA methylation genome-wide[173].

Methylation is also found in gene bodies in plants and animals[42]. In humans and other species, DNMT3A and DNMT3B bind to H3K36me3, which is associated with transcriptional elongation[174]. They are inhibited by H3K4me3, which is associated with transcription initiation[175]. In mouse embryonic stem cells (ESCs), H3K36me3 recruits DNMT3B to methylate gene bodies, which suppresses spurious transcription initiation within genes[176]. In *A. thaliana*, methylation excludes H2A.Z from gene bodies, which may stabilize the expression of constitutive genes[177]. In zebrafish, 'placeholder' nucleosomes with H2A.Z and H3K4me1 deter methylation of promoters of housekeeping and developmental genes during the germline-to-embryo transition[178].

This suggests that topoisomerase I has low activity at paused RNAPII and fails to counteract positive supercoils that contribute to RNAPII arrest. However, when released from pausing by CTD phosphorylation, enhanced topoisomerase I activity relaxes positive supercoils efficiently to enable elongation through the gene body.

As in bacteria, the generation of negative supercoils upstream of eukaryotic promoters can have regulatory consequences but with new 'twists'. Underwinding DNA can favour the formation of non-B-form DNA structures that exclude nucleosomes[54] and enable binding of specific regulatory factors[55]. Sequence motifs associated with such non-B structures are largely absent from prokaryotic genomes[54]. In yeast, the spread of transcription-coupled supercoils appears to co-regulate clusters of seven or more contiguous genes[56]. In human cells, genome-wide mapping of supercoils revealed variably sized supercoiling domains of median size 100 kb (REF. [57]), reminiscent of the chromatin interaction domains of *C. crescentus*. Underwound domains are associated with transcription, DNase I hypersensitive sites, and topoisomerase I and are depleted for topoisomerase II. They depend on both transcription and topoisomerase activity, have a more decondensed structure than overwound domains, and become compacted if transcription is inhibited[57]. Transcription of short initiating RNAs is sufficient to decompact the domains when elongation is prevented[57]. Supercoiling domains may bring enhancers and promoters together in proximity. Non-coding RNAs transcribed from enhancers, known as enhancer RNAs (eRNAs), have been proposed to promote a superhelicity that is favourable for transcription of the promoters on which they act[55].

The different superhelical environment imposed by nucleosomes possibly favoured different topoisomerases than those in bacteria and archaea. Eukaryotic type II topoisomerases (type IIA) are not related to the type II topoisomerases (type IIB) used by most archaea[58] but instead branch as sisters to the topoisomerases of the Marseilleviridae[31]. Indeed, topoisomerase IIA was present in the ancestor of all NCLDVs that predated LECA[23,59]. The relationship of giant viruses to eukaryotes is controversial[60], but most topoisomerases may have originated in viruses and been transferred to the three cellular domains after their divergence[58] (TABLE 1). Interestingly, topoisomerase II is found along the scaffold of mitotic chromatids[61] and is required for decatenation of sister chromatids. This activity is driven by positive supercoiling that is introduced by the combined action of condensin with topoisomerase II in the presence of mitotic spindles[62]. Topoisomerase II is found with cohesin and CTCF sites at the boundaries of topologically associating domains (TADs), where it may help regulate the superhelicity of TADs[63]. A model of chromatin condensation proposes that as transcription ceases in metaphase, topoisomerase II and condensin compact DNA into condensed, positively supercoiled domains that form the structure of mitotic chromosomes[64]. It seems possible that the role of nucleosomes working with condensin and topoisomerase II to compact and decatenate mitotic chromosomes[37,38] may have been a key factor that facilitated the expansion of eukaryotic genomes.

TFs and remodellers in nucleosome depletion

Bacterial and archaeal TFs share a common origin, with ~53% of archaeal TFs having at least one bacterial homologue[65]. By contrast, only ~2% of archaeal TFs have homologues uniquely in eukaryotes, with another 6% having homologues in both eukaryotes and bacteria. Among the latter are 6–11 families with helix–turn–helix (HTH) domains, which are abundant in prokaryotes[66]. HTH families in eukaryotes, such as the MYB and homeodomain families, are only distantly related to those in prokaryotes, and some eukaryotic HTH proteins may derive from transposons. In addition to HTH proteins, a profusion of ~55 distinct eukaryotic-specific DNA-binding domain families were present in LECA, such as HMG1, AT-hook, C_2H_2 zinc-fingers, and MADS-box proteins[12]. This diversification might be due, in part, to the expansion of genomes and gene regulatory targets made possible by nucleosomal packaging and perhaps also due to a reduced regulatory role of superhelicity in genomes largely constrained by nucleosomes. Lineage-specific amplifications of TF families are widespread in eukaryotes[12] and may underlie eukaryotic morphological diversity.

Diversification of TFs was accompanied by specialization and diversification of the Snf2 family of proteins that use the energy of ATP to slide or evict nucleosomes or other proteins to make TF-binding sites available. The Snf2 family of ATP-dependent remodellers is divided into four major subfamilies, SWI/SNF, ISWI/SNF2L,

Enhancer RNAs
(eRNAs). Non-coding RNAs transcribed from enhancers.

Topologically associating domains
(TADs). Regions of the genome that interact with themselves in 3D nuclear space more often than regions outside the TAD.

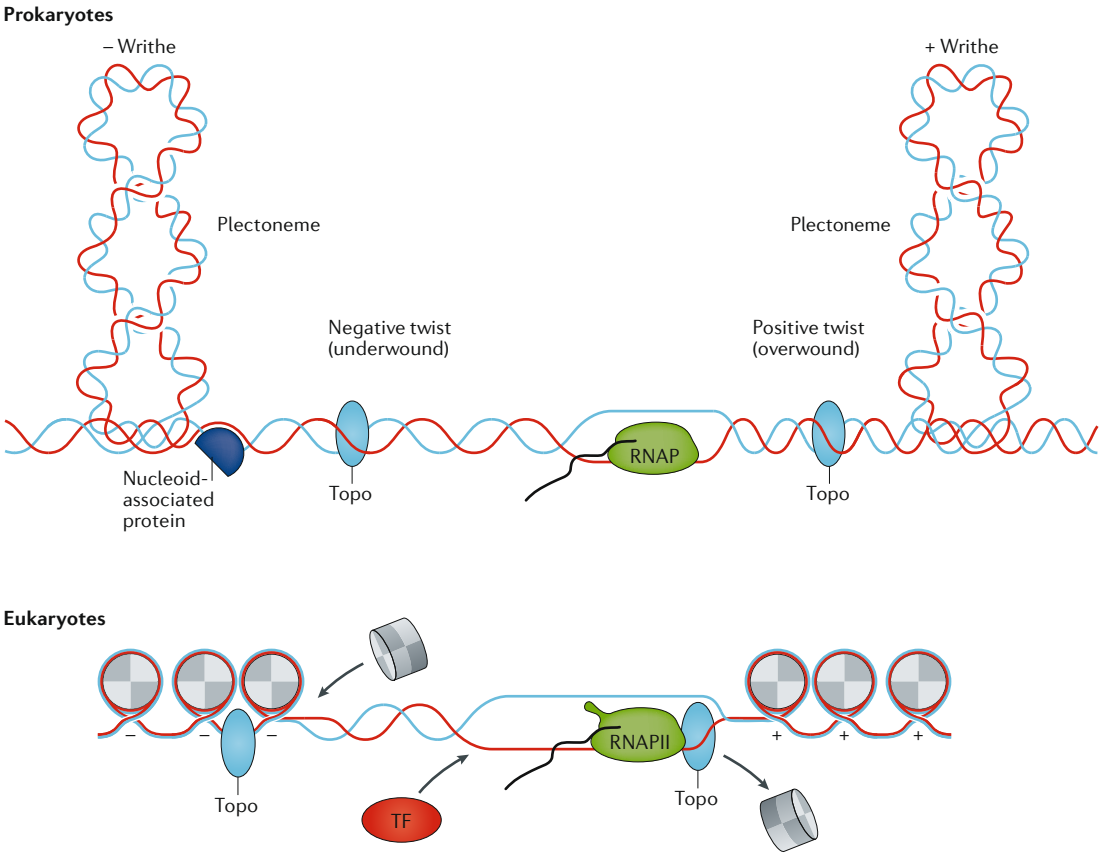


Fig. 2 | **Supercoiling in prokaryotes and eukaryotes.** Supercoiling can manifest as a twist of the DNA double helix about its axis or as a writhe of the DNA looping on itself. Melting and unwinding by polymerases progressing along the DNA produce positive supercoiling ahead of the polymerase and negative supercoiling behind the polymerase. Nucleoid-associated proteins in prokaryotes and nucleosomes in eukaryotes constrain supercoils. In eukaryotes, nucleosomes force supercoiling into the linker regions between nucleosomes. Positive supercoils ahead of a polymerase can evict nucleosomes, whereas negative supercoils behind can favour transcription factor (TF) binding or nucleosome assembly. Topoisomerases (Topos) break one or both strands of DNA to relax supercoils, relieving superhelical torsion. RNAP, RNA polymerase.

