Chromatin regulation at the frontier of synthetic biology

Albert J. Keung1,2, J. Keith Joung3,4, Ahmad S. Khalil1,5 and James J. Collins2,5,6

Abstract | As synthetic biology approaches are extended to diverse applications throughout medicine, biotechnology and basic biological research, there is an increasing need to engineer yeast, plant and mammalian cells. Eukaryotic genomes are regulated by the diverse biochemical and biophysical states of chromatin, which brings distinct challenges, as well as opportunities, over applications in bacteria. Recent synthetic approaches, including ‘epigenome editing’, have allowed the direct and functional dissection of many aspects of physiological chromatin regulation. These studies lay the foundation for biomedical and biotechnological engineering applications that could take advantage of the unique combinatorial and spatiotemporal layers of chromatin regulation to create synthetic systems of unprecedented sophistication.

Synthetic biology provides a powerful framework to understand and harness biology. Biological components are assembled into well-controlled systems, enabling the systematic study of emergent properties and complex behaviours. Synthetic components and systems and the regulatory behaviours they encode can then be adapted by engineers for numerous applications — for example, controlling the dynamic expression of biosynthetic genes in industrial organisms, engineering sensors of environmental state and creating therapeutically relevant cell types. The core components that synthetic biologists have come to rely on, particularly transcriptional repressors and inducible promoters, were assembled in bacteria by early pioneers into genetic networks with switching1 and oscillating behaviours2, and these components have since been applied to engineer bacterial cells that can sense and ‘remember’ the presence of antibiotics in the mouse gut3. Importantly, by engineering connections between basic genetic units, these pioneering studies demonstrated the functional power of synthetic systems to directly test hypotheses about how complex regulatory behaviours arise and to create useful cellular devices.

As we confront new challenges in biology, medicine and biotechnology, there are great opportunities to apply synthetic biology approaches to higher-order organisms such as yeast, plants and mammals. Indeed, many of the synthetic components and gene networks developed in bacterial systems have been demonstrated to have utility in eukaryotic systems4. However, it is becoming increasingly clear that these types of regulatory systems alone are unlikely to drive and recapitulate the biological complexity of eukaryotic organisms. Eukaryotic genes are regulated in fundamentally different ways from bacterial genes5. A central distinguishing feature is the packaging of eukaryotic DNA into chromatin. Chromatin underlies the greater complexity of eukaryotic gene regulation and has been implicated in a broad range of industrially and biomedically relevant behaviours, including cellular responses to environmental stresses, cancer and stem cell differentiation6–12. More than a decade ago, synthetic biology provided a functional approach to test hypotheses surrounding genetic networks. It now has bright prospects for functionally testing and expanding our understanding of chromatin and harnessing its diverse roles in cellular regulation.

Chromatin is a constellation of DNA, proteins and RNA components that exists in diverse biochemical and conformational states (Fig. 1). Genomic DNA is wound around octamers of histone proteins (Fig. 1b), called nucleosomes, which are arrayed to form the ‘bead-on-a-string’ backbone of chromatin (Fig. 1c). Nucleosomes can be biochemically modified (Fig. 1a) and spatially positioned on DNA, which itself can be methylated13, by the actions of hundreds of chromatin-modifying proteins. Nucleosomes also provide binding surfaces for these modifying proteins, as well as for transcription factors and nucleic acid polymerases14. Furthermore, nucleosomes alter the affinity of transcription factors for the underlying DNA through steric interactions15,16.
In fact, there is intimate crosstalk between transcription factors and the chromatin state, and the genomic locations of one are a significant predictor of the other\(^\text{17}\). At larger scales (FIG. 1d), regions of chromatin on the same or different chromosomes can interact with each other and with subnuclear structures, thus sharing regulatory factors that can influence and coordinate gene expression\(^\text{26}\). Much of our understanding of chromatin comes from decades of work in molecular cell biology and biophysics, and it has recently accelerated through the generation of many genome-wide maps of chromatin components and structure\(^\text{19,20}\). This work has revealed a rich realm of spatiotemporal and combinatorial regulatory potential\(^\text{19,21}\). Synthetic approaches are providing functional complements to these high-resolution maps and revealing the roles of different chromatin features in generating complex regulatory behaviours. Moreover, by expanding our understanding of chromatin and ability to functionally manipulate it, synthetic systems could soon be engineered for a wide variety of human health and industrial applications.

This Review discusses recent progress at the intersection of chromatin biology and synthetic biology, highlighting synthetic approaches to study and control three major features of chromatin structure: biochemical chromatin modifications, including their contribution to heritable and stable gene silencing through heterochromatin; nucleosome positioning; and spatial and topological conformations. We also consider the regulatory potential of non-coding RNAs in BOX 1. We focus on how fundamental regulatory motifs are revealed by the synthetic manipulation of chromatin. Native chromatin biology has been reviewed in greater depth and detail elsewhere\(^\text{21,22}\). We conclude by discussing potential applications in medicine and biotechnology, as well as new technologies that may accelerate future work in synthetic chromatin biology.

**Editing chromatin biochemistry**

In 1964, Vincent Allfrey and colleagues radioactively labelled histones in vitro with \(^{14}\text{C}\)-acetyl and \(^{14}\text{C}\)-methyl groups, which provided evidence for the
post-translational modification of histones. They went on to hypothesize that lysine acetylation reduces the electrostatic attraction of positively charged histones to negatively charged DNA: “as a charge neutralization mechanism, acetylation of the histones would be expected to modify DNA–histone interactions, and this may offer a molecular basis for the pronounced changes in histone acetylation and RNA synthesis during the course of gene activation” (REF. 24). This was an exciting hypothesis not only because of the mechanism it posited but also because it advocated a functional relationship between chromatin structure and gene expression. We now know that there are more than a dozen distinct types of histone modifications, more than 60 modified amino acid residues of histones, 2 of each histone type per nucleosome, and several different forms of DNA methylation. This combinatorial diversity creates a challenging problem for studying the individual and combinatorial roles of histone modifications; yet, it could underlie a powerful approach to synthetically set and tune expression states of synthetic genes without altering their DNA sequence. Thus, understanding the roles of these chromatin modifications has been a central goal in the past 50 years and will continue to be so in the foreseeable future.

