Please note the following before checking your manuscript

Dear Professor Beck,

Please note:

- Queries may be nested in the addresses section or in the reference list, please look out for these.
- References that could be misread — those that follow numbers, superscripted and subscripted characters, quotation marks and abbreviated units — are set in brackets, on the line (for example, we use 10 (REF. 10) instead of 10^{10}).
- Please check reference numbering carefully, as they may have been renumbered (for example, to avoid duplications).
- When marking up your corrections, please use the Commenting tools only (do not edit the text directly).
- If your article has Supplementary Information but you have not received the file(s) along with this proof, please let us know as soon as possible.
- We do not usually edit Supplementary Information in house, so should you require changes to your SI, please send a revised file and indicate what changes have been made.

EPIGENETICS

From profiles to function in epigenomics

Stefan H. Stricker[1,2], Anna Köferle[3] and Stephan Beck[3]

Abstract | Myriads of epigenomic features have been comprehensively profiled in health and disease across cell types, tissues and individuals. Although current epigenomic approaches can infer function for chromatin marks through correlation, it remains challenging to establish which marks actually have causative roles in gene regulation and other processes. After revisiting how classical approaches have addressed this question in the past, we discuss the current state of epigenomic profiling and how functional information can be indirectly inferred. We also present new approaches that promise definitive functional answers, which are collectively referred to as 'epigenome editing'. In particular, we explore CRISPR-based technologies for single-locus and multi-locus manipulation. Finally, we discuss which level of function can be achieved with each approach and introduce emerging strategies for high-throughput progression from profiles to function.

Non-genetic factors contribute to many cellular functions, traits and phenotypes[1]. Among the first to conceptually recognize this was Conrad Hal Waddington who coined the term 'epigenetics' in 1942 to describe the molecular mechanisms "by which the genes of the genotype bring about phenotypic effects" (REF. 2). Captured by the iconic image of the epigenetic landscape (FIG. 1a), he imagined its mode of action to be 'causal' (REF. 2), similar to the presumed deterministic effect that a topographic shape has on the movement of a marble[3].

More than half a century later, we have come a long way in our understanding of the molecular basis of epigenetics and its role in cellular and organismal plasticity and dynamics. Several ground-breaking studies have revealed that alterations to chromatin — the nuclear complex of macromolecules that consists of DNA, protein (histones) and RNA — can in some cases account for changes in gene expression (for a selection of classic experiments concerning DNA methylation see BOX 1). For the purposes of this Review, we therefore define modifications of DNA and histones as alterations to chromatin but distinguish between chromatin marks (individual chemical modifications) and features (multiple linked modifications and more complex elements).

Catalogues of chromatin marks and features obtained from cells and tissues at different stages of development and disease states have become an extremely useful resource. Epigenomic profiling was the key to discovering many significant associations between chromatin features and genomic function at the level of gene regulation and expression, cell identity, age and even disease[4–6]. However, correlation does not necessarily imply causation, and technical limitations had not previously allowed the interrogation of individual marks or combinations of marks to test for direct functional effects. Consequently, the majority of research focused on identifying what Adrian Bird defined as the unifying definition of epigenetic events: "the structural adaptation of chromosomal regions" that may "register, signal or perpetuate altered activity states" (REF. 7). Epigenetic research is currently at a turning point. New approaches, which benefit from the remarkable developments in genome editing, enable us to move forward and to finally elucidate which individual chromatin marks or features have causal roles in processes such as gene regulation, cellular memory, cellular differentiation and disease aetiology[8].

The terms 'function' and 'functional' mean different things to different people and, in our view, are at times used incorrectly in the literature. For clarification and in the context of this Review, we therefore differentiate between two levels of function — inferred and causal — as illustrated in FIG. 2. Inferred function is usually based on the correlation of aggregated marks or features with observed effects, for example, gene activity states or phenotypes, but cannot establish whether marks have truly causal roles. By contrast, causal function is based on direct evidence of individual marks or features driving the expression of a particular gene or regulating a particular phenotype.

Throughout this Review, we emphasize the level of function that can or that has been demonstrated using different experimental approaches. We discuss what can be learned from comparative chromatin profiling, and how associations of chromatin marks with phenotypes can identify candidate regions for functional testing. Although we fully appreciate the importance of plant

[1]*Institute of Stem Cell Research, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, Neuherberg 85764, Germany.*
[2]*Biomedical Center, Ludwig-Maximilian University, Grosshaderner Strasse 9, Planegg-Martinsried 82152, Germany.*
[3]*University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6BT, UK.*

Correspondence to S.B.
s.beck@ucl.ac.uk

doi:10.1038/nrg.2016.138
Published online DD Mmm 2016

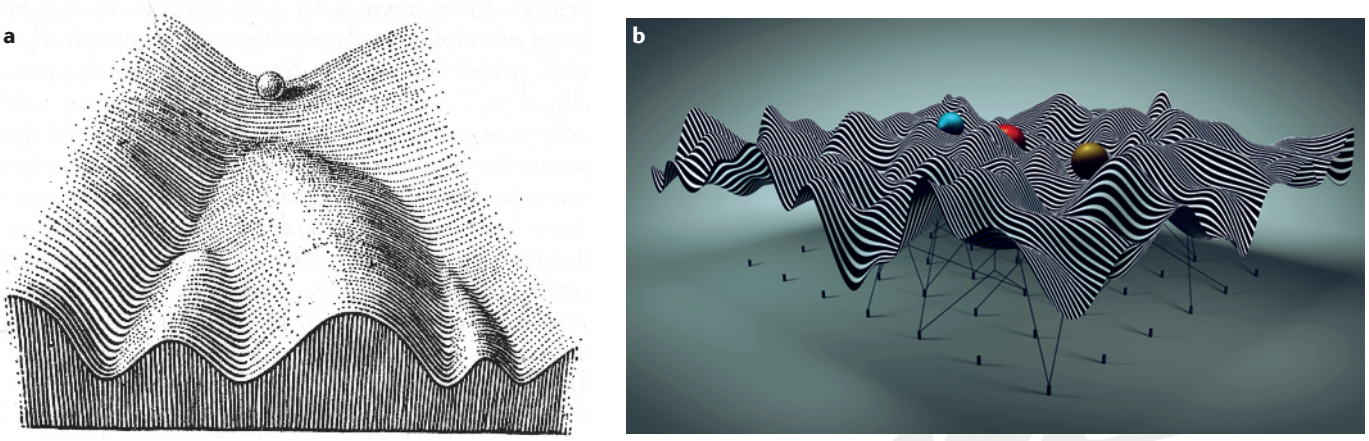


Figure 1 | **Evolving views on the Waddington landscape. a** | Epigenetic landscape as depicted by C. H. Waddington[3]. In this analogy, epigenetics influences cellular fate during development, analogously to gravitational forces on a defined landscape. **b** | A contemporary version of the Waddington landscape depicting epigenome editing. Epigenomic manipulation (represented by strings pulling the contours of the landscape) promises to dynamically change the landscape and thus cellular phenotypes. The contemporary image is derived from an animation by Paul Liam Harrison (see Supplementary information S1 (movie) for the animation and for further information on its generation). Part **a** is from *The Strategy of the Genes*, C. H. Waddington, ©1957 Allen & Unwin, reproduced by permission of Taylor & Francis Books UK. Part **b** image courtesy of P. L. Harrison, University of Dundee, UK, and EpiGeneSys. **[Au:OK?]**

epigenomics and associated resources[9–13], this Review focuses on mammalian chromatin. We briefly revisit insights gained from early knockout studies but mainly concentrate on recent epigenome-editing approaches[14] (sometimes also known as epigenetic engineering[15]), which directly test causality. A short overview of methods for epigenome editing will be provided. However, for a more detailed discussion of technological aspects, we refer the reader to excellent recent reviews on genome and epigenome editing[16–18]. Finally, we speculate how these approaches could be used to efficiently deduce causal function from profiles in the future. Although we recognize the importance of differential expression and the binding of transcription factors, nucleosome positioning and chromatin remodelers to gene regulation (which have previously been thoroughly reviewed)[19–21], we mostly focus on the contribution of chromatin marks to gene expression.

Epigenomic resources

Following the completion of the Human Genome Project[22] it became immediately evident that additional efforts would be required to understand how complex genomes are regulated. Driven by different technologies, new international resources (see Further information) were soon established to profile all aspects of the genome and epigenome that were thought to have functional relevance. Collectively, these resources have increased the amount of data per sample that we now have over and above the genome by several orders of magnitude.