CHD/Mi-2, and INO80 (REF.[67]), and was represented in LECA by at least six members[12]. Distant homologues of remodellers comprising two prokaryotic Snf2 subfamilies exist in bacteria and some archaea, where they are probably nonessential and show no relationship to the presence or absence of archaeal histones[68], indicating that the eukaryotic remodelling functions evolved de novo to accommodate the acquisition of nucleosomes. An Snf2 homologue has a conserved role in transcription initiation[69] in several NCLDVs[59], suggesting that the ancestral NCLDV Snf2 may also have been ancestral to eukaryotic remodellers (TABLE 1). The subfamilies of remodellers are distinguished by additional chromatin-binding domains and by their mechanisms of action (BOX 3) (reviewed elsewhere[67]). Members in each subfamily form multiple lineage-specific complexes with other subunits that can interact with DNA, TFs, and histone modifications and thereby determine their specificity[67,70,71].

A nucleosome-depleted region (NDR) is found at budding yeast promoters upstream of active transcription start sites (TSSs) and is flanked by two well-positioned nucleosomes[72,73], termed the −1 nucleosome (upstream) and +1 nucleosome (downstream). NDRs are maintained by poly(dA:dT) sequence elements that have evolved to resist nucleosome formation[74], by the action of remodellers, and by the binding of TFs, PIC components and RNAPII[75]. NDR length in yeast is correlated with levels of transcription, binding of TBP, and trimethylation of H3 on lysine 4 (H3K4me3) of adjacent nucleosomes and is anticorrelated with H1 occupancy and H2A.Z occupancy[76]. The NDR is widely conserved in eukaryotes, and TF binding and nucleosome depletion are highly concordant throughout plant and animal development[77,78]. In zebrafish, depleting nuclear histone concentration while retaining genome-wide nucleosome density causes early activation of zygotic transcription by enabling TFs to quickly outcompete nucleosomes bound at their target sites[79]. In *D. melanogaster*, DNA replication removes both nucleosomes and TFs, with recolonization behind the replication fork by nucleosomes that are subsequently outcompeted by TFs in a process that is likely to be both concentration-dependent and remodeller-dependent[80]. These observations suggest a simple mass-action model of DNA occupation by nucleosomes and TFs, with adjustments by remodellers, which could apply in diverse developmental contexts.

+1 Nucleosome
A highly conserved nucleosome positioned downstream of the nucleosome-depleted region, which acts as a barrier to transcription elongation by RNA polymerase.

Box 3 | **Mechanisms of ATP-dependent remodellers**

The Snf2 family catalytic subunits of ATP-dependent remodellers belong to the superfamily 2 DNA and RNA helicases but have lost helicase activity while retaining the ability to translocate on DNA[68]. The translocase domains of remodellers bind to DNA on the nucleosome surface and utilize ATP hydrolysis to introduce superhelical torsion or twist, generating the power stroke for remodelling that breaks histone–DNA contacts[67,179]. Other than INO80, which binds to DNA close to superhelical location 6 (SHL6) at the edge of nucleosomes[124], the translocase domains of all other remodellers bind at SHL2 close to the centre of nucleosomes to initiate DNA translocation[67].
The translocase tracks the nucleosomal DNA unidirectionally by pulling DNA towards itself and pumping DNA past the site of translocation while remaining tethered by another DNA-anchor or histone-anchor. Nicks or gaps in the DNA backbone of the tracking DNA strand greatly impede translocation[180].

Remodellers tune this basic DNA-tracking and DNA-propelling action to bring about specific nucleosomal changes such as nucleosome spacing (ISWI, CHD and INO80), octamer eviction (SWI/SNF), or histone dimer exchange (INO80). Remodeller-specific regulatory domains within the catalytic subunit or associated subunits in the remodeller complex positively or negatively regulate the ATPase and DNA-translocation activities[70] (reviewed elsewhere[67]), which are often modulated by histone interactions.
For example, the H4 amino-terminal tail and the acidic patch of H2A or H2A.Z[70] relieve autoinhibition of ISWI remodellers[181,182].

Besides translocating DNA on the histone octamer surface, remodellers can also deform the histone octamer core[183]. Preventing deformation near SHL2 inhibits sliding by the ISWI subfamily SNF2h ATPase, but not by INO80, and increases nucleosome eviction by the SWI/SNF subfamily RSC complex. In yeast, RSC-bound +1 nucleosomes are asymmetrically disrupted in their DNA contacts[120], consistent with a cryo-electron-microscopy reconstruction of an RSC–nucleosome complex in which DNA and an H2A–H2B dimer[184] are displaced from the nucleosome surface.

Yeast TFs differ widely in their intrinsic ability to deplete nucleosomes. Six of 104 TFs studied in vivo bind to DNA tightly at a single motif, are highly expressed, and have strong nucleosome depletion activity, including the ubiquitous so-called general regulatory factors Abf1, Rap1 and Reb1 (REF.[81]). Other TFs, such as Rsc3, a subunit of the RSC remodelling complex, generally require multiple binding sites to deplete nucleosomes, and two-thirds of the 104 TFs are unable to deplete nucleosomes, presumably relying on other TFs, remodellers, or transient unwrapping to access nucleosomal DNA. General regulatory factors and RSC can act independently to deplete nucleosomes at NDRs[82]. RSC slides the +1 nucleosome towards the gene body, oriented by GC-rich and poly(A) motifs in the NDR, to better expose the TATA promoter element to enable binding of TBP and PIC formation.

Similar to yeast general regulatory factors, pioneer factors in animals are thought to be able to displace nucleosomes at enhancers and promoters to initiate tissue-specific developmental programmes when other TFs cannot (reviewed elsewhere[83]). The pioneer factors FOXA1 (also known as HNF3A) and FOXA2 (also known as HNF3B) have domains that are structurally similar to histone H1, which binds to entry and exit DNA. They can displace H1, which may enable access by other TFs[84]. By contrast, in mouse 3134 cells, the glucocorticoid receptor, which binds to DNA upon activation by hormone, binds mostly to nucleosomes, but also in NDRs, and recruits the SWI/SNF remodeller BRG1 to establish an NDR at the receptor binding site[85]. The pioneer factors Zelda and Gaf act during zygotic transcriptional activation in *D. melanogaster* embryogenesis to deplete nucleosomes and facilitate acetylation and binding by other TFs[86].

Pioneer factors
Transcription factors that can bind to nucleosome-occluded DNA and promote accessibility to transcription, often at an early step in a developmental pathway.

Acetylation and nucleosome disruption

PTMs of histone amino-terminal tails of the four core histones constitute a major mechanism for directing chromatin proteins that bind to specific PTMs to their histone targets. Well-studied PTMs are predominantly found on the highly conserved tails of H3 and H4 (REF.[15]), perhaps because H3–H4 tetramers have lower turnover than H2A–H2B dimers[87]. The smaller number of PTMs on H2A and H2B tails might be the reason they are less strictly conserved[30] than H3 and H4 tails.

Acetylation was proposed decades ago to structurally alter nucleosomes and promote their disruption through neutralizing the charge on the lysines of the tails, reducing DNA–histone binding and limiting their ability to form hydrogen bonds[88]. Histone acetylation improves the efficiency with which RNAPII traverses chromatin in vitro, similar to the transcriptional effect of removing the tails entirely[89]. The discovery of acetylation turnover by HATs and HDACs on histones revealed histone acetylation to be an active regulatory mechanism[88].

N-lysine acetyltransferase homologues of HATs are found in archaea, bacteria and viruses[41,69], where they acetylate non-histone substrates. Indeed, HATs and HDACs are more properly known as lysine acetyl transferases (KATs) and lysine deacetylases (KDACs) because they acetylate and deacetylate thousands of other proteins besides histones (reviewed elsewhere[90]). In addition, they can add and remove longer-chain acyl groups to and from lysine, although in HeLa cells, except for propionylation, which is nearly as abundant as acetylation, longer-chain acylations are more than 200-fold less abundant than acetylations[91], meaning that any possibilities of metabolic functions are only speculative (reviewed elsewhere[92]).

Acetylation of the nucleoid-associated protein MtHU in *Mycobacterium tuberculosis* reduces DNA interaction and decompacts the genome[93]. Attempts to acetylate archaeal histones were unsuccessful[94], although archaea acetylate the chromatin protein Alba, which represses in vitro transcription after deacetylation by a sirtuin[95]. Eukaryotic histones have increased affinity for DNA and less flexibility than archaeal histones, suggesting these features may have necessitated their acetylation for mobilization. LECA had at least four families of HATs in the GNAT/MYST domain superfamily, as well as the unrelated TAF1 HAT family associated with TBP[41]. The RPD3 and sirtuin HDAC superfamilies were present in LECA, and additional HDACs have been acquired from bacteria at multiple different points in eukaryotic evolution (TABLE 1).

In human cells, both HATs and HDACs associate with active genes and correlate with H3 and H4 acetylation levels, RNAPII levels and gene expression levels[96], and inhibition of either HDACs or p300/CBP HAT activity inhibits gene expression and reduces RNAPII occupancy, suggesting that acetylation turnover is more important than static acetylation[97], perhaps to first loosen chromatin for transcription and then to reset the chromatin landscape to prevent inappropriate transcription initiation. Acetylation of promoters is not dependent on transcription in a yeast RNAPII mutant[98], and transcription-coupled acetylation, despite

being important for RNAPII transit[89], appears to constitute only a small fraction of steady-state histone acetylation[98].

HATs are frequently found in multisubunit complexes and can be recruited to promoters by the interaction of these subunits with the acidic activation domains of TFs[99]. Tethering of the p300 acetyltransferase core to promoters or enhancers of inactive genes is sufficient to drive their robust expression[100], suggesting that a major role of TFs is to target acetylation to the genes they activate. However, p300 may also have a role in recruiting RNAPII. The *D. melanogaster* homologue of p300, Cbp, is required to maintain paused RNAPII and to overcome the barrier of the +1 nucleosome, which it acetylates. At highly paused promoters, CBP may recruit RNAPII through an interaction with TFIIB[101].