Is there a histone code? An attractive hypothesis called the ‘histone code’ posits that particular combinations of histone modifications are ‘read’ by proteins to influence downstream functions. There have been impressive efforts that mapped the genome-wide locations of chromatin modifications and correlated them to transcriptional states. In addition, common experimental perturbations knock out or pharmacologically inhibit histone-modifying proteins to alter chromatin state. However, correlation and pleiotropy resulting from genome-wide perturbations often confound these approaches. Thus, for many specific histone modifications, it remains unclear whether they cause, result from or are unrelated to transcriptional activity. Two complementary synthetic approaches that tackle these challenges are the generation of synthetic histone proteins to study the effects of modifications on specific amino acid residues, and the biochemical modification of histones.
of nucleosomes at specific genomic locations in cells. In addressing the histone code hypothesis, these minimal synthetic approaches may also provide a ‘blueprint’ for assembling chromatin components to derive useful regulatory logic and behaviours from the vast combinatorial space of histone modifications.

**Residue-specific modifications in synthetic histones.**

Synthetic histones have been extremely powerful tools to control the specific type and residue of histone modifications (FIG. 2A). The most comprehensive approaches were derived from chemical biology and unnatural amino acid techniques, and are able to mimic almost all modification types at any histone residue49. Chemical ligation of peptides with specific post-translational modifications to recombinant histones provides control over the modification type and the modified residue49 (FIG. 2Aa). These histones can be reconstituted into nucleosomes for in vitro studies49. A high-throughput method was recently developed to synthesize 54 chemically defined nucleosomes with distinct combinations of post-translational modifications50. These nucleosomes were incubated in vitro with histone-binding and histone-modifying proteins to assess the effects of histone modifications on the binding affinity of these proteins and on their catalytic activity towards other histone residues. Binding of the nucleosomes to barcoded DNA sequences followed by next-generation sequencing allowed all the different synthetic nucleosomes to be pooled for each assay. This increased throughput to thousands of experimental samples. Unnatural amino acids also provide a strategy to control the modification of specific histone residues50,54 (FIG. 2Ab). For example, Nguyen and colleagues replaced the histone H3 lysine 9 (H3K9) residue with unnatural amino acids that could be chemically deprotected in vitro to reveal monomethyl51 and dimethyl52 lysines, and they used these recombinant histones to study the binding of heterochromatin protein 1 (HP1).

Although chemical ligation and artificial amino acid strategies generate highly homogeneous histones, they are difficult to implement in cells, but this remains a tantalizing possibility49. Instead, most methods in cells have relied on the generation of residue-specific histone mutants or the partial deletion of histone tails48,49 (FIG. 2Ac). In one of the most comprehensive studies so far, Dai et al.48 used gene synthesis to create a library of 486 *Saccharomyces cerevisiae* mutants of histones H3 and H4. They substituted every amino acid in the complete histone proteins with alanine and every alanine with serine, and further substituted all lysine residues with arginine or glutamine to mimic the electrostatic charges of constitutively deacetylated or acetylated states, respectively. They also deleted segments of varying lengths from the amino-terminal tails of histones H3 and H4. By DNA barcoding each mutant, the yeast strains could be efficiently screened for a range of phenotypes, including fitness, sensitivity to heat shock, sensitivity to DNA damage, altered transcriptional elongation and altered transcriptional silencing. Fitness-related genetic interactions were also identified between specific histone mutants and Ubp8 and Set2, which regulate H2BK123 ubiquitylation and H3K36 methylation, respectively. In the future, this type of high-throughput approach could be extended to identify additional interdependencies between histone residues and chromatin-regulating proteins. It is also possible that particular histone modifications have different and potentially counterproductive effects on industrially relevant cellular properties, such as fitness, heat shock response and ethanol tolerance. Libraries of industrial yeast strains expressing mutant histones could be phenotypically screened to identify biosynthetic strains with desirable characteristics. Thus, residue-specific modifications in synthetic histones provide an important avenue both to study the combinatorial complexity of the histone code and to generate useful cellular properties and industrial cell strains.

**Site-specific chromatin modifications: evolving technologies.**

Synthetic histones provide unparalleled control over the types and residues of post-translational modifications. However, they are limited to in vitro studies or to cellular studies of effects globally across the genome. As a complementary strategy, chromatin modifiers can be recruited to a particular genomic locus; such site-specific ‘epigenome editing’ approaches help to avoid pleiotropic effects and provide evidence of whether chromatin modifications directly cause changes in gene expression (FIG. 2B). This approach also has broader functionality in terms of its ability to recruit activities besides histone modifications, including DNA methylation and nucleosome remodelling. Furthermore, site-specific control will be necessary to regulate specific genes or genomic regions for biomedical41 and biotechnological applications. The origin of synthetic site-specific regulators can be traced to early work in synthetic yeast transcription factors, in which the DNA-binding bacterial repressor LexA was fused to the transcriptional activator Gal4 (REF. 42). Since then, many chromatin modifiers have also been fused to DNA-binding domains (DBDs), as reviewed elsewhere42 (FIG. 2Ba).