Projects and data types. Beginning in 2003, ENCODE (Encyclopedia of DNA Elements) was the first international project to use large-scale epigenomic profiling to identify regulatory elements in the human genome. ENCODE pioneered many of the required technologies (for example, for profiling histone modifications) and focused on cell lines rather than tissues or primary cells. The project was subsequently expanded to include model organisms (modENCODE), adding the power of comparative epigenomics. ENCODE became a member of the International Human Epigenome Consortium (IHEC), a project that was launched in 2010 and which aims to generate 1,000 reference epigenomes in primary tissues and cell types and which has become the umbrella organization under which national and international epigenome efforts are jointly coordinated. IHEC currently has nine members, ENCODE (United States), Roadmap Epigenomics (United States), BLUEPRINT (European Union), DEEP (Germany), Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC; Canada) and the national epigenome projects from Japan, Korea, Singapore and Hong Kong (see Further information). Within the context of this Review, the key IHEC achievements so far have been the introduction and the implementation of the IHEC quality standards for epigenomic data and the IHEC data portal, which provides access to the data of all IHEC projects. At the time of writing, more than 7,000 data sets from more than 350 tissues and cell types were available, including 228 complete and 982 partially complete reference epigenomes. Based on highly successful pilot projects[23], these data will soon be complemented by a new international effort (4D Nucleome), which aims to produce three-dimensional maps of mammalian genomes and to develop predictive models to infer function from mammalian genome architecture. As discussed below, the integration of epigenomic features with genetic variation (for example, from the Catalogue of published genome-wide

Genome-wide association studies
(GWAS). Studies that aim to identify genetic loci associated with an observable trait, disease or condition.

Single-nucleotide polymorphisms
(SNPs). Single base-pair differences in the DNA sequence between individual members of a species.

DNase I-hypersensitive sites
(DHSs). Regions of chromatin that are sensitive to digestion with DNase I, indicating that these sites are accessible and free of nucleosomes.

association studies (GWAS Catalogue)), gene expression (for example, from the Genotype–Tissue Expression (GTEx) project) and other data is tremendously useful for pinpointing candidate variants for functional analysis. However, one limitation that currently hampers the comprehensive exploitation of these resources for functional and other analyses is that much of the raw data are not available under open access and require prior approval by a Data Access Committee (DAC).

The descriptive data types that can be obtained by epigenomic profiling are still growing in both numbers and complexity. BOX 2 illustrates this complexity of profiles and provides details of the individual marks and features profiled so far. These can be divided into the following categories: DNA modifications, which includes C5-methylcytosine (5mC), the first and most well-studied epigenomic modification that was discovered in 1948 (REF. 24), as well as N6-methyladenine (6mA), which was only recently reported[25]. Accounting for oxidation products of 5mC, as well as 3mC and 6mA, there are currently six different known epigenomic modifications at the DNA level but this number is likely to increase in the future, and analysis of their chemical and biological functions is subject to intense ongoing research[26,27]. Histone modifications represent by far the largest category among profiled chromatin marks. With 12 currently known chemical modifications, which can occur at more than 130 post-translational modification sites[28] on five canonical and some 30 histone variants, the theoretical number of combinatorial possibilities is truly astronomical[29], and, consequently, our knowledge of their functional roles is still limited[30,31]. For some commonly studied marks, however, correlations between their presence and the activity of different genomic elements are apparent (for example, methylation of H3K9 and H3K27 at inactive (or poised) promoters, methylation of H3K4 and acetylation of H3K27 on active enhancers and promoters, and methylation of H3K36 in transcribed gene bodies; for a more comprehensive view, see the 'dashboard' of histone modifications from Zhou and colleagues[30]). Profiling of nucleosome occupancy along the genome can reveal regions of open chromatin that may have gene regulatory functions. Interestingly, common trait-associated single-nucleotide polymorphisms (SNPs) identified through GWAS approaches frequently lie outside of coding regions but fall into DNase I-hypersensitive sites (DHSs), where they are thought to regulate distal genes[32]. Profiling of RNA modifications on coding and non-coding RNAs is less advanced, owing to technological limitations. Although more than 100 different RNA modifications are known[33], new modifications are continuing to be discovered[34–36] and they have not yet been comprehensively profiled across the transcriptome[33,37]. Further advanced is the systematic profiling of chromatin architecture, which only recently became technically and economically feasible. Projects, such as the 4D Nucleome (mentioned above), aim to link genetic and epigenomic variants with the enhancers and promoters that they interact with in three-dimensional space, thereby defining gene-set interactomes and pathways as new candidates for functional analysis and therapeutic targeting. Such local short-range interactions have been shown to aggregate into higher-order chromatin domains, which can themselves have functional roles[38].

Data integration and interpretation. A wide variety of different marks and features have been profiled, resulting in an amount of published epigenomic data that can easily be overwhelming. Taken together, comparative approaches have resulted in reliable information about the composition and plasticity of mammalian epigenomes during development and disease. However, without additional context, it remains difficult to predict from these descriptive data which of the large number of marks, features and profiles are the most indicative for causal and quantitative effects (FIG. 2). Consequently, next-generation approaches for the integration and interpretation of chromatin features have attracted great interest. IHEC, for example, has identified a subset of nine profiles and assays (BOX 2) that are required for the generation of so-called reference epigenomes in order to bundle obtained epigenomic data and to maximize the potential of these data to infer function. The rapidly growing number of reference epigenomes registered in EpiRR (see Further information) constitutes an ideal starting point for integrative analysis. More recent approaches using high-level epigenomic data integration have been pioneered by the Roadmap Epigenomics Project and have so far been applied to 111 human epigenomes[39]. Typically, such integrated data sets consist of 20–50 genome-wide profiles making up a multi-dimensional data matrix, as illustrated in FIG. 3. To ensure consistency across the matrix, novel methods such as ChromImpute have been developed for large-scale imputation of epigenomic data[40], resulting in several improvements: the detection of low-quality data, the inference of missing data, and, as a consequence, a more accurate and complete annotation and interpretation of epigenomes.

Box 1 | **A summary of key early epigenetic experiments that addressed functions of DNA methylation [Au:please edit to 1 line]**

Chromatin can be altered in a large variety of ways, but only a few chromatin features have been shown to be functionally involved in gene expression. The first chromatin mark to gain attention (and which is still the best-studied) was DNA methylation. Discovered in the late 1940s as a modification of cytosine bases[24,118] it early on became a prime candidate for an epigenetic effector because of its uneven distribution in the genome and its heritability[119–121]. The first correlations between gene expression and DNA methylation were reported on a series of highly informative model loci (including chicken and mammalian globin genes, the X-chromosome inactivation centre (XIC), regions subjected to genomic imprinting, and virus, transgene or retrovirus silencing)[122–124]. However, it was only after experimental inhibition[125] or deletion of DNA methyltransferases[126], *in vitro* methylation of DNA[127] and genetic deletion of differentially methylated regions[51], that functional connections could be deduced. Because of epigenomics and transcriptomics we now know that the relationship between DNA methylation and gene expression is likely to be more complicated than was initially suspected. Although marks at certain positions correlate with the silencing of some genes (for example, in colon cancer)[128], others are rather uninformative or even occur at active genes[129–132].

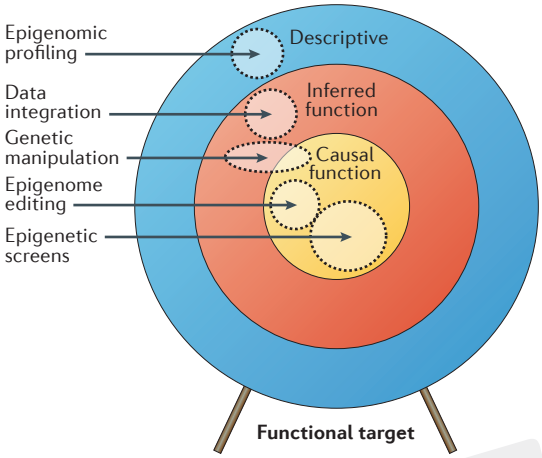


Figure 2 | **Experimental approaches and the level of function that they report.** **[Au:OK?]** Whereas epigenomic profiling alone results in descriptive information, the integration of multiple layers of information allows function to be inferred. The genetic manipulation of elements that harbour epigenetic features of interest can reveal the relevance of epigenetic features, albeit only indirectly. Currently used methods for epigenome editing of single marks or features can reveal causality, and future approaches using epigenetic screens (high-throughput epigenome editing) will enable us to identify novel functional marks on an epigenomic scale.

For the interpretation of such complex data, multi-dimensional matrices can then be aggregated or collapsed into a small number of chromatin states using computational programs such as ChromHMM[41] and Segway[42,43], which have been trained on a variety of data sets. These data integration approaches result in chromatin states being annotated with inferred function (including 'promoter', 'enhancer', 'insulator', 'transcribed' and 'repressed') at a particular genomic locus. Several of such inferred enhancers that are defined by strong H3K4 methylation and weak signals of RNA polymerase II (RNAPII) occupancy have been experimentally validated[41]. Furthermore, chromatin states have also been used in combination with Hi-C interaction maps to predict individual and cell type-specific enhancer–promoter interactions using TargetFinder[44]. On the basis of these and many more specialized tools such as Epigram, epiGRAPH, Epilogos, eFORGE, Epigwas, ChromNet and the Epigenetic Clock (see Further information), complex chromatin maps can be further segmented.