In addition to a structural role in loosening DNA contacts, acetyl-lysine provides binding sites for three different protein domains: bromodomains, double plant homeodomain fingers (DPFs), and YEATS domains. These acetyl-binding domains are often found in multidomain proteins or protein complexes with other chromatin-binding and enzymatic domains[102], consistent with the notion that acetyl-lysine plays a critical role in localizing and/or regulating these proteins at promoters. For HATs with acetyl-binding domains, this localization is self-reinforcing. At least four bromodomains were present in LECA[41]. The bromodomains of the HAT Gcn5 and the remodeller Swi2/Snf2 are required in yeast for the stable occupancy at promoter nucleosomes of the SAGA transcription complex and the SWI/SNF remodelling complex[103]. Similarly, DPFs are found in the HATs MOZ and MORF where they bind to H3K9ac or H3K14ac and facilitate association with chromatin[104]. The YEATS domain protein Yaf9 preferentially binds to H3K27ac[105] and is a component of NuA4 and of the SWR1 complex that deposits H2A.Z. Yaf9 is required for acetylation of H2A.Z by NuA4 and for H2A.Z incorporation at a third of yeast promoters[106]. Similarly, in mouse embryonic stem cells (ESCs), the YEATS domain of GAS41, a component of the TIP60 and SRCAP complexes that deposit H2A.Z, binds to H3K27ac and is necessary for H2A.Z deposition at bivalent promoters[107]. These findings suggest that acetylation was already an important promoter landmark in LECA.

SET domain proteins serve as histone methyltransferases in all eukaryotes. At least five were present in LECA, and SET domain proteins of unknown function are widespread in bacteria[41,108]. Trimethylation of H3K4 by SET1 family histone methyltransferases, which had a homologue in LECA[12], appears to be important for directing acetylation to promoters[15]. A minor fraction of HATs and HDACs are found at inactive human genes, and those inactive genes marked with H3K4me1, H3K4me2, or H3K4me3 at their promoters are more likely to become acetylated and occupied by RNAPII upon inhibition of HDACs[96], consistent with the fact that in animals, p300/CBP is necessary for dynamic acetylation on H3 tails marked with K4me3 (REF.[97]). Several other HATs and HDACs contain plant homeodomain (PHD) fingers or tandem tudor domains that bind to H3K4me3 (REF.[15]). In yeast, the NuA3 acetyltransferase has PHD-finger and PWWP domains that independently recruit NuA3 to H3K4me and H3K36me, respectively, although recruitment of NuA3 did not necessarily result in acetylation, suggesting an additional layer of regulation[109]. Mammalian BRWD2/PHIP colocalizes extensively with H3K4me through its cryptotudor domain, and depletion of the *D. melanogaster* homologue disrupts the pattern of H3K27ac[110]. H3K4 methyltransferases, in turn, can be directed to promoters by the Ser5-initiating form of the RNAPII CTD co-transcriptionally or by TFs, especially at mammalian CpG islands[15]. The mammalian MLL H3K4 methyltransferase contains a bromodomain[102], suggesting that HATs and MLL may recruit each other. H3K4me might be largely redundant with TFs in directing HATs, which could explain why loss of H3K4me has few phenotypic consequences other than a reduced proliferative rate in *D. melanogaster* cells[111], despite H3K4me being conserved throughout eukaryotes[112–114].

Chromatin diversification after LECA

While many features of chromatin regulation of promoters are highly conserved across plants, animals, and fungi, variations occur in early-diverging eukaryotes.

Promoter architecture. The root of the eukaryotic tree remains controversial, but metamonads branch near the likely root[10], and the kinetoplastid *Trypanosoma brucei* probably also represents an early branch. The metamonads *Giardia intestinalis* and *Trichomonas vaginalis* have simple promoter regions of generally <100 bp. They lack the conserved heptapeptide repeat of the RNAPII CTD, the general TF TFIIA, H2A.Z, H1, p300/CBP acetyltransferases, histone demethylases, and DNA methyltransferases[12,35,115], which could reflect either loss or a primitive absence of these proteins (TABLE 1). In *T. vaginalis*, H3K27ac and H3K4me3 are associated with active genes[112]. In *G. intestinalis*, HDACs and the silencing mark H3K9me are important for encystation and antigenic variation[116]. Kinetoplastids such as *T. brucei* have unique polycistronic transcription units, with dispersed initiation of transcripts over approximately the first 2 kb of an ~10 kb region of more accessible nucleosomes containing H4ac, H2A.Z, and H2B.V, which is an H2B variant specifically paired with H2A.Z[117,118]. The mRNAs for individual genes are processed with the addition of a 39-nucleotide capped *trans*-spliced leader, and NDRs are found at the splice site upstream of each gene start codon rather than in the region where transcripts are initiated. These observations suggest that the most common eukaryotic promoter architecture was not yet fixed in early-diverging eukaryotes, but the basic activating and silencing PTMs were already present in LECA.

H2A.Z and nucleosome stability. Histone variants replace their cognate core histones in nucleosomes and change nucleosome properties, including how much DNA a nucleosome wraps (BOX 4). We have previously reviewed histone variants and their dynamics[3,87], and here we focus on recent advances in understanding

CpG islands
Regions of more than 200 bp near many mammalian promoters that are enriched for CG dinucleotides.

Box 4 | **H2A wrapping variants**

Arabidopsis thaliana has four types of H2A variant: conventional H2A, H2A.X, H2A.Z and plant-specific H2A.W, which all form homotypic nucleosomes[132]. H2A.Z is the least stable to temperature, and H2A is the most stable, with stability largely dependent on their differing Loop1 regions and docking domains. The longer carboxyl terminus of H2A.W protects an additional 10–15 bp of linker DNA beyond the 147 bp of most nucleosomes. H2A.W is found in heterochromatin[185], where it may serve a silencing function similar to HP1 proteins found in animals. It can be phosphorylated in the DNA damage response, similar to H2A.X in euchromatin[186].

Like H2A.W, macroH2A in animals and their holozoan relatives[187] protects 10 bp of extranucleosomal DNA with its basic protein linker region that connects the histone fold domain (HFD) to the macrodomain that distinguishes macroH2A from other H2As[188]. The linker region facilitates condensation[189], and Loop1 stabilizes DNA binding to the histone octamer[190]. MacroH2A nucleosomes are often repressive, as they are less preferred by remodellers, inhibit acetylation by p300 (REFS[191,192]), and are found on the inactive X chromosome[193], but they stabilize both active and inactive cell-specific gene expression patterns, presenting a barrier to reprogramming cells[194,195].

Placental mammals have four families of short H2As (H2A.B, H2A.L, H2A.P, and H2A.Q) that have shortened docking domains and reduced acidic patches, and they wrap only 110–130 bp of DNA[196]. All are encoded on the X chromosome, have stage-specific expression in testes, and are evolving rapidly, suggesting a role in sperm and egg genetic conflict in the specialized reproduction of placental mammals. H2A.B.3 may have a role in spermatid mRNA splicing[197]. H2A.L.2 enables the loading of transition proteins, which recruits protamines that evict nucleosomes to form mature sperm nuclei[198]. Mammals also have H2A.J, differing from H2A mostly at its carboxyl terminus[199]. Through an unknown mechanism, H2A.J promotes senescence-associated inflammatory gene expression in cells with persistent DNA damage.

H2A.Z, which is present in nearly all eukaryotes and is enriched at the +1 nucleosome of genes[119], where it appears to poise genes for transcription. H2A.Z is absent in metamonads[35], raising the possibility that this variant was absent in LECA and originated after the divergence of metamonads, but its strong conservation in nearly all other eukaryotes attests to its key importance in gene regulation (TABLE 1). H2A.Z has an extended acidic patch that stimulates ATP-dependent remodellers[70], and, in yeast, a complex of RSC bound to H2A.Z-containing nucleosomes with asymmetric DNA contacts occurs at ~5% of +1 nucleosomes, which may represent remodelling intermediates that facilitate NDR formation and/or RNAPII transit[120]. The yeast SWR1 complex, which is a member of the INO80 subfamily of remodellers, binds to the NDR and replaces H2A–H2B dimers with H2A.Z–H2B dimers on the +1 and −1 nucleosomes[121]. Deposition is facilitated by nucleosome acetylation[106,107,122]. The DNA-binding Swc4 subunit can also direct SWR1 to specific genes in *Arabidopsis thaliana*[123]. The INO80 remodeller preferentially carries out the reverse exchange of H2A.Z–H2B for H2A–H2B in vitro[122,124,125], although this has been disputed[126]. Deletion of Ino80 in yeast results in global mislocalization of H2A.Z[125] and, more specifically, increased H2A.Z levels around unresolved DNA double-strand breaks[127].