More recently, DBDs that can be programmed to bind to almost any DNA sequence have been fused to chromatin modifiers (FIG. 2Ba). Programmable DBDs, such as zinc-finger (ZF) and transcription activator-like effector (TALE) proteins44, are important for several reasons. They enable modification of chromatin at essentially any endogenous locus without the need to insert synthetic DNA sequences, which is a key advancement over the use of artificial episomal or integrated DNA sequences (FIG. 2Bb). For example, endogenous genes were activated by ZF–Tet and TALE–Tet fusion proteins (which induce active demethylation of methylated cytosine)45,46 and were suppressed by ZF–DNA methyltransferases47–49 and ZF–H3K9 methyltransferases50,51. These studies provide strong evidence that DNA and H3K9 methylation are causative of gene repression, especially with the use of minimal H3K9 methyltransferase catalytic domains and catalytic mutants that failed to induce methylation and gene repression41. Endogenous targeting was also used to identify natural targets of enhancer regions52 (FIG. 2Bc).
Figure 2 | Synthetic control of biochemical chromatin modifications. Synthetic approaches provide two types of specificity: specificity in the histone residues that are modified (part A) and specificity in the genomic locations of histone modifications (part B). Aa | Synthetic histones can be created by ligating chemically defined peptides (red) to the amino termini of partial histone proteins that are produced recombinantly. Ab | Artificial amino acids that are already biochemically modified can be incorporated into specific residues of histones that are recombinantly produced in bacteria. Ac | Genetic mutations can change specific residues into ones that partially mimic the charge and shape of a modified histone residue. Histones that are chemically synthesized or that contain artificial amino acids can be used in in vitro assays to measure changes in binding affinities to regulatory proteins or altered reactivity of specific histone residues due to the presence of modifications on other residues. Histones that were genetically mutated can be used in cellular assays to measure the effects of specific residues on global gene expression profiles. Bb–c | The fusion proteins bind to specific locations in the genome, altering gene expression of a downstream gene or genes that are targeted by a bound enhancer. Bd | Chromatin modifiers can also be dynamically recruited to a genomic locus, allowing temporal measurements of changes in chromatin modifications.
This was achieved by reducing the activities of enhancers with TALE–LSD1 (also known as KDM1A) fusions that demethylated H3K4, and identifying native targets of the enhancer by observing the resulting downregulation of proximal genes. Programmable DBDs also enable multiplex and combinatorial targeting, which has proved useful in several cases. For example, using ZFs, we recently co-recruited a library of 223 diverse chromatin-regulating proteins together with the commonly used transcriptional activator VP16, which provided a set of distinct, two-input gene expression logic. Other groups have used multiple ZFs to increase the site specificity of synthetic DNA methylation by fusing two co-dependent parts of a DNA methyltransferase to two distinct ZFs that bind to adjacent genomic sites. Future work is likely to take advantage of highly multiplexible CRISPR–Cas9 systems to increase the throughput and technical ease of introducing multiple site-specific chromatin modifications. Indeed, Cas9 could enable large-scale ‘rewiring’ or reprogramming of the chromatin network for research in stem cells, development, genomic stability and evolution. Furthermore, libraries of multiplexed site-specific chromatin modifiers could be used to tune the expression of biosynthetic or endogenous genes without the need to alter gene sequences, which enable rapid screening and engineering of production organisms.

Site-specific chromatin modifications: heterochromatin and stable gene repression. Site-specific recruitment of chromatin modifiers also enables mechanistic studies of other complex behaviours such as gene expression ‘memory’. Indeed, chromatin has been extensively implicated in the establishment and maintenance of gene expression memory, so much so that the term epigenetic is often co-opted to refer to chromatin modifications in general. However, by a stricter definition, epigenetic refers only to gene expression states that are heritable through cell divisions. This is a crucial point, as it is clear that chromatin modifications alone do not transmit heritable ‘epigenetic’ information. Instead, they must act in concert with RNA interference (RNAi) and protein factors through self-reinforcing feedback mechanisms to maintain memory. As reviewed elsewhere, much effort over the past several decades has focused on identifying these mechanisms and minimal components that give rise to heritable gene regulation, mainly in stable repressive regions called heterochromatin. Heterochromatin describes compact genomic regions that are often characterized by reduced histone acetylation, H3K9 tri-methylation (H3K9me3) enrichment and occupancy of repressive proteins such as HP1.

Synthetic approaches have contributed to this area of research by demonstrating the causative functions of various chromatin regulators in the establishment of heterochromatic silencing. In particular, synthetically recruited heterochromatin proteins can demonstrate the requirement or sufficiency of DNA sequence elements, RNAi components, histone and DNA modifications, and regulatory proteins for memory, and such proteins can even induce heritable repression. More recently, Ragunathan et al. showed that histone modifications and regulatory proteins are sufficient to establish stable heterochromatin in Schizosaccharomyces pombe, and that DNA sequence elements and RNAi are dispensable. They did so by transiently recruiting the S. pombe H3K9 methyltransferase Clr4 to a specific genomic location. This established a large domain of repressive H3K9me3 that was maintained over several mitotic and meiotic cell divisions in the absence of any DNA sequence elements, RNAi, DNA methylation or recruited Clr4. This supports a mechanism in which the histone modification, H3K9me3, and proteins, including Clr4, that both read and ‘write’ H3K9me3 are able to propagate histone modifications over time and maintain heterochromatic memory. Interestingly, proteins that catalyse and remove H3K9me3 marks regulated the stability of this heterochromatic region, suggesting a potential way to tune this form of memory in engineering applications.

Recently, Hathaway et al. developed an inducible system that dynamically regulated (Fig. 2Bd) the recruitment of HP1 and the strong synthetic transcriptional activator VP64 to a reporter locus in mouse embryonic stem cells. The associations of HP1 with the synthetic ZFHD1 DBD and of VP64 with the Gal4 DBD were controlled by the inducible dimerization domains of the FKBP–FRB system and the PYL–ABI system, respectively. This approach provided unprecedented temporal resolution of the kinetics of heterochromatin formation and reactivation and, through an accompanying quantitative model, insights into molecular mechanisms of heterochromatin establishment. Notably, the authors showed that the recruited HP1 protein could induce 10-kb regions of heterochromatic H3K9me3 after 5 days, and that this region was stable for at least 8 days after HP1 recruitment was washed out. Their model and data showed that the stability and boundaries of the heterochromatic region were determined by a dynamic competition between placement of H3K9me3 marks and nucleosome turnover. An increase in DNA methylation was also measured; based on previous work that linked DNA methylation to stable heterochromatin, this probably contributed to the observed gene repression memory. The continued development of this type of synthetic approach will be of particular interest to both chromatin biologists and synthetic biologists because memory and switch-like elements are of tremendous importance in cell biology, as well as in biomedical and industrial applications.