In the context of disease, recent examples using integrative epigenomic analysis include the discovery of pathogenic rewiring of cell type-specific enhancer circuits in obesity[45] and type 1 diabetes[46], as well as the finding that epigenomic changes accompany innate immunity in humans[47]. Of these, the first study[45] best exemplifies the profiles-to-function approach (known as P2F) presented in this Review. Using integrative analysis (as illustrated in FIG. 3) the authors predicted the cell type and regulatory element (enhancer) in and through which a genetic variant that was identified by GWAS was likely to exert its function in fat mass and obesity (FTO)-associated obesity. They achieved this by mapping GWAS-identified risk variants onto chromatin state annotations that were generated by profiling 127 human cell types to predict the regulatory nature of the target region and the cell type in which this region was most likely to be functional. They then used haplotype-specific enhancer assays to validate the enhancer status of the predicted element, Hi-C to link the predicted enhancer to two target genes involved in early adipocyte differentiation, and expression quantitative-trait-locus (eQTL) analysis in primary human adipocytes from both risk-allele and non-risk-allele carriers to assess changes in gene expression. Finally, they restored the correct expression of the affected target genes in cells isolated from patients and a mouse model using CRISPR–Cas9 genome editing. In this case, a genetic variant was shown to be causally involved in a pathway for adipocyte thermogenesis regulation linked to pro-obesity and anti-obesity effects. There is no reason why the same P2F approach in combination with epigenome editing (as outlined below) should not work equally well for elucidating causal functions of epigenomic modifications and variants. Indeed, initial attempts to follow this strategy using a general pipeline are extremely encouraging but existing experimental and computational limitations, as well as currently unknown future challenges, will need to be overcome as the field moves forwards[48]. Together with many other studies, these profiling and data integration efforts have resulted in a fantastic resource that already allows us to infer which marks and features may be functional and which forms the starting point for future analyses of causal function.

Towards genetic analysis of causality

Epigenomic profiling has aided the discovery of a plethora of coordinated chromatin changes that occur during development and disease. Data integration enables these candidate sites to be reduced to a subset with inferred function (FIG. 2). However, the experimental validation of their relevance remains difficult. To some extent, genetic approaches have successfully provided evidence for the importance of chromatin marks. We discuss below two widely used approaches: genetic manipulation of the DNA domains underlying an epigenomic feature; and genetic manipulation of the enzymes that are responsible for the establishment or removal of these features.

Genetic manipulation of sites of chromatin marks or features. Individual epigenomic features can be removed through the manipulation of the underlying DNA sequence. Although it is possible to mutate or to delete single bases that harbour DNA modifications, this approach is not applicable to histone modifications or to larger epigenomic features. In many cases, entire genomic domains containing the feature of interest must be excised instead. Gene targeting has been an exceptionally successful approach to functionally link several epigenetic mechanisms (DNA methylation, chromatin insulation and non-coding transcription) to genomic imprinting[49–52]. However, in most cases, genetic

Hi-C
Experimental method to map contacts formed between segments of DNA in three-dimensional space on a genome-wide scale.

CRISPR–Cas9
(Clustered regularly interspaced short palindromic repeats–CRISPR-associated protein 9). Components of a bacterial defence system against viruses.

manipulation only provides indirect evidence for causality (FIG. 2), because functional consequences could also be attributed to the loss of the genomic DNA sequence rather than to the loss of the epigenomic feature.

Currently, genetic manipulation is often the only available option for conducting functional experiments (for example, see Fanuci *et al.*[53]). Improved methods using targetable nucleases[54] have made it easier to

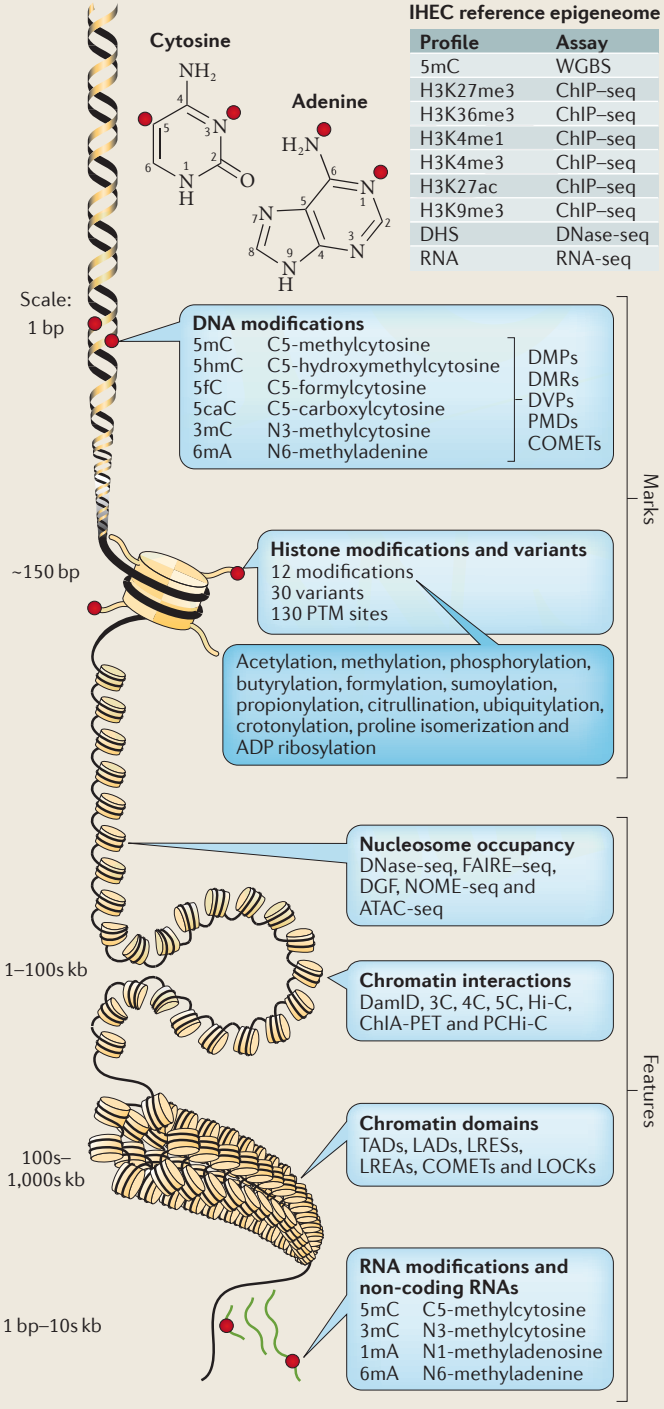
Box 2 | Profile types and categories

The number of epigenomic marks and features that can be profiled is an inherently moving target. Consequently, the profile types and six categories shown here (see the figure) are incomplete and subject to change. The diagram of chromatin depicts common marks and features that are further grouped into six boxed categories that are listed on the right-hand side.

At the DNA level, modifications have been shown to occur at position C5 or N3 on cytosines and at position N6 on adenines and have been shown to be catalysed either enzymatically by DNA methyltransferases (DNMTs) for 5 mC or chemically for 3 mC. The mechanism of modification for 6 mA is still unknown. As part of an active demethylation pathway that is mediated by TET enzymes, 5 mC can be further modified by stepwise oxidation to 5hmC, 5 fC and 5caC. Because of their versatility, profiling of DNA modifications can be configured for multiple readouts, including differentially methylated positions (DMPs) and differentially methylated regions (DMRs), differentially variable positions (DVPs), partially methylated domains (PMDs) and blocks of co-methylation (COMETs)[133,134].

At the histone level, 12 enzymatically catalysed modifications have been shown to occur at more than 130 post-transcriptional modification (PTM) sites at the tails of the four core histones (H2A, H2B, H3 and H4) and some 30 histone variants[135]. Recently, further modifications (H3K64ac and H3K122ac) were also observed in the globular domain of H3 and were shown by epigenome editing to define a new class of enhancers[136], adding to the evidence that suggests that our current knowledge of epigenetic modifications is far from complete. Despite their extraordinary variety, histone modifications are profiled by a single assay and readout (chromatin immunoprecipitation followed by sequencing (ChIP–seq)).