In yeast H2A.Z sits over the TSS, where it promotes RNAPII recruitment[128], and is displaced by PIC components[129]. At the mating-type locus and telomeres, it resists the spread of silent chromatin[130]. In vitro, H2A.Z nucleosomes are more mobile than H2A nucleosomes, with a lower breaking force[131] and lower thermal stability[132], although an earlier in vitro study found that they have greater stability towards salt than H2A nucleosomes do, even when acetylated[133]. In vivo, in *D. melanogaster* cells, where the +1 nucleosome is downstream of the TSS, H2A.Z lowers the barrier to transcription presented by the +1 nucleosome[134]. Progression of RNAPII through the +1 nucleosome can result in loss of an H2A.Z–H2B dimer and/or its DNA contacts on the promoter-proximal side in association with stalled RNAPII or on the distal side in association with elongation and positive torsion[50]. Such loss of dimer–DNA contact may underlie the eviction of H2A.Z without loss of H3 in *A. thaliana* temperature-responsive gene promoters[135] and the enrichment of subnucleosomal DNA fragments and increased accessibility found around sites occupied by H2A.Z and p300 in enhancers in mouse ESCs[136].

It has long been puzzling that H2A.Z can be either activating or inhibiting for transcription, but if H2A.Z is more easily disrupted at the +1 nucleosome and stimulates remodellers, these different responses may reflect the different contexts in which H2A.Z nucleosomes reside. In the *A. thaliana* thermal response, H2A.Z is inhibitory until temperature-stimulated heat shock factors bind to promoters and stimulate transcription[135]. In mouse pituitary gonadotropes, the position of H2A.Z relative to the TSS regulates high or low expression of luteinizing hormone subunits[131]. In mouse ESCs, H2A.Z colocalizes with H3K4me3 at both promoters and enhancers and facilitates targeting of the histone methyltransferases for the active mark H3K4me3 and the repressive mark H3K27me3, especially at enhancers[136]. In turn, H3K4me3 promotes H2A.Z incorporation at enhancers in an apparent positive feedback loop. We speculate that such a feedback loop could explain why both H2A.Z and H3K4me2 and/or H3K4me3 are anticorrelated with DNA methylation in *A. thaliana*[137,138], as H3K4me3 inhibits DNA methylation, and DNA methylation excludes H2A.Z (BOX 2).

Interaction domains and complexes. The domain architectures of HATs, HDACs, and especially histone methyltransferases in metamonads are the simplest in eukaryotes, indicating that the interaction domains of chromatin proteins have increased with organismal complexity in later-diverging eukaryotes[12], both through the addition of PTM recognition domains to proteins and through the addition of subunits to complexes. For example, in *Candida albicans*, the NuA4 acetyltransferase and the SWR1 remodeller are separate complexes in the hyphal state but are combined in the yeast state, and domains from both complexes have been combined into one protein independently in several eukaryotes, including in human p400 of the TIP60 complex[139].

A more complex example comes from Polycomb repressive complex 1 (PRC1) and PRC2, which have essential roles in controlling cell-type-specific developmental gene expression in multicellular eukaryotes. Diversification of these complexes may have facilitated the advent of cell differentiation in multicellular organisms by serving as a flexible, modular silencing apparatus that selectively inactivates a range of *cis* elements in response to developmental cues.

Gonadotropes
Cells of the pituitary gland that secrete luteinizing hormone and follicle-stimulating hormone.

Homotypic nucleosomes
Nucleosomes in which both members of any particular histone family are the same histone variant.

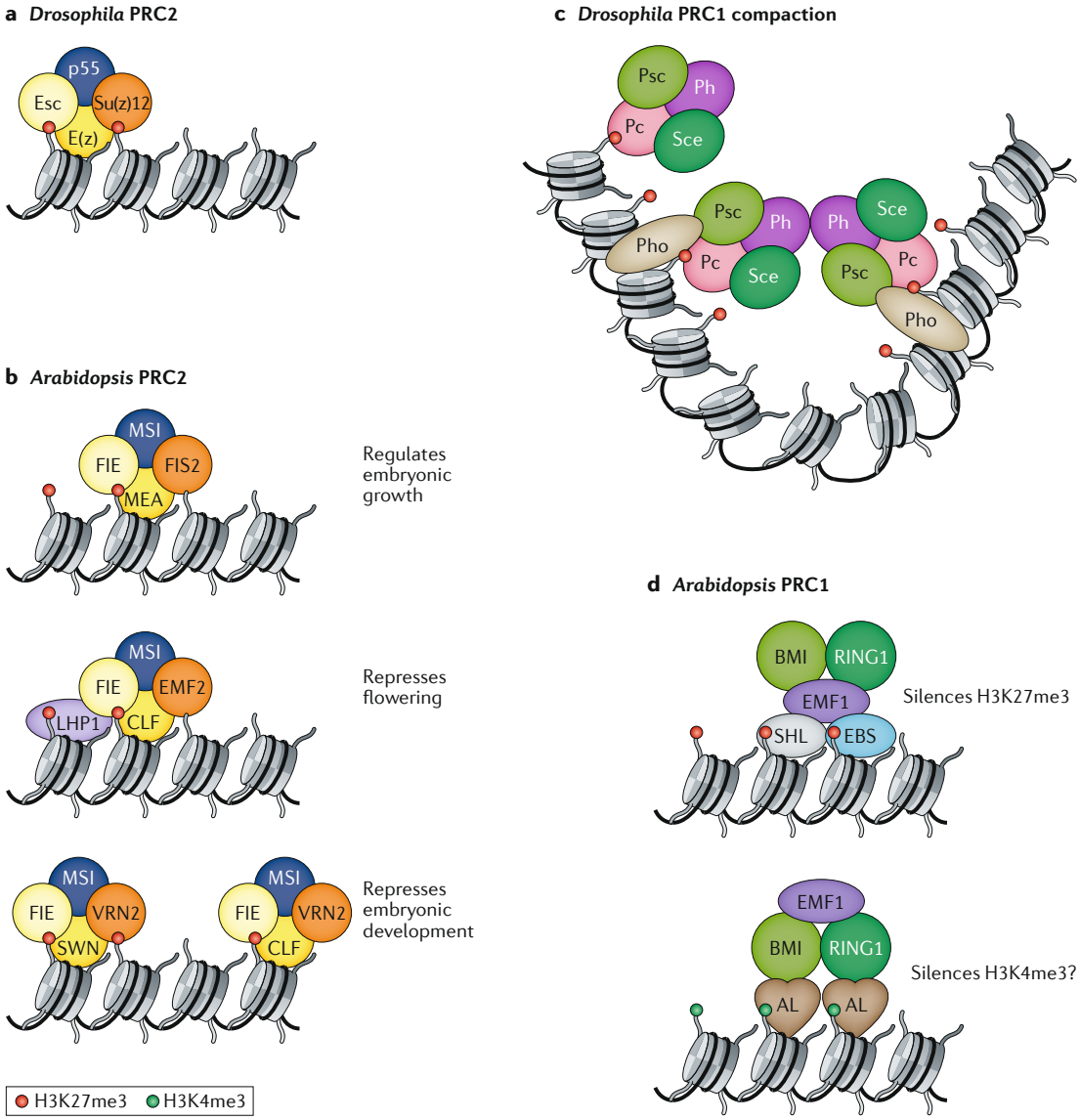


Fig. 3 | **PRC1 and PRC2 in animals and plants.** Although both plants and animals share many components of the Polycomb repressive complexes PRC1 and PRC2, they are differently deployed. **a**,**b** | PRC2 is widely conserved in eukaryotes and has four core components in *Drosophila melanogaster*, Esc, Su(z)12, p55, and the histone methyltransferase E(z), most of which are present in multiple copies in *Arabidopsis thaliana*, forming at least three distinct complexes that regulate distinct developmental programmes. The *A. thaliana* chromodomain protein LHP1 binds to H3K27me3 and together with enhancer of zeste (E(z)) homologue curly leaf (CLF) acts to spread H3K27me3. **c** | PRC1 in *D. melanogaster* contains the chromodomain protein Polycomb (Pc) that binds to H3K27me3 and ring finger E3 ubiquitin ligases Psc and Sce that ubiquitylate H2AK119; however, silencing depends on chromatin compaction by polymerization of Polyhomeotic (Ph). The complex can be localized by transcription factors (TFs) such as Pleiohomeotic (Pho). **d** | In *A. thaliana*, PRC1 complexes are not well characterized, but two complexes have been proposed containing BMI and RING1, homologues of Psc and Sce, respectively, along with plant-specific components with PHD fingers that can bind to H3K27me3 (SHL and EBS) or H3K4me3 (AL). The latter complex is proposed to shut off active genes to transition to repressed chromatin marked with H3K27me3 and H2AKub. Shapes coloured identically represent homologous proteins.

In animals, developmental silencing is dependent on methylation of H3K27 by PRC2 and H3K27me3-directed gene silencing by PRC1 (REFS[140,141]). A homologue of enhancer of zeste (E(z)) — the methyltransferase component of PRC2 that catalyses methylation of H3K27 — was present in LECA[12] and might have originally served merely to block H3K27 acetylation, perhaps to prevent transcription of transposons and repetitive sequences, as in *Chlamydomonas reinhardtii*[142], or of telomeric repeats, as in fungi[143]. Classical PRC2 in *D. melanogaster* (FIG. 3a) contains four subunits, which are present in most eukaryotes, although individual proteins have been lost in some lineages[142]. Flowering plants and vertebrates have multiple E(z) paralogues, which form multiple PRC2-like complexes that differ in cell-specific expression (FIG. 3b) and can dynamically exchange subunits to target particular genes[144,145]. Specific subunits can target PRC2 to CpG islands in mouse cells[146] or even

direct methylation to H1K26 instead of H3K27 (REF.[147]). Human PRC2 member EED (Esc homologue) binds to H3K27me3 and positions the E(z) homologue EZH2 for methylation of an adjacent nucleosome, facilitating H3K27me3 spreading to form H3K27me3 domains[148].