Site-specific chromatin modifications: ongoing challenges. Future work in site-specific chromatin modification should address the substrate specificity of synthetic chromatin modifiers. For example, a histone acetyltransferase may acetylate many lysine residues on a histone and recruit other regulatory proteins through binding interactions. These potential nonspecific modifications and interactions need to be fully characterized. The use of catalytically dead mutant forms of chromatin modifiers could reveal their residue specificity, whereas the truncation of chromatin modifiers to minimal catalytic domains could reduce the recruitment of other regulatory proteins.
Defining the binding specificities of the various customizable DNA-targeting platforms is also of considerable interest in the field. Although gene regulation activities tend to be the strongest at targeted sites, ZFs, TALEs and Cas9 can all exhibit substantial off-target binding as observed by various methods, and even minor off-target activities could be oncogenic or toxic\(^{27-79}\). Some strategies to improve the specificity of binding (or of effector activity) may include obligate pairs of DNA-binding or effector domains\(^{54,55,80,81}\); weakening binding affinities to individual nucleosides while increasing the overall strength of the target DNA sequence\(^{62}\); and altering the length of DNA-binding arrays for ZFs and TALEs or the guide RNA for Cas9 targeting\(^{72,83,84}\).

Most work has focused on writing chromatin modifications, but complementary technologies for reading these modifications\(^{85}\) will be equally important if conditional feedback of basal or altered chromatin states is desired, for example, to inform downstream actions such as the writing of additional modifications. The ability to sense aberrant chromatin states in live cells could also be useful in diagnosing disease states such as cancer.

These chromatin-editing and chromatin-sensing technologies will require creative, new strategies to enhance control over substrate modifications and sensing in diverse chromatin contexts; address typically modest protein–nucleosome binding energies\(^{86}\); improve genomic site specificity; and stabilize chromatin-based memory. For example, to address weak binding energies, inspiration could be drawn from natural systems and proteins that bind to chromatin through multiple domains or increased avidity\(^{87}\).

In addition to these improvements, the prospects for site-specific chromatin modifiers are bright and are likely to involve the incorporation of diverse molecular engineering techniques that will provide enhanced spatiotemporal resolution. In one recent example, Konermann et al.\(^{88}\) fused chromatin modifiers and TALEs to the Cry2 and CIB1 protein domains, respectively. Illumination with blue light dimerized the Cry2 and CIB1 domains, and dynamically localized chromatin modifiers to the DNA-bound TALE protein. These and other future systems may also adapt interesting biological systems that exhibit useful regulatory behaviours. For example, lowly expressed or non-expressed development genes are derived from the interactions between these components through temporally dynamic genetic and biochemical networks. The chromatin scaffold provides a unique opportunity to engineer cellular systems that

Modulating nucleosome positioning

Beyond biochemical modifications, nucleosome positioning on DNA influences gene expression. Nucleosomes generally inhibit transcription by reducing accessibility of transcription factors and RNA polymerases to DNA\(^{90-95}\), thus creating a basally repressed state that contrasts with highly accessible bacterial DNA\(^3\). Both nucleosome-remodelling proteins and the underlying DNA sequence control nucleosome positioning\(^{96-99}\). Owing to the ease of manipulating genetic sequences, synthetic approaches have focused on the manipulation of DNA sequences to control nucleosome positioning and gene expression. For example, Ravesh-Sadka et al.\(^{100}\) created a library of more than 70 promoter variants, each with different positioning and lengths of a (poly(dAdT)) tract that is known to disfavor nucleosome binding. Using this system, the positioning of nucleosomes regulated the accessibility of transcription factors to DNA. This strategy provided a dynamic range of more than an order of magnitude in transcriptional activity. This highly tunable approach to design the strength of synthetic promoters could be useful in situations for which balancing the expression strength of different components is important for its behaviour, such as in genetic oscillators\(^7\).

In addition to controlling transcription levels, the positioning of nucleosomes can drive complex nonlinear behaviours. For example, Lam et al.\(^{101}\) showed that, in yeast, nucleosomes can decouple the activation threshold of the PHO5 promoter from its dynamic range of expression. The activation threshold is controlled by the DNA-binding affinity of a protein that moves a nucleosome which originally blocks access of transcription factors to the promoter. When this nucleosome remodelling event is achieved, transcription factor binding sites uncovered in the promoter then determine the dynamic range. This type of regulation is reminiscent of a transistor in electrical engineering, in which the application of one signal allows the passage of a second signal, and could be used as a biological switch. Furthermore, the signal or protein expression level required to move the nucleosome and activate the PHO5 promoter could be engineered to be weaker than the protein expression level driven by the activated PHO5 promoter. This would essentially act as a signal amplifier that could be useful in circuit and cellular engineering. Future efforts could synthetically reposition nucleosomes by recruiting nucleosome remodellers (for example, SWI/SNF) to test other mechanisms that have been hypothesized to underlie complex behaviours, including the cooperative and switch-like activation or amplification of some genes\(^{102}\). These types of experiments demonstrate how nucleosome formation and repositioning could contribute to sophisticated and potentially useful gene expression programmes.

Spatial features of chromatin

Synthetic biology has primarily abstracted genomic material as a repository of regulatory components and genes that produce proteins and RNA. Complex behaviours are derived from the interactions between these components through temporally dynamic genetic and biochemical networks. The chromatin scaffold provides a unique opportunity to engineer cellular systems that
Nuclear lamina
A dense fibrillar network of intermediate filament proteins at the periphery of the nucleus.

Telomeres
The regions at the ends of chromosomes comprised of repetitive nucleotide sequences that are typically repressed by heterochromatin.

Auxotrophic markers
Genes absent in an organism that normally produce organic compounds required for survival of the organism.