On the feature level, we distinguish three categories of structurally different features for which multiple profiling assays have been developed depending on the complexity of the readout. For profiling nucleosome occupancy, DNase I footprinting (DNase-seq) has been the assay of choice in the past to identify cell type-specific regulatory elements but the assay for transposase-accessible chromatin sequencing (ATAC-seq)[137] is now becoming increasingly popular owing to its simpler workflow and its ability to be used with substantially fewer cells. Other nucleosome-occupancy techniques include formaldehyde-assisted isolation of regulatory elements followed by sequencing (FAIRE–seq), digital genomic footprinting (DGF), and nucleosome occupancy and methylome sequencing (NOME-seq). Different implementations of chromosome conformation capture assays (3C, 4C, 5C, Hi-C, chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) and promoter capture Hi-C (PCHi-C)) are being used to connect enhancers to the promoters that they control, as well as for profiling of chromatin insulators that block those interactions. Especially when integrated with DNA and/or histone modification profiles, profiling with Hi-C[138] allows the segmentation of the epigenome into a variety of chromatin domains (for example, topologically associated domains (TADs)). On the transcriptome level, many non-coding RNAs have been mapped, which in some cases can regulate gene expression[139], and profiling of RNA modifications (for example, modifications occurring at position C5 or N3 on cytosines and at positions N1 and N6 on adenines) has not yet been systemically analysed. In the upper left-hand panel, base positions at which methylation has been found in DNA or RNA are marked in red. Because of the complexity of combinatorial possibilities, there is currently no tissue or cell type for which all marks and features have been profiled. The largest collection of tissues and cell types for which at least nine core marks (shown in the upper right-hand panel) have been consistently profiled are those also referred to as International Human Epigenome Consortium (IHEC) reference epigenomes.



DamID, DNA adenine methyltransferase identification; DHS, DNase I-hypersensitive sites; LADs, lamina-associated domains; LOCKs, large organized chromatin K9 modifications; LREAs, long-range epigenetic activation domains; LRESs, long-range epigenetic silencing domains; WGBS, whole-genome bisulfite sequencing.

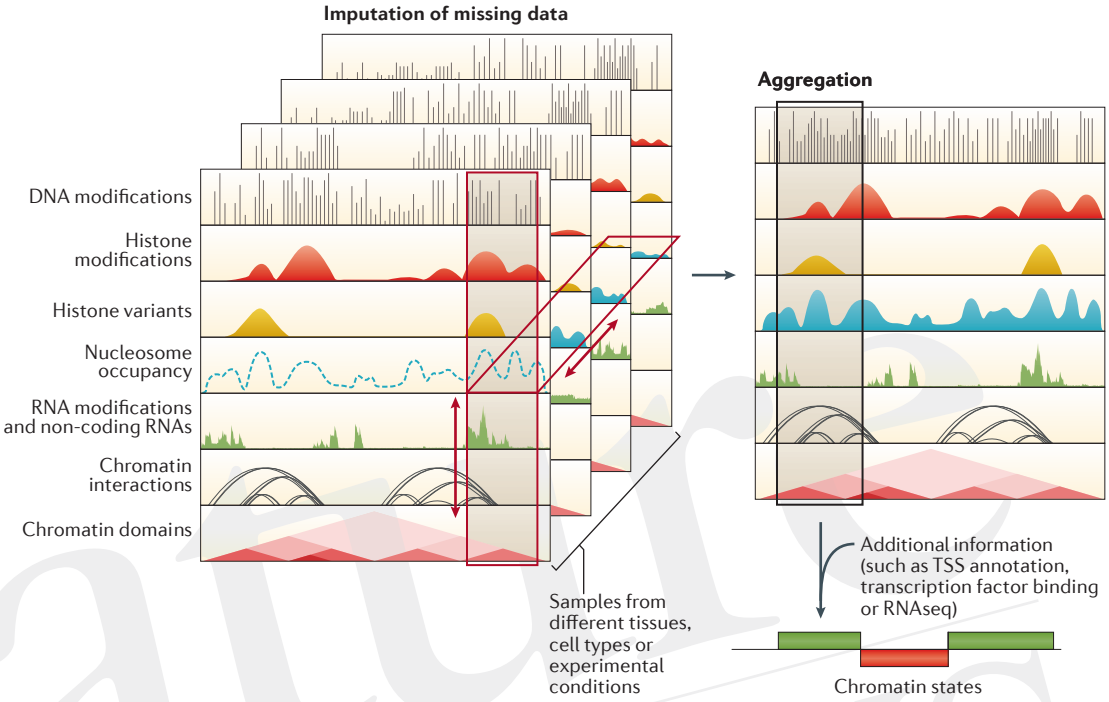


Figure 3 | **Multi-dimensional epigenome profile integration.** Integration is achieved through two steps. First, missing data are imputed using profiles from the same sample (vertical red box) and/or closely related samples (horizontal red box). Currently, histone modification and transcription factor binding profiles are mostly used for imputation but there is no reason why other profiles cannot also be used. Second, additional, non-epigenomic data can be added as appropriate. Although not the focus of this Review, such non-epigenomic data are equally important and include information on gene expression levels and transcription start sites (TSSs) derived from RNA sequencing (RNA-seq) studies. The entire data set per sample is then aggregated and segmented into chromatin states, ranging from two (as shown here by the filled red and green rectangles) to >50, depending on context and complexity[140]. Although most current epigenomic data have been derived from cell and tissue samples from healthy individuals, disease-relevant data from clinical samples and further integration with multi-omics data and pathways can be expected to follow soon, for example, as part of the recently established H2020 MultipleMS consortium for multiple sclerosis and the SYSCID consortium for chronic inflammatory diseases. For more details on the current strategy please refer to the fat-mass and obesity (FTO) example[45] given in the section on 'Data integration and interpretation'.

COMETs
Blocks of co-methylation identified by methylome segmentation.

Chromatin immunoprecipitation followed by sequencing
(ChIP–seq). A method for mapping the distribution of histone modifications or chromatin-associated proteins or transcription factors along the genome. DNA and proteins are crosslinked and an antibody specific to the protein of interest is used to enrich for DNA sequences bound to this protein. These are then identified by sequencing, revealing the genome-wide profile of the protein of interest.

Assay for transposase-accessible chromatin sequencing
(ATAC-seq). A method to identify regions of open chromatin in cells using an engineered Tn5 transposase to both cleave DNA and integrate primer sequences into the cleaved DNA.

Formaldehyde-assisted isolation of regulatory elements followed by sequencing
(FAIRE–seq). A technique that uses the solubility of open chromatin in the aqueous phase during phenol–chloroform extraction to identify sites of open chromatin.

Chromosome conformation capture assays
A group of techniques (including 3C, 4C, 5C, Hi-C and ChIA-PET) that are used to map physical interactions between segments of DNA in three-dimensional space.

Topologically associating domains
(TADs). Regions of chromatin in which loci frequently interact with each other, usually based on evidence from chromosome conformation capture techniques. Loci located in different TADs do not frequently come into contact.

experimentally generate precise modifications of genetic sequence. Such approaches have been successfully used to interrogate enhancer regions[55–59] and to investigate the function of local chromatin architecture. As mentioned in BOX 2, local chromatin architecture can be profiled and segmented into topologically associating domains (TADs), the boundaries of which seem to be genetically defined by orientation-specific CTCF-binding sites. Genetic inversion of CTCF-binding sites has been used to shift domain boundaries at the protocadherin gene cluster, leading to a re-configuration of enhancer–promoter interactions and to the reduced expression of some of the associated genes[60]. Similarly, a TAD structure at the *EPHA4* locus was modified in a mouse model by introducing deletions and inversions that mimic those observed in patients with limb malformations[38].

Whereas in the examples mentioned above a handful of candidate features within small and well-defined regions (imprinted domains, individual enhancers or single topological domains such as TADs) were manipulated, most epigenomic profiles contain many hundreds or thousands of candidate marks that are distributed across the entire genome. A strategy for how genetic manipulation can nevertheless be used to interrogate many epigenomic candidate sites simultaneously was recently introduced[61]. Korkmaz *et al.* integrated different published chromatin profiles (for example, H3K4me2 and H3K27ac) and transcription factor-binding sites to generate a candidate list of active enhancers that were bound by p53. To reveal which of these enhancers was necessary for a specific function of p53, namely, the induction of oncogene-induced senescence, the authors introduced targeted mutations in 685 regions and found that, surprisingly, most of the p53-bound enhancers were dispensable for triggering senescence. Instead, the authors were able to show that only two genomic binding sites of p53 are mandatory for this disease-relevant mechanism[61].