Classical PRC1 in *D. melanogaster* also contains four core subunits, including the chromodomain protein Polycomb, which binds to H3K27me3, and two RING-type zinc-finger E3 ubiquitin ligases (Sce and Psc) that together ubiquitylate H2AK119, as well as variable substoichiometric subunits including HDACs and the DNA-binding protein Pleiohomeotic that targets PRC1 to specific genes[144] (FIG. 3c). H2A119ub is not essential for PRC1-mediated silencing in animals[149,150]. H3K27me3 silencing is thought to occur through inhibition of transcription initiation and elongation by PRC1-mediated chromatin compaction[151], which changes during differentiation and depends on the PRC1 component Polyhomeotic[150], which can polymerize[152], potentially bringing disparate PRC1 complexes together (FIG. 3c). Vertebrates have 3–5 Polycomb homologues and several partially redundant paralogues of PRC1 Sce (RING1) and Psc (BMI) proteins[144]. Although LECA had several chromodomain proteins and RING E3 ligases, a homologue of the RING1 proteins of PRC1 was not among them[12,41], but RING1 homologues are present in plants, animals and other eukaryotes. However, fungi including *Cryptococcus neoformans* and *Neurospora crassa* lack PRC1 homologues, indicating that PRC1 is dispensable for general PRC-mediated silencing[143,153].

Plants have no polyhomeotic homologue, and other PRC1 components are deployed differently in plants[154] (FIG. 3d). The plant chromodomain protein LHP1 binds to H3K27me3 (REF.[145]) similarly to Polycomb but has a chromoshadow domain like animal H3K9me3-binding HP1, to which it is more closely related[154]. It copurifies with one of the PRC2-like complexes containing E(z) homologue curly leaf (CLF), rather than with PRC1 components, and aids in spreading H3K27me3, repressing floral genes in seedlings[145]. By contrast, RING1 and BMI in combination with CLF and the E(z) homologue SWN (also known as EZA1) repress embryonic genes. These two CLF complexes are associated with different sets of TF-binding motifs, suggesting they may be targeted by specific TFs. Two PRC1 complexes containing RING1, BMI, and plant-specific proteins have been proposed. To ubiquitylate H2A, RING1 and BMI require EMF1, which can bind to plant-specific SHL and EBS, which each bind to H3K27me3 through their PHD fingers[155]. RING1 and BMI can also bind to AL paralogues that bind to H3K4me3 with their PHD fingers and are proposed to switch off active H3K4-containing genes[156].

Conclusions and perspectives

Proto-eukaryotes assembled the enzymes and domains of chromatin proteins from precursors in bacteria, archaea, viruses and perhaps urkaryotes and invented new uses for them in response to the acquisition or invention of nucleosomes, which may have first served a genome defence role. Nucleosomes altered the role of DNA superhelicity, required mobilization by remodellers and HATs, and probably enabled the expansion of genomes through more efficient packaging and condensation. Such opportunities for genome expansion are likely to have facilitated dramatic evolution in the form of gene duplication, regulatory element proliferation, and deployment of an expanding and progressively interactive set of chromatin domains and proteins both before and since LECA, including the diversification of PRC complexes to control developmental programmes in multicellular eukaryotes. Chromatin evolution is ongoing in the rapid evolution of short H2A histone variants in mammals (BOX 4).

Continuing investigation of cell-type-specific chromatin changes in both model and early-diverging eukaryotes will better illuminate chromatin regulation and evolution. The interaction of HATs, HDACs, and remodellers with TFs, PIC components, PTMs, and each other is a very active area of investigation that is likely to flesh out details of chromatin regulation in the immediate future. Similarly, the association of different PRC complexes with specific developmental programmes in animals and plants is proceeding apace. Further investigation of viral genomes, which are less intensively studied, using existing technologies may provide insight into the 'missing' proto-eukaryote stages in the evolution of chromatin proteins. Similarly, investigation of transcriptional regulation in metamonads, particularly free-living metamonads, may illuminate details of a more primitive state of chromatin regulation of transcription, and knowledge of the chromatin proteins in other early-branching eukaryotes, such as species of *Malawimonas*, *Collodictyon*, *Ancyromonas*, and others, would help to illuminate the early stages of eukaryotic chromatin evolution. Our understanding of supercoiling domains in both gene regulation and chromosome compaction is in its infancy. Psoralen-based methods for mapping negative supercoils[51,57] and recent technologies for detecting single-strand and double-strand breaks[157] and non-B-form DNA[54] on a genome-wide scale show promise for better illuminating the role of DNA torsion, although development of additional new methods at both genomic and gene-specific scales is likely to speed progress in this area.

Published online: 18 March 2019

1. Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1–4 (1999).
2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251–260 (1997).
3. Talbert, P. B. & Henikoff, S. Histone variants — ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
4. Schopf, J. W. Fossil evidence of Archaean life. *Phil. Trans. R. Soc. B* **361**, 869–885 (2006).
5. Nutman, A. P., Bennett, V. C., Friend, C. R., Van Kranendonk, M. J. & Chivas, A. R. Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* **537**, 535–538 (2016).
6. Dodd, M. S. et al. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* **543**, 60–64 (2017).
7. Waldbauer, J. R., Sherman, L. S., Sumner, D. Y. & Summon, R. E. Late Archean molecular fossils from the Transvaal Supergroup record the antiquity of microbial diversity and aerobiosis. *Precambrian Res.* **169**, 28–47 (2009).
8. Bengtson, S., Sallstedt, T., Belivanova, V. & Whitehouse, M. Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLOS Biol.* **15**, e2000735 (2017).
9. Butterfield, N. J. Bangiomorpha pubescens n. gen., n. sp.:implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiatin of eukaryotes. *Paleobiology* **26**, 386–404 (2000).
10. Brown, M. W. et al. Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biol. Evol.* **10**, 427–433 (2018).