Yeast knockout library
A collection of yeast strains, each of which harbour a knockout allele for a single gene. Strains are either haploid and have a non-essential gene knocked out, or diploid and have the knockout allele in a heterozygous state.

also function in spatial dimensions. The repetitive array of nucleosomes on DNA and the polymeric nature of chromatin provide the physical properties needed to regulate genes at both small and large spatial scales in the nucleus. For example, domains of DNA and histone modifications (such as H3K9me3) often cover multiple tandem genes. At larger spatial scales, genomic loci interact within and between chromosomes and also localize to subnuclear structures such as the nuclear lamina. Synthetic approaches that control these spatial features are deepening our understanding of chromatin biology and revealing emergent behaviours that could be useful in cellular engineering (FIG. 3).

Controlling boundaries of genomic activity. The juxtaposition of sequence elements (DNA) and non-sequence elements (nucleosomes and DNA methylation) is an intriguing property of chromatin with profound and broad implications for gene regulation. For example, nucleosomes and methylated DNA bases provide a repetitive scaffold to which regulatory proteins can bind and propagate over large spatial regions that contain multiple genes. However, these regions and their boundaries must ultimately be controlled by DNA sequence elements. Synthetic tools have provided insights into this crosstalk between sequence and non-sequence elements. In particular, many studies have focused on how the boundaries of large regions of silent heterochromatin, which are often marked by DNA methylation and H3K9 methylation, are specified. It is thought that these regions are established by self-reinforcing mechanisms in which protein complexes bind to heterochromatic modifications and then catalyse the same modifications on adjacent nucleosomes or DNA.

Synthetic approaches have revealed various protein and sequence elements that can control the spreading of heterochromatic regions. Telomeres have been used as a natural source of heterochromatin, with auxotrophic markers inserted nearby as sensors of heterochromatic spreading. Using similar systems, various sequence elements have been inserted between the markers and telomeres, and identified as barriers to heterochromatic spreading. These include nucleosome-disrupting sequences, transcriptionally active tRNA genes, binding sequences for synthetic transcriptional activators such as SAS2 fused to the DBD of GAL4 (REF. 101), and sequences tethered to the nuclear pore complex (REF. 102).

There are also boundaries of transcriptional activity within individual genes that can and need to be controlled. For example, most promoters have a limited extent of divergent transcription of non-coding sequences. Marquadt et al.103 integrated a synthetic bidirectional promoter, which drives divergent transcription of two distinct fluorescent proteins, into a yeast knockout library and found that divergent transcription increased in CAF-I-knockout strains. As CAF-I regulates chromatin assembly, nucleosome positioning upstream of promoters probably limits divergent non-coding transcription. As more sophisticated synthetic gene expression systems are developed, both intragenic and intergenic boundary elements — such as CAF-I and nucleosome-disrupting sequences, respectively — will need to be used to avoid unintended spatial regulation.

Chromatin conformation and subnuclear interactions. Chromosomes exist in semi-restricted spatial domains within the nucleus called chromosome territories. In addition, the positioning of genomic loci within each

Figure 3 | Synthetic spatial control of chromatin. A | Several different types of barrier elements (red) can block the spreading of repressive heterochromatin. These include recruitment of transcriptional activators101, tethering to a nuclear pore complex102, insertion of actively transcribed tRNA genes or insertion of nucleosome-disfavouring DNA sequences95,100. B | The recruitment of a self-dimerizing protein can induce chromatin looping and localization of an enhancer or a locus control region (LCR) to a promoter102. This could localize factors that phosphorylate (P) and activate RNA polymerase II (Pol II). C | DNA-binding proteins can be used to answer biological questions, for example, whether the spindle checkpoint senses displacement between sister chromatids or within kinetochores116 (part Ca). Fusing DNA-binding domains to nuclear envelope proteins can localize (and often silence) specific genomic regions117 (part Cb). Part C adapted from REF. 116.
chromosome is nonrandom, as is their positioning relative to subnuclear structures such as the nuclear lamina. However, it is unclear whether these forms of spatial organization are just by-products of evolution or whether they have evolved for specific dynamic regulatory functions. Recent synthetic approaches have recapitulated some of these interactions, demonstrating that at least in some synthetic and physiological situations long-range interactions and chromatin conformation exhibit function.

Synthetic approaches have demonstrated that transcriptional effectors recruited to sites more than 1 kb downstream of promoters can both activate and repress yeast genes through intrachromosomal looping. There is also evidence that looping can regulate genes at much greater distances. One prominent example is the β-globin gene that is regulated by its locus control region (LCR) positioned 40 kb away. Deng et al. demonstrated that this looping was causative of gene activation. They designed a ZF that binds to the β-globin promoter and fused it to Ldb1, a protein that is present at the LCR. The synthetic tethering of Ldb1 to the β-globin promoter was enough to recruit the LCR region through Ldb1–Ldb1 binding and resulted in activation of β-globin transcription. The authors also used the same strategy to reactivate the developmentally silenced fetal γ-globin promoter in primary adult murine erythroblasts. This strategy could yield therapeutic benefits in patients with sickle cell anemia or β-thalassemia by using gene therapy approaches to introduce looping factors into cells, thus inducing transcription of non-sickling γ-globin to replace defective adult β-globin. Such potential therapeutics could be within reach in a decade or two, as there has been a recent resurgence in clinical trials of gene therapies. This looping approach has several advantages, including the requirement of only a single ZF–Ldb1 factor for introduction into cells. Furthermore, the looping approach activated γ-globin expression by more than 1,000-fold, which is much stronger than the levels typically observed with synthetic activators. Thus, this strategy could lead to therapies that are more potent than the simple recruitment of transcriptional activators directly to the γ-globin promoter or the introduction of exogenous globin genes.

Chromatin–chromatin interactions also drive complex functions beyond gene activation. Noordermeer et al. tested whether the insertion of an artificial LCR onto a separate chromosome from the β-globin gene could activate β-globin transcription. They observed that interchromosomal interactions between the artificial LCR and the native β-globin gene were able to activate transcription, although this activation occurred only in rare ‘jackpot’ cells. The authors hypothesized that rare long-range interactions might be a way for cells to stochastically generate phenotypic diversity and described this effect as ‘spatial effect variegation’. In addition to activating individual genes such as β-globin, chromatin–chromatin interactions can also temporally coordinate the expression of multiple genes, which could be useful in engineering the dynamic and ordered expression of different components in biosynthetic pathways.