Genetic manipulation of chromatin-modifying enzymes. Despite being generally successful in attributing causal functions to genomic domains hosting epigenomic marks and features, the approaches mentioned above cannot establish the extent of the contribution

of epigenetics to the observed effects. A second experimental strategy using genetic targeting, namely, the deletion or mutation of chromatin-modifying enzymes, overcomes this drawback and thus has been most instrumental in demonstrating the participation of the corresponding epigenomic marks **[Au:OK? Or the participation of the enzymes per se?]** in gene regulation. Epigenetic model systems that are applicable to early embryogenesis-based or embryonic stem (ES) cell-based experiments (genomic imprinting and retro-transposon silencing, among others) are especially useful for attributing crucial roles to chromatin modifications, as these systems allow the study of the acute effect that the loss of certain chromatin marks (for example, H3K9 methylation and DNA methylation) has on the expression of candidate loci (imprinted genes and retro-transposons, respectively)[62–65]. Moreover, knockout studies have been able to clearly establish that a large variety of chromatin-modifying enzymes are essential for normal animal development, as their loss induces embryonic lethality, sometimes quite early (for example DNMT1 (REF. 66), DNMT3A and DNMT3B[67], G9A[68], SUV39H1 and SUV39H2 (REF. 69), HDAC1 (REF. 70), EZH2 (REF. 71), SETDB1 (REF. 72) and LSD1 (REF. 73)). However, these experiments are less informative about the frequency of functional chromatin marks. Embryogenesis is a highly complex process, which can be disturbed in many ways. The loss of expression of a single protein can easily trigger lethality (even early lethality[74,75]). Consequently, it is difficult to deduce the functional relevance of individual marks from the elimination of many thousands. Beyond the difficulties in distinguishing the local versus global epigenomic consequences when chromatin-modifying enzymes are mutated (or pharmacologically inhibited), there are more aspects to consider. Chromatin-modifying enzymes have a much larger range of substrates than is often presumed. Most, if not all, histone-modifying enzymes also have non-histone targets[76,77]. Therefore, the resulting embryonic phenotypes cannot always be attributed to the misregulation of histone marks alone.

Embryonic lethality that arises from the germline depletion of genes can be circumvented by conditional knockouts. This strategy has been so successfully applied to chromatin-modifying enzymes that they can only be incompletely discussed here. The well-studied haematopoietic[78–82], muscular[83] and cardiac[84,85] systems are typical examples that have been used to show the crucial roles of DNA methylation[78–80], H4K20 methylation[83], H3K27 methylation[81,82,84] and histone acetylation[85] in somatic stem cell homeostasis, lineage specification and progression. However, it should also be mentioned that phenotypes detected in such studies are rarely driven by the deregulation of a multitude of genes as they are often surprisingly specific (for example, affecting only certain lineage choices[78,81] or cellular phenotypes[83]) and in some cases have even been rescued by normalizing the expression levels of single genes[82–84].

Although these observations of focused transcriptional consequences and specific phenotypes can in many cases be explained by the incomplete loss of chromatin marks or widespread compensation by redundant chromatin complexes, it could also indicate that only a small number of chromatin marks mediate functional effects that are large enough to cause cellular phenotypic changes and that those might strongly depend on the cellular context. Consistent with this idea is the fact that homeostatic cells often remain relatively unaffected by the pharmacological inhibition of chromatin-modifying enzymes, and many cancer cells show enhanced 'epigenetic vulnerability' (for a concise review see Dawson *et al.*[86]). Another example is the finding that, *in vitro*, even the global loss of canonical epigenetic marks does not necessarily result in major transcriptomic changes. The (almost) complete loss of DNA methylation[87], H3K27me3 (REF. 88) or an artificial induction of H3K4me3 marks[89] in ES cells, for example, results only in minor transcriptional changes, despite affecting their differentiating progeny. Taken together, these data indicate that a majority of epigenetic marks may not have decisive roles in stable cell populations and that the causality of chromatin marks is only revealed when accompanied by major cellular transitions such as differentiation, reprogramming and transformation. An alternative (but not mutually exclusive) explanation would be that many chromatin marks in the epigenome have opposing causal roles, which are often 'cancelled out' when marks are globally altered but which may result in more pronounced phenotypes when the manipulation of chromatin modifications is restricted to a few individual loci. In summary, the genetic manipulation of chromatin-modifying enzymes has been crucial to implicating their causal involvement in many biological processes; however, the functional involvement of individual chromatin marks can mostly still only be inferred (FIG. 2).

Site-specific epigenome editing

Recruitment of chromatin-modifying enzymes to specific loci. The genetic experiments described above indicate that only a small proportion of marks detected in epigenomic profiles may have direct causal roles (FIG. 2). Consequently, new experimental approaches able to directly test the causality of individual epigenetic marks are in high demand. Several new approaches are currently emerging, including the exploitation of naturally occurring or engineered histone mutations[90,91] and targetable chromatin remodelers[92]. Another approach is the fusion of chromatin-modifying enzymes (or catalytic domains) to targetable DNA-binding domains, which has made it possible to change single chromatin marks at particular genomic sites. This constitutes a substantial technological advance, as it is now possible to interrogate the function of individual marks instead of removing the underlying DNA sequences or all instances of a particular mark across the entire genome by genetic or pharmacological approaches. There is now a range of systems that allow the targeting of a chromatin-modifying enzyme to specific DNA sequences by fusing it to either a zinc finger, a transcription activator-like effector (TALE) or a catalytically inactive variant of the bacterial Cas9 nuclease (dCas9). Specifying the genomic

Zinc finger
A modular DNA-binding protein that can be engineered to bind to a sequence of choice.

Transcription activator-like effector
(TALE). DNA-binding protein that has a modular architecture ,with each module (~34 amino acids) recognizing a single nucleotide in a DNA sequence and that can therefore be engineered to bind to a DNA sequence of choice.

Cas9
(CRISPR-associated protein 9). Useful for genome engineering because it can be guided (by a guide RNA) to a particular site in the genome where it makes a DNA double-strand break.

dCas9
The nuclease-dead version of Cas9, which can no longer produce DNA double-strand breaks.

target sequence using zinc finger or TALE architectures involves assembling multiple repetitive protein domains that each recognize a particular DNA base in the target sequence. By contrast, the CRISPR–Cas system can be targeted to a precise genomic location by specifying the base sequence of part of a synthetic RNA known as a guide RNA (gRNA). Remarkably, Cas9 is able to target genomic sites, even when they are functionally silenced or structurally condensed, although this influences the dynamics of DNA recognition[93]. It is easier and faster to generate large numbers of gRNAs that target different sequences than it is to assemble a large number of different zinc finger or TALE domains. The main advantage of the CRISPR–Cas system for epigenomic editing thus lies in the ease of generating targeting constructs and its potential for multiplexing.

Using these platforms to target chromatin-modifying enzymatic domains to particular sites in the genome enables the testing of whether individual chromatin marks have causal effects on gene expression (as illustrated in FIG. 4a). In addition to more general transcriptional transactivator and repressor proteins[94], a range of chromatin-modifying enzymes have already been attached to DNA-binding domains and shown to successfully add or remove chromatin marks at the target sites (TABLE 1). Collectively, these pioneering studies have shown that the catalytic domains of chromatin-modifying enzymes can be sufficient to induce transcriptional changes when directed to specific target sites. Adequate controls were included in most of these studies, including catalytic mutants which ensured that the observed effect is due to enzymatic activity and not merely due to chromatin binding. For example, demethylation of several sites in the *RHOXF2* promoter leads to the transcriptional upregulation of this gene[95]. Similarly, a dCas9–p300 histone acetyltransferase fusion has been used to activate the transcription of *MYOD* and *OCT4* from proximal promoters and distal enhancers. In many cases, the induction of mRNA production achieved with dCas9–p300 is stronger than that achieved with a classical transactivator domain without enzymatic activities (VP64) at the same site[94,96]. Additionally, lysine demethylase LSD1 has been shown to silence genes when targeted to known enhancer regions[14,97], whereas various targetable constructs of the DNMT3A DNA methyltransferase can decrease transcript levels when targeted to promoters[98–100]. Thus, targetable chromatin modifiers have been used both to upregulate and to downregulate mRNA levels, providing direct evidence that chromatin modifiers can regulate transcription. Whether the observed effects are exclusively mediated through epigenomic marks or whether local modifications of other chromatin proteins can sometimes contribute[101] has yet to be firmly established. Furthermore, effects on transcription are detected following the modification of some, but not all, targeted sites. This indicates inherent differences in the regulatory potential of genomic loci and, consistent with results from genetic experiments, that certain chromatin marks may only be functionally relevant at a subset of sites at which they occur. To further investigate, it will be necessary to study how the catalytic activity of the chromatin modifier at a particular site affects transcription and whether the engineered chromatin changes recruit known 'readers' of chromatin marks.