11. Koumandou, V. L. et al. Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* **48**, 373–396 (2013).
12. Iyer, L. M., Anantharaman, V., Wolf, M. Y. & Aravind, L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* **38**, 1–31 (2008).
13. Nasir, A., Kim, K. M., Da Cunha, V. & Caetano-Anolles, G. Arguments reinforcing the three-domain view of diversified cellular life. *Archaea* **2016**, 1851865 (2016).
The sources of eukaryotic protein superfold families are analysed and arguments are presented for three cellular domains, in contrast to Zaremba-Niedzwiedzka et al. (2017).
14. Tanny, J. C. Chromatin modification by the RNA Polymerase II elongation complex. *Transcription* **5**, e988093 (2014).
15. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications - writers that read. *EMBO Rep.* **16**, 1467–1481 (2015).
16. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
17. Sherafatian, M. & Mowla, S. J. The origins and evolutionary history of human non-coding RNA regulatory networks. *J. Bioinform. Comput. Biol.* **15**, 1750005 (2017).
18. Ishihama, A. Building a complete image of genome regulation in the model organism *Escherichia coli*. *J. Gen. Appl. Microbiol.* **63**, 311–324 (2018).
19. Blombach, F. & Grohmann, D. Same same but different: the evolution of TBP in archaea and their eukaryotic offspring. *Transcription* **8**, 162–168 (2017).
20. Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
21. Jun, S. H., Reichlen, M. J., Tajiri, M. & Murakami, K. S. Archaeal RNA polymerase and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* **46**, 27–40 (2011).
22. Zentner, G. E. & Henikoff, S. Mot1 redistributes TBP from TATA-containing to TATA-less promoters. *Mol. Cell. Biol.* **33**, 4996–5004 (2013).
23. Guglielmini, J., Woo, A., Krupovic, M., Forterre, P. & Gaia, M. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. Preprint at *bioRxiv* https://doi.org/10.1101/455816 (2018).
24. Alva, V. & Lupas, A. N. Histones predate the split between bacteria and archaea. *Bioinformatics* https://doi.org/10.1093/bioinformatics/bty1000 (2018).
25. Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R. T. Structure and function of archaeal histones. *PLOS Genet.* **14**, e1007582 (2018).
26. Mattiroli, F. et al. Structure of histone-based chromatin in Archaea. *Science* **357**, 609–612 (2017).
This paper shows that archaeal histones can form extended polymers that wrap DNA and affect gene regulation.
27. Xie, Y. & Reeve, J. N. Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J. Bacteriol.* **186**, 3492–3498 (2004).
28. Wilkinson, S. P., Ouhammouch, M. & Geiduschek, E. P. Transcriptional activation in the context of repression mediated by archaeal histones. *Proc. Natl Acad. Sci. USA* **107**, 6777–6781 (2010).
29. Ammar, R. et al. Chromatin is an ancient innovation conserved between Archaea and Eukarya. *eLife* **1**, e00078 (2012).
30. Malik, H. S. & Henikoff, S. Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10**, 882–891 (2003).
31. Erives, A. J. Phylogenetic analysis of the core histone doublet and DNA topo II genes of Marseilleviridae: evidence of proto-eukaryotic provenance. *Epigenetics Chromatin* **10**, 55 (2017).
This paper analyses the phylogeny of histones and topoisomerase II in giant viruses and proposes that they reflect a proto-eukaryotic stage of histone evolution.
32. Marinov, G. K. & Lynch, M. Diversity and divergence of dinoflagellate histone proteins. *G3 (Bethesda)* **6**, 397–422 (2015).
33. Gornik, S. G. et al. Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Curr. Biol.* **22**, 2303–2312 (2012).
34. Irwin, N. A. T. et al. Viral proteins as a potential driver of histone depletion in dinoflagellates. *Nat. Commun.* **9**, 1535 (2018).
35. Dalmasso, M. C., Sullivan, W. J. Jr & Angel, S. O. Canonical and variant histones of protozoan parasites. *Front. Biosci. (Landmark Ed)* **16**, 2086–2105 (2011).
36. Kasinsky, H. E., Lewis, J. D., Dacks, J. B. & Ausio, J. Origin of H1 linker histones. *FASEB J.* **15**, 34–42 (2001).
37. Shintomi, K. et al. Mitotic chromosome assembly despite nucleosome depletion in *Xenopus* egg extracts. *Science* **356**, 1284–1287 (2017).
The authors show that condensins and topoisomerase II can form chromosome axes without nucleosomes, but nucleosomes are needed for full condensation and to prevent fragility.
38. Shintomi, K., Takahashi, T. S. & Hirano, T. Reconstitution of mitotic chromatids with a minimum set of purified factors. *Nat. Cell Biol.* **17**, 1014–1023 (2015).
39. Madhani, H. D. The frustrated gene: origins of eukaryotic gene expression. *Cell* **155**, 744–749 (2013).
40. Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J. & Craig, N. L. DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. *Proc. Natl Acad. Sci. USA* **107**, 21966–21972 (2010).
41. Aravind, L., Burroughs, A. M., Zhang, D. & Iyer, L. M. Protein and DNA modifications: evolutionary imprints of bacterial biochemical diversification and geochemistry on the provenance of eukaryotic epigenetics. *Cold Spring Harb. Perspect. Biol.* **6**, a016063 (2014).
42. Zemach, A. & Zilberman, D. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Biol.* **20**, R780–785 (2010).
43. Blot, N., Mavathur, R., Geertz, M., Travers, A. & Muskhelishvili, G. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.* **7**, 710–715 (2006).
44. Muskhelishvili, G. & Travers, A. The regulatory role of DNA supercoiling in nucleoprotein complex assembly and genetic activity. *Biophys. Rev.* **8**, 5–22 (2016).
45. Sobetzko, P. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. *Nucleic Acids Res.* **44**, 1514–1524 (2016).
46. Le, T. B., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731–734 (2013).
47. Sperling, A. S., Jeong, K. S., Kitada, T. & Grunstein, M. Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc. Natl Acad. Sci. USA* **108**, 12693–12698 (2011).
48. Sivolob, A. & Prunell, A. Linker histone-dependent organization and dynamics of nucleosome entry/exit DNAs. *J. Mol. Biol.* **331**, 1025–1040 (2003).
49. Sheinin, M. Y., Li, M., Soltani, M., Luger, K. & Wang, M. D. Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. *Nat. Commun.* **4**, 2579 (2013).
50. Ramachandran, S., Ahmad, K. & Henikoff, S. Transcription and remodeling produce asymmetrically unwrapped nucleosomal intermediates. *Mol. Cell* **68**, 1038–1053 (2017).
Mapping of protected subnucleosomal DNA fragments in *D. melanogaster* cells demonstrates that positive torsion generated by RNAPII elongation disrupts DNA contacts with the +1 nucleosome and that subnucleosomal DNA fragments from cell-free DNA in human blood plasma can be used to infer transcriptional status.
51. Teves, S. S. & Henikoff, S. Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* **21**, 88–94 (2014).
52. Gaykalova, D. A. et al. Structural analysis of nucleosomal barrier to transcription. *Proc. Natl Acad. Sci. USA* **112**, E5787–E5795 (2015).
53. Baranello, L. et al. RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell* **165**, 357–371 (2016).
54. Kouzine, F. et al. Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356 (2017).
The non-B-form DNA is globally mapped in mammalian cells by potassium permanganate footprinting and found to be enriched in the promoters of developmentally regulated genes.
55. Kouzine, F., Levens, D. & Baranello, L. DNA topology and transcription. *Nucleus* **5**, 195–202 (2014).
56. Tsochatzidou, M., Malliarou, M., Papanikolaou, N., Roca, J. & Nikolaou, C. Genome urbanization: clusters of topologically co-regulated genes delineate functional compartments in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **45**, 5818–5828 (2017).
This paper describes topologically co-regulated gene clusters that are concurrently either upregulated or downregulated in response to topological stress and shows that these clusters are organized nonrandomly in the budding yeast genome.
57. Naughton, C. et al. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* **20**, 387–395 (2013).
58. Forterre, P. & Gadelle, D. Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res.* **37**, 679–692 (2009).
59. Iyer, L. M., Balaji, S., Koonin, E. V. & Aravind, L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* **117**, 156–184 (2006).
60. Forterre, P. & Gaia, M. Giant viruses and the origin of modern eukaryotes. *Curr. Opin. Microbiol.* **31**, 44–49 (2016).
61. Earnshaw, W. C., Halligan, B., Cooke, C. A., Heck, M. M. & Liu, L. F. Topoisomerase II is a structural component of mitotic chromosome scaffolds. *J. Cell Biol.* **100**, 1706–1715 (1985).
62. Baxter, J. et al. Positive supercoiling of mitotic DNA drives decatenation by topoisomerase II in eukaryotes. *Science* **331**, 1328–1332 (2011).
63. Uuskula-Reimand, L. et al. Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **17**, 182 (2016).
64. Baxter, J. & Aragon, L. A model for chromosome condensation based on the interplay between condensin and topoisomerase II. *Trends Genet.* **28**, 110–117 (2012).
65. Perez-Rueda, E. & Janga, S. C. Identification and genomic analysis of transcription factors in archaeal genomes exemplifies their functional architecture and evolutionary origin. *Mol. Biol. Evol.* **27**, 1449–1459 (2010).
66. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262 (2005).
67. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-ependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18**, 407–422 (2017).
68. Flaus, A., Martin, D. M., Barton, G. J. & Owen-Hughes, T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res.* **34**, 2887–2905 (2006).
69. de Souza, R. F., Iyer, L. M. & Aravind, L. Diversity and evolution of chromatin proteins encoded by DNA viruses. *Biochim. Biophys. Acta* **1799**, 302–318 (2010).
70. Dann, G. P. et al. ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature* **548**, 607–611 (2017).
The authors demonstrate that remodelling complexes with the same ATPase but different accessory subunits respond differently to the same histone modification and that all remodellers require the acidic patch of H2A or H2A.Z for efficient remodelling. They propose that modifications around the acidic patch can modulate remodelling efficiency.
71. Lessard, J. et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201–215 (2007).
72. Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905 (2004).
73. Yuan, G. C. et al. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
74. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20**, 267–273 (2013).
75. Chereji, R. V. & Clark, D. J. Major determinants of nucleosome positioning. *Biophys. J.* **114**, 2279–2289 (2018).
76. Chereji, R. V., Ramachandran, S., Bryson, T. D. & Henikoff, S. Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol.* **19**, 19 (2018).
77. Song, L. et al. A transcription factor hierarchy defines an environmental stress response network. *Science* **354**, aag1550 (2016).