For example, active and inactive regions of the human genome tend to cluster together through predominantly intrachromosomal, as well as interchromosomal, contacts. It is thought that regulatory factors, either activating or repressive, are shared between spatially clustered genes, thus coordinating transcription regulation. Fanucchi et al. revealed an example of this coordination between the SAMD4A, TNFAIP2 and SLC6A5 genes, which are known to cluster in response to tumour necrosis factor-α (TNFα) exposure. They designed TALE nucleases to cleave each of the three loops that brought these genes together and measured changes in gene expression. They found that disrupting the chromatin contacts interfered with transcription of the other genes. Interestingly, this co-transcriptional effect was hierarchical. Some cleavage sites disrupted transcription of all three genes, whereas others affected only one or two genes.

In addition to intrachromatin interactions, chromatin also localizes with various subnuclear structures. Some subnuclear structures can be induced to form on chromatin by the tethering of specific proteins. For example, artificial kinetochores or centromeres can be created by the synthetic recruitment of specific proteins to DNA. When Ask1, a member of the Dam1–DASH microtube complex in S. cerevisiae, was fused to LacI and recruited to eight copies of genomically integrated lacO operator sites, it was able to trigger assembly of a synthetic kinetochore. This kinetochore carried out many native functions, including bi-orienting sister kinetochores at metaphase, segregating sister chromatids and repairing errors in chromosome attachment. In another example, synthetic recruitment of the Clr4 H3K9 methyltransferase to a euchromatic region in S. pombe promoted de novo centromere formation. Synthetic control over centromere and kinetochore formation, coupled with knockouts of different centromere-related proteins, could reveal different roles of proteins in centromere establishment and function, including in regulating the fidelity and stability of the chromosome. Furthermore, the ability to engineer stable synthetic centromeres in synthetic chromosomes could be useful in expressing large cassettes that contain entire synthetic circuits, full biosynthetic pathways or large therapeutic genes that would be difficult to package in viruses for gene delivery (such as titin for various myopathies).

Synthetic tools can also be used to control the biochemical parameters of subnuclear structures, such as centromeres and kinetochores, to answer fundamental biological questions. In particular, it had been unclear how, during mitosis, the spindle checkpoint senses that paired chromosomes are attached to separate spindles. The general hypothesis was that when chromosomes were properly attached to separate spindles, the spindles would pull in opposite directions, which leads to a displacement either within kinetochore protein complexes or between the paired centromeres. To address this question, Nannas et al. synthetically restricted the ability of centromeres to stretch by integrating lacO sites on either end of the centromere of chromosome III in haploid S. cerevisiae. By expressing tetrameric
LacI complexes during mitosis, the outside ends of the paired centromeres were tethered together and restricted from moving apart. Tethering did not alter the spindle checkpoint or chromatid segregation, suggesting the checkpoint senses intrachromosomal kinetochore strain rather than interchromosomal displacement between paired centromeres

Chromatin also localizes to other subnuclear structures that do not function as part of chromosomes. Some of these interactions are correlated with transcriptional activity. For example, the nuclear lamina is generally associated with repressed genes, whereas the nuclear periphery is associated with active genes. Although the mechanisms underlying these correlations remain unclear, synthetic tethering of reporter genes to subnuclear structures has been shown to alter gene expression. In one example, Reddy et al. inserted an array of lacO operator sites downstream of a reporter gene. They also fused LacI to emerin, an inner nuclear membrane protein (FIG. 3Cb). Through interaction with LacI, the reporter was recruited to the nuclear membrane and silenced. In a related experimental system, Kind et al. fused the nuclear lamina protein lamin B1 to the adenine methyltransferase dam. When they expressed the methyl-adenine-binding protein dpl fused to the transcriptional activator VP16, the fusion protein was recruited and bound to methylated adenine bases located at the nuclear lamina. The dpl–VP16 fusion protein then induced delamination of lamina-associated chromatin domains from the nuclear periphery. The authors hypothesized that activating transcription in lamina-associated chromatin domains induced their movement away from the nuclear periphery, which is consistent with the trend of a silent nuclear periphery. Finally, in addition to controlling chromatin localization and conformations, synthetic DBDs such as LacI or Cas9 can also be used to recruit fluorescent proteins to observe the dynamic 3D structure of chromatin. In summary, many rich behaviours arise from the conformations and subnuclear interactions of chromatin, and existing and future synthetic tools that control the spatial aspects of chromatin may reveal even more unique regulatory properties.

A bright future for synthetic chromatin biology

More than a decade ago, systems and synthetic biology enhanced our understanding of gene networks. It is appealing to speculate how these fields may once again provide insights into the regulatory logic of chromatin and address the many open questions in chromatin and cell biology. The key advantages that synthetic approaches have provided are their ability to functionally perturb chromatin states with specificity in both histone residues and genomic locations. These attributes should continue to help researchers to determine the extent and validity of a histone code, as well as its functional role in the regulation of various molecular processes from gene expression activity and kinetics to DNA repair. Furthermore, although synthetic approaches often force the recruitment or manipulation of chromatin components in ways that are physiologically unnatural, their advantage is that they can identify the roles of context — for example, contexts in which regulation at one genomic site versus another or with specific combinations of components leads to distinct outcomes. Additionally, there has been recent and rapid development of technologies that expand our experimental capabilities, and it is exciting to speculate how the development and incorporation of these technologies could have substantial impacts on future research directions. For example, the synthetic construction of entire chromosomes could lend itself to studies that interrogate the relationships between regulatory DNA sequences, distances between genes and spatial chromosome conformation. Moreover, CRISPR–Cas9 technologies could enable tractable experiments that induce genome-wide rewiring of chromatin states and subsequent analyses of their effects on many biological processes, including cellular differentiation, proliferation and oncogenesis. Finally, optical tools could probe chromatin dynamics at unprecedented spatiotemporal resolution, thus revealing the kinetics of gene activation and repression, memory formation and the relative order of action of different components in executing biological processes such as transcription.