As discussed above, the term 'function' can take different meanings ranging from inferred to causal, whereby causal function could manifest itself in several ways. Some epigenomic features might be dominant in their effect (for example, directly affecting RNA polymerase activities), whereas others might be dependent on certain pre-requisites to reveal a functional involvement (for example, transcriptional priming: poising the cellular response spectrum by forming a transcription factor binding platform)[19,102]. Thus, in some cases (and quite similar to most other biological mechanisms), the function of an epigenomic feature could depend on the cell type, culture condition or the developmental window studied. Furthermore, causal effects could also reveal themselves on several levels, as a change in transcript level, protein level or cellular phenotype. It is often difficult to judge whether statistically significant but sometimes relatively small engineered changes in transcript levels are biologically relevant. However, it is encouraging that several studies have already achieved changes in protein level through epigenomic editing[97,98,103–105]. Ultimately, however, it will be important to directly test whether engineered chromatin modifiers can influence cellular or organismal phenotypes. Some reports have made such a connection already, showing, for example, that the addition or removal of single chromatin marks is sufficient to alter cell proliferation, colony-forming ability of cancer cells[98], the self-renewal of pluripotent stem cells[97] and even addiction-related behaviour in living mice[103].

One important question that remains unanswered is how common such functional marks are and whether engineered changes can be sustained by cells and mitotically inherited. Although DNA methylation is thought to be the most heritable and stable mark, there is emerging evidence that cells may in fact counteract engineered changes. Engineered DNA methylation marks have been observed in some studies to decrease to background levels *in vitro*[106,107], indicating that they are either actively or passively lost; however, in another report, engineered marks were found to persist[108]. As the targeted sites differed (and in the study with the persistent marks, they were located on a human artificial chromosome) it is possible that endogenous chromatin 'context' determines whether an engineered change can be maintained, but this requires further investigation. If engineered changes are found to be transient — and this may need to be established independently for each type of chromatin modification at each targeted site — negative results with regard to functional effects need to be examined with care. Expression of the targetable chromatin modifier, engineered modifications, transcriptional and phenotypic changes should be monitored over time. One recent study monitored silencing up to 50 days after transfection and showed stable gene silencing using a combination of transiently expressed

Guide RNA
(gRNA). An artificial fusion of CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA) used to target the Cas9 protein to a target site in the genome.

repressors (KRAB, DNMT3A and DNMT3L) targeted either through dCas9 or through TALE DNA-binding domains to the endogenous B2M locus in K562 cells[104]. The maintenance of silencing seemed to be dependent on DNA methylation at this locus, as treatment with 5-azacytidine or targeted demethylation with a dCas9–TET1 construct re-activated the locus in a proportion of cells, whereas targeting with dCas9–p300, the transcriptional activator dCas9–VP160 or treatment with interferon-γ had no effect.

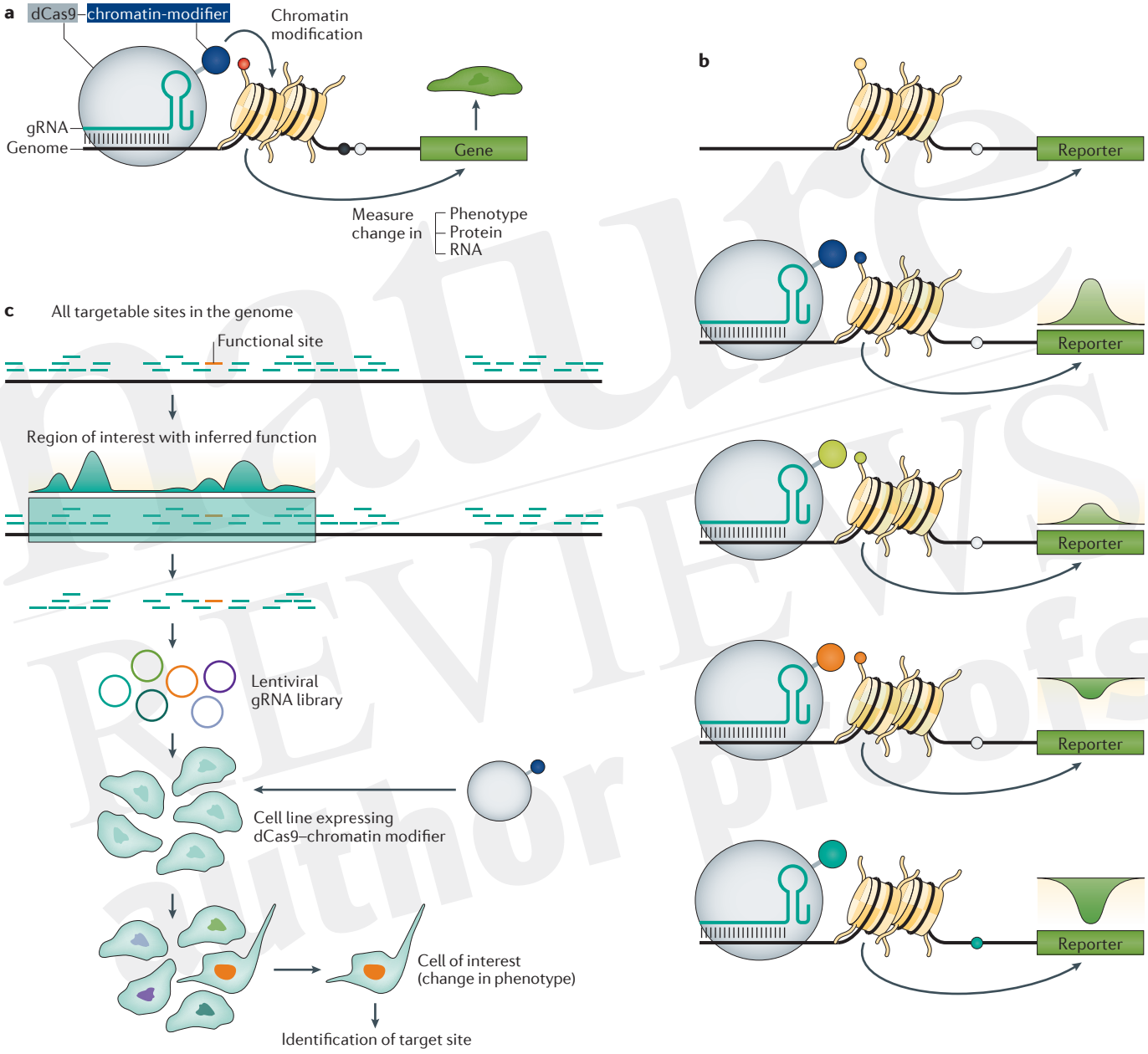


Figure 4 | **Strategies for epigenome editing. a** | A chromatin modifier (or its minimal catalytic domain) is fused to a targetable DNA-binding domain (shown here is nuclease-dead Cas9 (dCas9)). The enzymatic activity of the chromatin modifier is directed towards a particular DNA sequence where it can either add or remove chromatin marks from histones or DNA depending on the nature of the chromatin modifier. This system allows the investigation of how the editing of a mark at a particular site affects the expression of associated genes. Readouts can be at the level of RNA, protein or phenotype (as shown). **b** | Illustration of how targeted epigenome editing can be used to build a hierarchy of functional marks. Various different chromatin modifiers (shown in different colours) are fused to the same DNA-binding domain and targeted to the same site. The effect of the engineered chromatin modification on associated genes can be measured to establish a hierarchy of chromatin features. Upward-pointing and downward-pointing bell-shaped curves represent upregulation and downregulation of the target gene, respectively. **c** | Using epigenome editing allows the identification of chromatin modifications that affect cellular phenotypes. Prior knowledge about the genes involved in regulating particular phenotypes and the location of regulatory elements can be used to design pooled guide RNA (gRNA) libraries that enable targeted screens. If particular gRNAs target the chromatin modifier to functional sites, phenotypic changes are induced; this allows the selection of the cells of interest, determination of gRNA sequences and hence the identification of the regulatory element where an altered epigenomic state controls the phenotype of interest.