78. Uyehara, C. M. et al. Hormone-dependent control of developmental timing through regulation of chromatin accessibility. *Genes Dev.* **31**, 862–875 (2017).
79. Joseph, S. R. et al. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *eLife* **6**, e23326 (2017).
In zebrafish embryos, the concentration of TFs and of histones determines the timing of zygotic transcription without altering nucleosome density, supporting the idea of competition between nucleosomes and TFs in transcription activation.
80. Ramachandran, S. & Henikoff, S. Transcriptional regulators compete with nucleosomes post-replication. *Cell* **165**, 580–592 (2016).
81. Yan, C., Chen, H. & Bai, L. Systematic study of nucleosome-displacing factors in budding yeast. *Mol. Cell* **71**, 294–305 (2018).
The authors systematically evaluate the nucleosome-displacing properties of 104 TFs.
82. Kubik, S. et al. Sequence-directed action of RSC remodeler and general regulatory factors modulates +1 nucleosome position to facilitate transcription. *Mol. Cell* **71**, 89–102 (2018).
83. Sartorelli, V. & Puri, P. L. Shaping gene expression by landscaping chromatin architecture: lessons from a master. *Mol. Cell* **71**, 375–388 (2018).
84. Iwafuchi-Doi, M. et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).
85. Johnson, T. A. et al. Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. *Nucleic Acids Res.* **46**, 203–214 (2018).
86. Schulz, K. N. et al. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early *Drosophila* embryo. *Genome Res.* **25**, 1715–1726 (2015).
87. Talbert, P. B. & Henikoff, S. Histone variants on the move: substrates for chromatin dynamics. *Nat. Rev. Mol. Cell Biol.* **18**, 115–126 (2017).
88. Verdin, E. & Ott, M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat. Rev. Mol. Cell Biol.* **16**, 258–264 (2015).
89. Protacio, R. U., Li, G., Lowary, P. T. & Widom, J. Effects of histone tail domains on the rate of transcriptional elongation through a nucleosome. *Mol. Cell. Biol.* **20**, 8866–8878 (2000).
90. Narita, T., Weinert, B. T. & Choudhary, C. Functions and mechanisms of non-histone protein acetylation. *Nat. Rev. Mol. Cell Biol.* https://doi.org/10.1038/s41580-018-0081-3 (2018).
91. Klein, B. J. et al. Recognition of histone H3K14 acylation by MORF. *Structure* **25**, 650–654 (2017).
92. Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nat. Rev. Mol. Cell Biol.* **18**, 90–101 (2017).
93. Ghosh, S., Padmanabhan, B., Anand, C. & Nagaraja, V. Lysine acetylation of the *Mycobacterium tuberculosis* HU protein modulates its DNA binding and genome organization. *Mol. Microbiol.* **100**, 577–588 (2016).
94. Soares, D. J., Marc, F. & Reeve, J. N. Conserved eukaryotic histone-fold residues substituted into an archaeal histone increase DNA affinity but reduce complex flexibility. *J. Bacteriol.* **185**, 3453–3457 (2003).
95. Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P. & White, M. F. The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* **296**, 148–151 (2002).
96. Wang, Z. et al. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019–1031 (2009).
97. Crump, N. T. et al. Dynamic acetylation of all lysine-4 trimethylated histone H3 is evolutionarily conserved and mediated by p300/CBP. *Proc. Natl Acad. Sci. USA* **108**, 7814–7819 (2011).
98. Durant, M. & Pugh, B. F. Genome-wide relationships between TAF1 and histone acetyltransferases in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **26**, 2791–2802 (2006).
99. Brown, C. E. et al. Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit. *Science* **292**, 2333–2337 (2001).
100. Hilton, I. B. et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33**, 510–517 (2015).
101. Boija, A. et al. CBP regulates recruitment and release of promoter-proximal RNA polymerase II. *Mol. Cell* **68**, 491–503 (2017).
In *D. melanogaster* promoters, the authors find that Cbp maintains RNAPII at the pause site, acetylates the +1 nucleosome and is required to overcome the +1 nucleosome barrier to transcription.
102. Fujisawa, T. & Filippakopoulos, P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.* **18**, 246–262 (2017).
103. Hassan, A. H. et al. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111**, 369–379 (2002).
104. Ali, M. et al. Tandem PHD fingers of MORF/MOZ acetyltransferases display selectivity for acetylated histone H3 and are required for the association with chromatin. *J. Mol. Biol.* **424**, 328–338 (2012).
105. Klein, B. J. et al. Yaf9 subunit of the NuA4 and SWR1 complexes targets histone H3K27ac through its YEATS domain. *Nucleic Acids Res.* **46**, 421–430 (2018).
106. Wang, A. Y. et al. Asf1-like structure of the conserved Yaf9 YEATS domain and role in H2A. Z deposition and acetylation. *Proc. Natl Acad. Sci. USA* **106**, 21573–21578 (2009).
107. Hsu, C. C. et al. Gas41 links histone acetylation to H2A. Z deposition and maintenance of embryonic stem cell identity. *Cell Discov.* **4**, 28 (2018).
108. Alvarez-Venegas, R., Sadder, M., Tikhonov, A. & Avramova, Z. Origin of the bacterial SET domain genes: vertical or horizontal? *Mol. Biol. Evol.* **24**, 482–497 (2007).
109. Martin, B. J. et al. Histone H3K4 and H3K36 methylation independently recruit the NuA3 histone acetyltransferase in *Saccharomyces cerevisiae*. *Genetics* **205**, 1113–1123 (2017).
110. Morgan, M. A. J. et al. A cryptic Tudor domain links BRWD2/PHIP to COMPASS-mediated histone H3K4 methylation. *Genes Dev.* **31**, 2003–2014 (2017).
111. Hodl, M. & Basler, K. Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr. Biol.* **22**, 2253–2257 (2012).
112. Song, M. J. et al. Epigenome mapping highlights chromatin-mediated gene regulation in the protozoan parasite *Trichomonas vaginalis*. *Sci. Rep.* **7**, 45365 (2017).
113. Gupta, A. P. & Bozdech, Z. Epigenetic landscapes underlining global patterns of gene expression in the human malaria parasite. *Plasmodium falciparum*. *Int. J. Parasitol.* **47**, 399–407 (2017).
114. Hsu, D. W., Chubb, J. R., Muramoto, T., Pears, C. J. & Mahadevan, L. C. Dynamic acetylation of lysine-4-trimethylated histone H3 and H3 variant biology in a simple multicellular eukaryote. *Nucleic Acids Res.* **40**, 7247–7256 (2012).
115. Vanacova, S., Liston, D. R., Tachezy, J. & Johnson, P. J. Molecular biology of the amitochondriate parasites, *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis*. *Int. J. Parasitol.* **33**, 235–255 (2003).
116. Carranza, P. G. et al. Specific histone modifications play critical roles in the control of encystation and antigenic variation in the early-branching eukaryote. *Giardia lamblia*. *Int. J. Biochem. Cell Biol.* **81**, 32–43 (2016).
117. Wedel, C., Forstner, K. U., Derr, R. & Siegel, T. N. GT-rich promoters can drive RNA pol II transcription and deposition of H2A. Z in African trypanosomes. *EMBO J.* **36**, 2581–2594 (2017).
118. Siegel, T. N. et al. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.* **23**, 1063–1076 (2009).
119. Raisner, R. M. et al. Histone variant H2A. Z marks the 5′ ends of both active and inactive genes in euchromatin. *Cell* **123**, 233–248 (2005).
120. Ramachandran, S., Zentner, G. E. & Henikoff, S. Asymmetric nucleosomes flank promoters in the budding yeast genome. *Genome Res.* **25**, 381–390 (2015).
121. Ranjan, A. et al. Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A. Z replacement. *Cell* **154**, 1232–1245 (2013).
122. Watanabe, S., Radman-Livaja, M., Rando, O. J. & Peterson, C. L. A histone acetylation switch regulates H2A. Z deposition by the SWR-C remodeling enzyme. *Science* **340**, 195–199 (2013).
123. Gomez-Zambrano, A. et al. Arabidopsis SWC4 binds DNA and recruits the SWR1 complex to modulate histone H2A.Z deposition at key regulatory genes. *Mol. Plant* **11**, 815–832 (2018).
124. Brahma, S. et al. INO80 exchanges H2A. Z for H2A by translocating on DNA proximal to histone dimers. *Nat. Commun.* **8**, 15616 (2017).
125. Papamichos-Chronakis, M., Watanabe, S., Rando, O. J. & Peterson, C. L. Global regulation of H2A. Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200–213 (2011).
126. Wang, F., Ranjan, A., Wei, D. & Wu, C. Comment on "A histone acetylation switch regulates H2A. Z deposition by the SWR-C remodeling enzyme". *Science* **353**, 358 (2016).
127. Papamichos-Chronakis, M., Krebs, J. E. & Peterson, C. L. Interplay between Ino80 and Swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptation in response to DNA damage. *Genes Dev.* **20**, 2437–2449 (2006).
128. Adam, M., Robert, F., Larochelle, M. & Gaudreau, L. H2A. Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. *Mol. Cell. Biol.* **21**, 6270–6279 (2001).
129. Tramantano, M. et al. Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *eLife* **5**, e14243 (2016).
130. Meneghini, M. D., Wu, M. & Madhani, H. D. Conserved histone variant H2A. Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**, 725–736 (2003).
131. Rudnizky, S. et al. H2A. Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. *Nat. Commun.* **7**, 12958 (2016).
Optical tweezers are used to show that H2A.Z nucleosomes have a lower breaking force than H2A nucleosomes and that positioning of H2A.Z nucleosomes relative to the TSS can lead to distinct regulation strategies.
132. Osakabe, A. et al. Histone H2A variants confer specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic Acids Res.* **46**, 7675–7685 (2018).
133. Ishibashi, T. et al. Acetylation of vertebrate H2A. Z and its effect on the structure of the nucleosome. *Biochemistry* **48**, 5007–5017 (2009).
134. Weber, C. M., Ramachandran, S. & Henikoff, S. Nucleosomes are context-specific, H2A. Z-modulated barriers to RNA polymerase. *Mol. Cell* **53**, 819–830 (2014).
135. Cortijo, S. et al. Transcriptional regulation of the ambient temperature response by H2A. Z nucleosomes and HSF1 transcription factors in *Arabidopsis*. *Mol. Plant* **10**, 1258–1273 (2017).
136. Hu, G. et al. H2A. Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **12**, 180–192 (2013).
137. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A. Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129 (2008).
138. Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M. & Jacobsen, S. E. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **10**, R62 (2009).
139. Wang, X. et al. Merge and separation of NuA4 and SWR1 complexes control cell fate plasticity in *Candida albicans*. *Cell Discov.* **4**, 45 (2018).
140. Muller, J. et al. Histone methyltransferase activity of a *Drosophila* polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
141. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
142. Shaver, S., Casas-Mollano, J. A., Cerny, R. L. & Cerutti, H. Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics* **5**, 301–312 (2010).
143. Jamieson, K. et al. Telomere repeats induce domains of H3K27 methylation in *Neurospora*. *eLife* **7**, e31216 (2018).
144. Grossniklaus, U. & Paro, R. Transcriptional silencing by polycomb-group proteins. *Cold Spring Harb. Perspect. Biol.* **6**, a019331 (2014).
145. Wang, H. et al. *Arabidopsis* flower and embryo developmental genes are repressed in seedlings by different combinations of polycomb group proteins in association with distinct sets of cis-regulatory elements. *PLOS Genet.* **12**, e1005771 (2016).
The authors show that the H3K27me3-binding chromodomain protein LHP1 in *A. thaliana* is associated with the PRC2 histone