In addition to expanding our fundamental understanding of biology, an ongoing achievement of synthetic biologists is the abstraction of biology into modular components with defined properties, which greatly streamlines design cycles for cellular engineering and applications. An intriguing question that remains is whether an analogous design framework is possible at the chromatin level and, if so, what form this framework might take. The most effective synthetic approaches will continue to benefit from the wealth of knowledge and conceptual thinking developed by chromatin biologists over the past 50 years. In fact, our fundamental understanding of chromatin has already had a substantial impact in identifying regulatory properties that could be harnessed in artificially engineered cellular systems for bioindustrial and biomedical applications. For example, spatially coordinated gene expression and multigene silencing could be used by metabolic and microbial engineers to efficiently regulate large bio-synthetic cassettes in a manner that cannot be achieved by bacterial systems. The many barrier elements that have been identified could also be used to control these regions of long-range regulation and prevent the spreading of regulation to endogenous genes. The direct role of chromatin in gene expression memory has been synthetically recapitulated and could be used to record and respond to environmental events to which cells are subjected, such as exposure to antibiotics, heat shock or toxins, or changes in cell density in a bioreactor. Finally, given the many ways to now target, modulate and modify chromatin, these approaches are likely to be used to correct disease-associated chromatin states in cancer, neurodegenerative diseases and many other conditions, or to artificially alter chromatin changes that are involved in developmental and differentiation pathways. Given that several epigenetic therapeutics, typically small-molecule inhibitors of chromatin-modifying proteins, have already been used in the clinic, more-specific strategies derived from synthetic
Controlling biosynthetic pathways

Aa

Endogenous gene

Biosynthetic gene

Repressor

Barrier element

Repressive heterochromatin

Temporal control

Production OFF

Production ON

Degradation or dilution

Time

Bioreactor

Yeast cell expressing site-specific repressive chromatin fusion protein

Figure 4 | Potential applications of synthetic chromatin biology. 

A | Eukaryotic organisms such as yeast are widely used in industry to produce diverse molecules that range from food flavourings to pharmaceuticals. Often, multiple biosynthetic genes need to be introduced into an organism’s genome. Regulating many genes can be difficult, especially with many distinct promoters and other regulatory elements. Aa | In chromatin-based regulation, chromatin states such as heterochromatin can exert regulation over several kilobases of DNA through the simple recruitment of a repressor to a single location in the genome. Barrier elements can be used to prevent spreading of repressive regulation into endogenous genes. Ab | Repression of biosynthetic genes when cell densities are low promotes cell growth and fitness by avoiding taxing of cell resources. When cell densities in a bioreactor are optimal, the repressor can be degraded or diluted out, thus turning on expression of the biosynthetic genes. B | Chromatin has been widely implicated in many areas relevant to human health, including cell differentiation, cell proliferation, cell survival in different environmental conditions, epithelial-to-mesenchymal transition (EMT, which is relevant to cancer metastasis) and oncogenic potential. Functional tools are lacking to directly perturb chromatin states at specific locations in the genome and to test hypotheses of their roles in these biomedical processes. Type II CRISPR–Cas9 DNA targeting technologies could be used to edit chromatin states at multiple genomic locations by fusing chromatin regulators to the Cas9 protein and expressing a library of guide RNAs that are complementary to genomic target sites.