Investigating quantitative contributions and hierarchies of regulatory epigenetic marks. Now that tools to manipulate individual chromatin marks have become widely available, the time has come to move on from qualitative descriptions (such as 'silencing' or 'activating') to comprehensively quantify the contribution of individual marks in defining endogenous transcriptional states. Therefore, it would be useful to establish

Table 1 | **Epigenetic modifiers used in epigenomic editing**

Chromatin-modifying enzyme	Function	Targeting protein	Locus targeted	Model system or cell lines	Observed modification and effect size	Effect on transcription and size of effect	Effect at the level of phenotype (protein or other)	Refs
Gene activation								
PRDM9 (catalytic domain)	K4 methylase	dCas9 and ZF	Promoters: *ICAM1*, *RASSF1A*, *EPCAM* and *PLOD2*	HEK293 and A549	Increase of H3K4me3 (up to 60%)	Upregulation of EPCAM (up to 8-fold)	NA	141
p300 (catalytic domain; amino acids 1,048–1,664)	HAT	dCas9, ZF and TALE	• Promoters: *IL1RN*, *MYOD*, *OCT4*, β-globin (*HBE* and *HBG*) and *ICAM1* • Enhancers: *MYOD*, *OCT4* and β-globin	HEK293T	Increased H3K27ac (up to 10-fold increase relative to D1399 catalytic mutant)	Increase in transcription (10–10,000-fold increase in RNA levels relative to D1399Y catalytic mutant)	NA	96
TET1 (catalytic domain)	DNA demethylase	dCas9	Promoter: *BRCA1*	HeLa and MCF7	• DNA demethylation • 10–50% decrease in methylation levels	Increase in transcription up to 2.5-fold	Reduction in cell proliferation	142
TET1 (catalytic domain)	DNA demethylase	dCas9	Promoters: *RANKL*, *MAGEB2* and *MMP2*	HeLa and 293T	DNA demethylation variable	Increase in transcription up to 10-fold	Reduction in cell proliferation	143
TET1 (catalytic domain; amino acids 1,418–2,136)	DNA demethylase	TALE	Promoters: *HBB*, *KLF4* and *RHOXF2*	K562, HEK293 and HeLa	• DNA demethylation • 10–80% decrease in methylation levels (measured relative to off-target TALE or H1671Y/D1673A catalytic mutant)	Increase in transcription ~2–1,000-fold relative to off-target (measured relative to off-target TALE or H1671Y/D1673A catalytic mutant)	NA	95
TET1 (catalytic domain)	DNA demethylase	dCas9	• Promoter: *Bdnf* • Enhancer: *MyoD*	• Mouse ES cells (reporter genes) • Post-mitotic neurons • C3H10T1/2 MEF cells • *In vivo* mouse model with paternally imprinted SNRNP–GFP reporter	• Up to 60% demethylation (reporter locus, *MyoD*) • Up to 35% at the *Bdnf* promoter	3-fold increase in *Bdnf-IV* and *MyoD* mRNA	• Re-expression of GFP in 25% of transduced cells (by FACS) and in up to 70% *in vivo* (lentiviral delivery to brain) • Expression of BDNF (immuno-fluorescence), fibroblast-to-myoblast conversion (MYOD, by immuno-fluorescence required addition of 5-azacytidine)	105

Table 1 (cont.) | **Epigenetic modifiers used in epigenomic editing**

Chromatin-modifying enzyme	Function	Targeting protein	Locus targeted	Model system or cell lines	Observed modification and effect size	Effect on transcription and size of effect	Effect at the level of phenotype (protein or other)	Refs
Gene repression								
32 repressive histone effector or recruiter domains (including HDAC8 (*X. laevis*, amino acids 1–325), NUE (*C. trachomatis*, full length), SET8 (*T. gondii*, amino acids 1,590–1,893), KYP (*A. thaliana*, amino acids 1–331), RPD3 (*S. cerevisiae*, amino acids 19–340), Sir2a (*P. falciparum*, amino acids 1–273))	• HDAC8 histone deacetylase • NUE histone methyltransferase • SET8 H4K20 methyltransferase • KYP H3K9 methyltransferase • RPD3 histone deacetylase • Sir2a histone deacetylase	TALE	Promoters: *Grn2* and *NeuroG2*	Primary neurons and Neuro2a cells	• KYP increased H3K9me1 (~1.4-fold) • SET8 increased H4K20me3 (~2.4-fold) • NUE increased H3K27me3 (~2.2-fold) • HDACs (HDAC8, RPD3 and Sir2a) reduced H4K8Ac 50–60%	Up to 50–75% decrease in RNA level	NA	110
LSD1 (full length)	Histone H3K4 demethylase	TALE	• Candidate enhancer in *SCL* locus and 40 additional candidate enhancers	K562	• 65% loss of H3K4me2 and 60% loss of H3K27ac (relative to TALE alone and scrambled TALE controls) • Up to 80% loss of H3K4me2 and 90% loss of H3K27ac measured relative to an mCherry transfection control	Up to 50% decrease in RNA level (effect monitored for known targets or nearest expressed genes) **[Au: text moved from earlier column, OK?]**	NA	14
LSD1 (full length)	Histone H3K4 demethylase	dCas9	• *Oct4* distal enhancer • 8 candidate enhancers thought to regulate pluripotency in ES cells • *Tbx3*	Mouse ES cells	• Up to 85% H3K4me2 loss • >90% loss of H3K27ac	>90% loss of mRNA	ES cell morphology changes	97
SUV39H1 (full length and shorter constructs)	HMT	ZF	Promoter: *VEGF*	HEK293	Increased H3K9 methylation (up to 2.8-fold)	40% loss of mRNA	NA	144
G9A (amino acids 829–1,210)	HMT	ZF	Promoter: *VEGF*	HEK293	Increased H3K9 methylation (up to 2.7-fold)	40% loss of mRNA	NA	144
DNMT3A (amino acids 598–908)	DNA methyltransferase	ZF	Promoters: *MASPIN* and *SOX2*	SUM159 and MCF7 cells	Increased DNA methylation	60% downregulation of RNA	• Protein (up to 80% downregulation) • Reduced breast cancer colony formation • Reduced proliferation	98

Table 1 (cont.) | **Epigenetic modifiers used in epigenomic editing**

Chromatin-modifying enzyme	Function	Targeting protein	Locus targeted	Model system or cell lines	Observed modification and effect size	Effect on transcription and size of effect	Effect at the level of phenotype (protein or other)	Refs
Gene repression (cont.)								
DNMT3A–DNMT3L (DNMT3L, C-terminal domain; DNMT3A, catalytic domain)	DNA methyltransferase	ZF and TALE	CpG islands: *VEGFA* and *CDKN2A* ($p16^{INK4A}$–$p14^{ARF}$ locus)	• SKOV3 cells • HeLa cells • Primary human fibroblasts	Increased DNA methylation	40–60% downregulation	Increased proliferation	99
DNMT3A (amino acids 602–912)	DNA methyltransferase	dCas9	Promoters: *IL6ST* and *BACH2*	HEK293	Increased DNA methylation	40–50% downregulation	NA	106
Triple combination of DNMT3A, DNMT3L and KRAB (DNMT3L full length and DNMT3A catalytic domain)	DNA methyltransferase	dCas9 and TALE	• Promoter: *IFNAR1* and *VEGFA* • Promoter and enhancer: B2M –tdTomato	K562 and HEK293T	• Up to 100% DNA methylation • Loss of H3K4me3 and RNAPII signal • Increased H3K9me3 (B2M)	500-fold downregulation of B2M mRNA, and up to 80% reduction in *IFNAR1* and *VEGFA* mRNA	• Stable loss of B2M–tdTomato in 78% (K562) and 25% (HEK293T) of cells (by FACS) • Loss of MHC-1 expression **[Au:OK?]**	104
DNMT3A (full length)	DNA methyltransferase	dCas9	• SNRNP–GFP reporter inserted into *Gapdh* promoter • CTCF binding sites in *miR-290* and *Pou5f1* gene loops	Mouse ES cells	Up to 70% increase in DNA methylation (reporter), 35% at the *miR-290* CTCF site, up to 40% at the *Pou5f1* CTCF site	Up to 3-fold mRNA increase of some genes inside and outside the gene loop	Silencing of GFP in up to 70% of cells (by FACS)	105

A. thaliana, *Arabidopsis thaliana*; *C. trachomatis*, *Chlamydia trachomatis*; dCas9; nuclease-dead Cas9; EPCAM, epithelial cell adhesion molecule; ES, embryonic stem; FACS, fluorescence-activated cell sorting; HAT, histone acetyltransferase; HDAC, histone deacetylase; HMT, histone methyltransferase; *ICAM1*, intercellular adhesion molecule 1; *IL1RN*, interleukin 1 receptor antagonist; *IL6ST*, interleukin 6 signal transducer; *KLF4*, Kruppel like factor 4; LSD1, lysine-specific histone demethylase 1; MEF, mouse embryo fibroblast; *MMP2*, matrix metallopeptidase 2; NA, not applicable; *P. falciparum*, *Plasmodium falciparum*; RHOXF2, Rhox homeobox family member 2; RNAPII, RNA polymerase II; *S. cerevisiae*, *Saccharomyces cerevisiae*; SCL, stem cell leukaemia; TALE, transcription activator-like effector; *T. gondii*, *Toxoplasma gondii*; *VEGF*, vascular endothelial growth factor; *X. laevis*, *Xenopus laevis*; ZF, zinc finger.

a hierarchical order of these marks; that is, to elucidate which functional chromatin marks are primary triggers (influencing other epigenomic features) and which marks usually occur as secondary consequences. In this way, it would be possible to pinpoint the proportion of transcription that is strictly defined by chromatin features, how other gene-regulatory mechanisms (transcription factors, topological structure and non-coding transcription) are interlinked and where the molecular switches can be found that functionally turn genes on and off.