methyltransferase CLF and aids spreading of methylation to repress floral development, while PRC1 components BMI1 and RING1 associate with histone methyltransferases CLF and SWR and suppress embryo-specific genes.
146. Li, H. et al. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* **549**, 287–291 (2017).
147. Kuzmichev, A., Jenuwein, T., Tempst, P. & Reinberg, D. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* **14**, 183–193 (2004).
148. Poepsel, S., Kasinath, V. & Nogales, E. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* **25**, 154–162 (2018).
149. Pengelly, A. R., Kalb, R., Finkl, K. & Muller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev.* **29**, 1487–1492 (2015).
150. Kundu, S. et al. Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. *Mol. Cell* **65**, 432–446 (2017).
Super-resolution microscopy of mammalian cells identifies PRC1 compacted chromatin domains that depend on Polyhomeotic but not on PRC1-dependent ubiquitylation and that are lost as PRC1 binding is lost during differentiation.
151. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
152. Robinson, A. K. et al. The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. *J. Biol. Chem.* **287**, 8702–8713 (2012).
153. Dumesic, P. A. et al. Product binding enforces the genomic specificity of a yeast polycomb repressive complex. *Cell* **160**, 204–218 (2015).
154. Berke, L. & Snel, B. The plant polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. *BMC Evol. Biol.* **15**, 44 (2015).
155. Li, Z., Fu, X., Wang, Y., Liu, R. & He, Y. Polycomb-mediated gene silencing by the BAH-EMF1 complex in plants. *Nat. Genet.* **50**, 1254–1261 (2018).
156. Peng, L. et al. Structural analysis of the arabidopsis AL2-PAL and PRC1 complex provides mechanistic insight into active-to-repressive chromatin state switch. *J. Mol. Biol.* **430**, 4245–4259 (2018).
157. Baranello, L. et al. DNA break mapping reveals topoisomerase II activity genome-wide. *Int. J. Mol. Sci.* **15**, 13111–13122 (2014).
158. Gray, M. W. & Doolittle, W. F. Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* **46**, 1–42 (1982).
159. Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. & Woese, C. R. Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82**, 4443–4447 (1985).
160. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
161. Woese, C. R., Magrum, L. J. & Fox, G. E. Archaebacteria. *J. Mol. Evol.* **11**, 245–251 (1978).
162. Harish, A. What is an archaeon and are the Archaea really unique? *PeerJ* **6**, e5770 (2018).
163. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
New species of archaea are identified with proteins homologous to eukaryotic endomembrane system components, and a two-domain model is supported in which eukaryotes are rooted within archaea, in contrast to Nasir et al. (2008).
164. Da Cunha, V., Gaia, M., Nasir, A. & Forterre, P. Asgard archaea do not close the debate about the universal tree of life topology. *PLOS Genet.* **14**, e1007215 (2018).
165. Moreira, D. & Lopez-Garcia, P. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
166. Pittis, A. A. & Gabaldon, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
167. Filee, J. Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: the visible part of the iceberg? *Virology* **466–467**, 53–59 (2014).
168. Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, 1286–1297 (2014).
169. Ambrosi, C., Manzo, M. & Baubec, T. Dynamics and context-dependent roles of DNA methylation. *J. Mol. Biol.* **429**, 1459–1475 (2017).
170. Lyons, D. B. & Zilberman, D. DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. *eLife* **6**, e30674 (2017).
171. Kato, M., Miura, A., Bender, J., Jacobsen, S. E. & Kakutani, T. Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr. Biol.* **13**, 421–426 (2003).
172. Satyaki, P. R. & Gehring, M. DNA methylation and imprinting in plants: machinery and mechanisms. *Crit. Rev. Biochem. Mol. Biol.* **52**, 163–175 (2017).
173. Williams, B. P. & Gehring, M. Stable transgenerational epigenetic inheritance requires a DNA methylation-sensing circuit. *Nat. Commun.* **8**, 2124 (2017).
174. Rondelet, G., Dal Maso, T., Willems, L. & Wouters, J. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *J. Struct. Biol.* **194**, 357–367 (2016).
175. Otani, J. et al. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep.* **10**, 1235–1241 (2009).
176. Neri, F. et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).
177. Coleman-Derr, D. & Zilberman, D. Deposition of histone variant H2A. Z within gene bodies regulates responsive genes. *PLOS Genet.* **8**, e1002988 (2012).
178. Murphy, P. J., Wu, S. F., James, C. R., Wike, C. L. & Cairns, B. R. Placeholder nucleosomes underlie germline-to-embryo DNA methylation reprogramming. *Cell* **172**, 993–1006 (2018).
179. Havas, K. et al. Generation of superhelical torsion by ATP-dependent chromatin remodeling activities. *Cell* **103**, 1133–1142 (2000).
180. Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev.* **16**, 2120–2134 (2002).
181. Gamarra, N., Johnson, S. L., Trnka, M. J., Burlingame, A. L. & Narlikar, G. J. The nucleosomal acidic patch relieves auto-inhibition by the ISWI remodeler SNF2h. *eLife* **7**, e35322 (2018).
182. Yan, L., Wang, L., Tian, Y., Xia, X. & Chen, Z. Structure and regulation of the chromatin remodeller ISWI. *Nature* **540**, 466–469 (2016).
183. Sinha, K. K., Gross, J. D. & Narlikar, G. J. Distortion of histone octamer core promotes nucleosome mobilization by a chromatin remodeler. *Science* **355**, eaaa3761 (2017).
Deformation of the histone octamer is shown to be required for remodelling activity of the SNF2h remodeller but not INO80, and it reduces nucleosome eviction by RSC. The authors propose that altered nucleosome conformations may be relevant in other processes such as eviction by pioneer factors.
184. Chaban, Y. et al. Structure of a RSC-nucleosome complex and insights into chromatin remodeling. *Nat. Struct. Mol. Biol.* **15**, 1272–1277 (2008).
185. Yelagandula, R. et al. The histone variant H2A. W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* **158**, 98–109 (2014).
186. Lorkovic, Z. J. et al. Compartmentalization of DNA damage response between heterochromatin and euchromatin is mediated by distinct H2A histone variants. *Curr. Biol.* **27**, 1192–1199 (2017).
187. Rivera-Casas, C., Gonzalez-Romero, R., Cheema, M. S., Ausio, J. & Eirin-Lopez, J. M. The characterization of macroH2A beyond vertebrates supports an ancestral origin and conserved role for histone variants in chromatin. *Epigenetics* **11**, 415–425 (2016).
188. Chakravarthy, S., Patel, A. & Bowman, G. D. The basic linker of macroH2A stabilizes DNA at the entry/exit site of the nucleosome. *Nucleic Acids Res.* **40**, 8285–8295 (2012).
189. Muthurajan, U. M., McBryant, S. J., Lu, X., Hansen, J. C. & Luger, K. The linker region of macroH2A promotes self-association of nucleosomal arrays. *J. Biol. Chem.* **286**, 23852–23864 (2011).
190. Bowerman, S. & Wereszczynski, J. Effects of macroH2A and H2A.Z on nucleosome dynamics as elucidated by molecular dynamics simulations. *Biophys. J.* **110**, 327–337 (2016).
191. Doyen, C. M. et al. Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Mol. Cell. Biol.* **26**, 1156–1164 (2006).
192. Chang, E. Y. et al. MacroH2A allows ATP-dependent chromatin remodeling by SWI/SNF and ACF complexes but specifically reduces recruitment of SWI/SNF. *Biochemistry* **47**, 13726–13732 (2008).
193. Chadwick, B. P., Valley, C. M. & Willard, H. F. Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome. *Nucleic Acids Res.* **29**, 2699–2705 (2001).
194. Lavigne, M. D. et al. Composite macroH2A/NRF-1 nucleosomes suppress noise and generate robustness in gene expression. *Cell Rep.* **11**, 1090–1101 (2015).
195. Pliatska, M., Kapasa, M., Kokkalis, A., Polyzos, A. & Thanos, D. The histone variant macroH2A blocks cellular reprogramming by inhibiting mesenchymal-to-epithelial transition. *Mol. Cell. Biol.* **38**, e00669–17 (2018).
196. Molaro, A., Young, J. M. & Malik, H. S. Evolutionary origins and diversification of testis-specific short histone H2A variants in mammals. *Genome Res.* **28**, 460–473 (2018).
197. Soboleva, T. A. et al. A new link between transcriptional initiation and pre-mRNA splicing: the RNA binding histone variant H2A. B. *PLOS Genet.* **13**, e1006633 (2017).
198. Barral, S. et al. Histone variant H2A. L.2 guides transition protein-dependent protamine assembly in male germ cells. *Mol. Cell* **66**, 89–101 (2017).
199. Contrepois, K. et al. Histone variant H2A. J accumulates in senescent cells and promotes inflammatory gene expression. *Nat. Commun.* **8**, 14995 (2017).
200. Ponger, L. & Li, W. H. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol. Biol. Evol.* **22**, 1119–1128 (2005).

Acknowledgements
The authors thank S. Brahma, K. Ahmad and the reviewers for helpful suggestions on the manuscript.

Author contributions
P.B.T. and M.P.M. researched content for the article. All authors contributed to the discussion of content, writing the article and reviewing and editing the manuscript before submission.

Competing interests
The authors declare no competing interests.

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

Reviewer information
Nature Reviews Genetics thanks K. Struhl and the other, anonymous reviewer(s) for their contribution to the peer review of this work.