In conclusion, synthetic biology has great potential to advance two broad goals: first, in providing tools and approaches to complement studies that address fundamental questions in biology; second, in developing an engineering framework to harness chromatin components for cellular engineering applications. Our path to both goals will be accelerated by the continued collaboration between scientists and engineers. Given the broad importance of chromatin in most cellular processes and the many diseases with epigenetic components in their aetiology, collaborative efforts could have profound effects on our understanding of cell biology and our treatment of a vast number of diseases.
28. Pick, H., Kile, S. & Frizz, B. Engineering chromatin states: chemistry and synthetic biology approaches to investigate histone modification function. Biochemistry. Biol. 1839, 644–656 (2014).
29. He, S. et al. Facilitating catalytically dead acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. Proc. Natl Acad. Sci. U.S.A. 1038A (2003).
30. Shimko, J. C., Hovard, J. C., Pouirier, M. G. & Ottensen, J. J. Preparing semisynthetic and fully synthetic histones H4 and H5 to modify the nucleosome core. Methods Mol. Biol. 181, 177–192 (2013).
31. Nguyen, U. T. et al. Accelerated chromatin biochemistry using nucleosome libraries. Nature Methods 11, 834–840 (2014). By reconstituting chemically synthesized histones with barcoded DNA, this study was able to pool diverse histone–DNA complexes, thus greatly reducing the number of biochemical assays needed to explore a large combinatorial histone modification space.
32. Neumann, H. et al. A method for genetically installing site-specific acetylative in recombinant histones defines the effects of H5K6 acetylation. Mol. Cell 36, 155–163 (2009).
33. Davis, L. C. & Chin, J. W. Designer proteins: applications of genetic code expansion in biological and medicinal chemistry. Cell 130, 168–182 (2007).
34. Nguyen, D. P., Garcia Alai, M. M., Kapadis, P. B., Neumann, H. & Chin, J. W. Chemically synthesized Ne-k-tetrahydroacetylated lysine in recombinant histones. J. Am. Chem. Soc. 131, 14194–14195 (2009).
35. Nguyen, D. P., Garcia Alai, M. M., Virdie, S. C. & Chin, J. W. Chemical synthesis of a histone H3-ΔDNA-like mutant. Mol. Cell 51, 1072–1076 (2013).
36. Chin, J. W. et al. An expanded euacetylatable histone code. Science 301, 964–967 (2003).
37. Dion, M. F., Altschuler, S. J., Wu, L. F. & Rando, O. J. A site-specific acetyltransferase that reverses simple histone H4 acetylation code. Proc. Natl Acad. Sci. U.S.A. 102, 5501–5506 (2005).
38. Kim, J. A., Hutabarat, M. S. M. & Allfrey, V. G. Mutagenesis of pairwise combinations of histone amino-terminal tails reveals functional redundancy in budding yeast. Proc. Natl Acad. Sci. U.S.A. 109, 3779–3784 (2012).
39. Dai, J. et al. Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. Cell 150, 1066–1078 (2012).
40. Siddique, A. N. et al. Targeted methylation and gene silencing of VEGF-A in human cells by using a designed DemiSA-DoxA directed acetylase modifies histone modifications in human cell lines. Proc. Natl Acad. Sci. U.S.A. 111, 15357–15372 (2014).
41. Laci, V. et al. Nucleosome-encoding transgenic factor binding predicts histone modifications in human cell lines. J. Cell Sci. 117, 1272–1278 (2004).
42. Bogenhagen, D. F., Li, P., Liao, C., Carroll, D. J. & Grunstein, M. Nucleosome architecture in the nuclear space: regulation of gene expression in three dimensions. Nature Rev. Genet. 8, 104–115 (2007).
43. Dunham, I. et al. A very large-scale ciliated eukaryotic DNA element in the human genome. Nature 489, 57–74 (2012).
44. Ram, O. et al. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell 147, 1628–1639 (2011).
45. Zhou, W. V., Ge, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. Nature Rev. Genet. 12, 7–18 (2011).
46. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. Annu. Rev. Biochem. 78, 273–304 (2009).
47. Allen, V. C., Faulkner, R. & Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl Acad. Sci. U.S.A. 51, 767–784 (1964).
48. Vidali, G., Gershey, E. L. & Allfrey, V. G. Chemical studies of histone acetylation. The acetylation of ε-N-acetylcytosine in calf thymus histones. J. Biol. Chem. 243, 6361–6366 (1968).
49. Tan, M. et al. Identification of 67 histone marks and histone lysine acetylation as a new type of histone modification that regulates gene expression. Proc. Natl Acad. Sci. U.S.A. 108, 1704–1709 (2011).
50. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. Nature 403, 41–45 (2000).
51. Herron, D. T. & Allfrey, V. G. Histone modification: cause or cagy? Trends Genet. 27, 389–396 (2011).
52. Turner, B. M. The adjustable nucleosome: an epigenetic epigenetic module. Trends Genet. 26, 436–442 (2011).
11.Shen, E. et al. Nuclear aggregation of off-target recombinase proteins follows their monogenic expression. Cell 151, 724–737 (2012).

12.Lasica, S., Lao, D. T. & Murray, A. W. Recruiting a microRNA binding complex to DNA directs chromosome segregation in budding yeast. Nature Cell Biol. 11, 1116–1120 (2009).

13.Kapranov, P. et al. Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. Science 324, 1716–1719 (2009).

14.Nanav, N. J. & Murray, A. W. Tethering sister centromeres to each other suggests the spindle checkpoint detects stretch within the kinetochore. Pnas Genet. 10, e100442 (2014).

15.Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. Nature 452, 243–247 (2008).

16.Kind, J. et al. Single-cell dynamics of genome–nuclear lamina interactions. Cell 153, 178–192 (2013).

17.Aragon-Alcaide, L. & Strunnikov, A. V. Functional dissection of in vivo interchromosome association in Saccharomyces cerevisiae. Nature Cell Biol. 2, 812–818 (2000).

18.Chen, B. et al. Dynamic imaging of genic loci in living human cells by an optimized CRISPR–Cas system. Cell 155, 1479–1491 (2013).

19.Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. Nature 397, 168–171 (1999).

20.Barkai, N. & Leibler, S. Robustness in simple biochemical networks. Nature 387, 913–917 (2000).

21.Dymond, J. S. et al. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477, 471–476 (2011).

22.Sander, J. D. & Joung, J. K. CRISPR–Cas systems for editing, regulating and targeting genomes. Nature Biotechnol. 32, 347–355 (2014).

23.Tischer, D. & Wein, O. D. Illuminating cell signalling with optogenetic tools. Nature Rev. Mol. Cell Biol. 15, 551–558 (2014).

24.Siva, M. L., Riggio, N. & Bernstein, B. E. Epigenetic reprogramming in cancer. Science 339, 1567–1570 (2013).

25.Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. Nature 28, 1079–1088 (2010).

26.Buhler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-independent gene silencing. Cell 125, 873–886 (2006).

27.Jiang, J. et al. Translation dosage compensation to trimRNA2. Nature 500, 295–300 (2013).

28.By inserting the large coding RNA XIST into one of three copies of chromosome 21 in Down syndrome pluripotent stem cells, this study was able to completely silence the extra chromosome, suggesting a potential therapeutic avenue for many polydysplasia diseases.

29.Keller, C., Halaszegar-Szilyni, R., Shimada, Y., Hotz, H. R. & Buhler, M. Noncoding RNAs prevent spreading of a repressive histone mark. Nature Struct. Mol. Biol. 20, 1540 (2013).

Acknowledgements

This work was supported by a US National Institutes of Health (NIH) National Institute of General Medical Sciences (NIGMS) Ruth L. Kirschstein Postdoctoral Fellowship (A.J.K.), an NIH Director’s Pioneer Award (J.K.), a Defense Advanced Research Projects Agency grant (J.K.J., A.S.K. and J.J.C.), and start-up funds from the Department of Biomedical Engineering at Boston University (A.S.K.), a National Science Foundation CAREER Award (A.S.K.), a NIH R24 (J.J.C.), the Wyss Institute for Biologically Inspired Engineering (J.J.C.) and the Howard Hughes Medical Institute (J.J.C.). Owing to space limitations, the authors regret that many important publications were not able to be discussed in this Review.

Competing interests statement

The authors declare competing interests: see Web version for details.