To quantify the individual contribution of different epigenetic modifications, many different types of engineered chromatin marks should be directly compared (FIG. 4b). To this end, epigenomic editing approaches should not solely concentrate on re-assessing the importance of well-studied chromatin marks but should instead include some of those marks and features that have so far been less comprehensively analysed (depicted in BOX 2). However, this will require the swift development of a series of new engineering tools. In yeast, the first important steps in this direction have already been taken[109]. In this organism, more than 100 different chromatin factors were fused to the same zinc finger DNA-binding domain targeting the promoter of a reporter gene. Some of the targeted proteins were found to act as activators, whereas others acted as inhibitors, which allowed subsequent ranking based on effect size; that is, changes in protein expression in response to targeted chromatin modification[109]. Studies in mammalian cells have not yet been as comprehensive but some recent publications have followed a similar strategy to compare the effects of a series of chromatin domains[110] and modifications[108] on candidate genes in mammalian cells[94].

Such approaches will be the basis for elucidating the hierarchical order of chromatin marks. Targeting more than one chromatin modification to the same locus will allow the elucidation of which marks are causing others to change and which are functionally dominant, antagonistic, additive and synergistic. Sequential expression of dCas9–chromatin-modifier constructs will show whether the timing of modifications is important in establishing chromatin states and, in combination with overexpression and knockdown constructs, will pinpoint the relationship of marks with other gene-regulatory

mechanisms (such as transcription factors). Thus, expanding the molecular toolbox of epigenome editing will be of lasting benefit, owing to the large amount of possible questions to tackle in the near future.

Epigenome editing of higher-order chromatin architecture. Local chromatin architecture has been extensively manipulated by genomic targeting through deletions or mutation of regulatory regions, insulators and border elements (for example, see REFS 38,49,111,112). Strategies are currently emerging that can alter domain boundaries without affecting the underlying genomic DNA sequence. Although targetable CTCF proteins have not yet been reported, a zinc finger–LDB1 fusion has been used to target the inactive β-globin locus in an erythroblast cell line lacking GATA1 (REF. 113). Binding of LDB1 induced the formation of a chromatin loop between the promoter and a locus control region and was sufficient to activate the transcription of β-globin, although the expression levels did not reach wild-type levels. A similar approach has been used to activate the expression of developmentally silenced fetal globin genes in mouse and human erythroid cells, which might have translational value for the treatment of sickle-cell anaemia[113].

A vision for high-throughput epigenetic screens. It has now been convincingly demonstrated that particular chromatin modifications can affect the expression of reporter loci and, in some cases, even endogenous genes. One of the currently pressing questions in epigenetics is how much of the genome will be controllable in a similar way. The availability of a wide range of epigenome editing tools could be used in the future to screen for individual epigenomic marks that are either necessary or sufficient for specific cellular phenotypes (FIG. 4c). This would require appropriate libraries of gRNAs, each targeting dCas9–chromatin-modifying enzymes to different genomic sites. In order to avoid screening complete genomes, knowledge gained from epigenomic profiles should be integrated into library design. For example, designing gRNA libraries that focus on informative regions that have been identified from integrative analysis of epigenomic profiles would greatly reduce the complexity of libraries and would thus enhance the power of such epigenomic screens.

As outlined in FIG. 4c, such pooled screens will enable the identification of the subset of causal epigenetic marks among the many with inferred function. gRNA libraries representing loci identified in epigenomic profiles would be introduced into cells expressing the relevant dCas9–modifier fusion. Cells that responded to an individual modification with a suspected phenotypic change will be separated from the bulk population (through selection, fluorescence-activated cell sorting (FACS) or cellular behaviour) and used to gain information about the individual gRNAs the cells received. It is difficult to predict which cellular phenotypes will be most susceptible to these approaches, but one way forwards would be to start with the cellular phenotypes that are clearly epigenetic, easy and accurate to measure and ideally reversible (for example, cell identity, cell cycle control and migration).

In principle, epigenetic screens are already feasible. For example, rather than genetically mutating a large number of enhancers through CRISPR-based genetic screening[61], epigenome editing tools could be used to manipulate chromatin modifications at these sites. A small-scale epigenetic screen such as this has already been conducted. Kearns *et al.* used published profiles of ES- and epiblast-like cells (EpiLCs) to generate a list of candidate enhancers with potential roles in pluripotency. Targeting dCas9–LSD1 to one of these active enhancers mediated H3K4 demethylation and abrogated the transcription of associated genes[97], and the authors then used this system to screen eight candidate enhancers to investigate the effect of H3K4 demethylation on ESC self-renewal[97]. Scaling up such epigenome-editing approaches to epigenome-wide screens has the potential to reveal many (and eventually all) epigenetic marks and features that have causal roles in a given cellular phenotype (FIG. 4c).

To make epigenome editing universally applicable, several issues have yet to be resolved: first and foremost, information about the distribution of the majority of chromatin features is still missing. Only a small minority of epigenomic features has been profiled, and some of the profiling technologies used have since become outdated and/or have evolved (for example, epigenomics on the single-cell level)[114], indicating a continued need for profiling efforts. Moreover, the toolset of efficient dCas9–chromatin-modifier fusion proteins needs to be expanded. There is, for example, no current validated tool for the successful addition or removal of H3K27 methylation, which is one of the most frequently profiled chromatin marks in human cells. Furthermore, gRNA libraries specifically targeting regions that harbour particular chromatin modifications are not yet available. However, simple methods for the generation of ultra-high-complexity or even genome-wide gRNA libraries have already been established[115,116]. Finally, to make the most of the new molecular tools discussed in this Review, reliable *in vitro* models or *in vivo* approaches that allow the assessment (and selection) of induced phenotypic changes need to be developed[103]. Considering the rapid progression of CRISPR-based technology during recent years it is conceivable that comprehensive functional interrogation of chromatin marks and features could become a common component of epigenomic-profiling studies in the near future.

Conclusions

In this Review, we have traced some of the seminal studies and approaches that have been leading the way towards the functional analysis of epigenomic marks and features, which remains one of the main challenges for epigenomics. Based on current evidence, the key innovation to deliver this breakthrough will almost certainly be based on epigenome editing and, in particular, on the ability to conduct epigenome-wide screens to identify causal chromatin features from the myriad of features with inferred function identified through epigenomic profiling and data integration. Returning to the analogy of an epigenetic landscape, epigenomic

Pooled screens
Approaches in which cells receiving the screening library (for example, pools of guide RNAs) are grown and selected together for a phenotypic change.

Fluorescence-activated cell sorting
(FACS). An experimental method that measures a fluorescence-based signal (from a reporter or antibody staining) emitted from individual cells of a population, and uses these fluorescence signals to isolate single cells of interest.

engineering promises to turn the static landscape, as depicted by Waddington, into a dynamic environment, as illustrated by a contemporary animation (FIG. 1b; see Supplementary information S1 (movie)) created by the resident artists of EpiGeneSys (see Further information), the European Network of Excellence for Epigenetics and Systems Biology. In the context of this Review, the pulling of the strings to alter the contours of the landscape represents the approaches discussed here to screen for chromatin marks that causally influence cellular fate in health and disease. On the basis of the tremendous progress made to date, it is perhaps not surprising that expectations are running high to translate any fledgling new insights into novel medicines and treatments. Although epigenetic marks and drugs are already in clinical use as biomarkers and treatments, respectively, for certain types of cancer, epigenome editing has not yet been used therapeutically. In addition to functional candidates and technical improvements, this step would require ethical considerations similar to those currently discussed for genome-editing technology[117]. With these promising developments in mind, does this mean that epigenomic profiling is essentially completed and a thing of the past? Certainly not, because new marks and features are still being discovered and new and improved profiling technologies are still being developed. It will, however, be interesting to explore which type of profiling turns out to be the most informative for which field of research and, in particular, for the discovery of causal functions hidden in chromatin.

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**, 245–254 (2003).
2. Waddington, C. H. The epigenotype. 1942. *Int. J. Epidemiol.* **41**, 10–13 (2012).
3. Waddington, C. H. *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology* (Allen & Unwin, 1957).
The first occurrence of the epigenetic landscape as a metaphor of epigenetic mechanisms.
4. Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell Biol.* **16**, 593–610 (2015).
5. Clark, S. J., Lee, H. J., Smallwood, S. A., Kelsey, G. & Reik, W. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol.* **17**, 72 (2016).
6. Berdasco, M. & Esteller, M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev. Cell* **19**, 698–711 (2010).
7. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
8. Henikoff, S. & Greally, J. M. Epigenetics, cellular memory and gene regulation. *Curr. Biol.* **26**, R644–R648 (2016).
9. Taudt, A., Colome-Tatche, M. & Johannes, F. Genetic sources of population epigenomic variation. *Nat. Rev. Genet.* **17**, 319–332 (2016).
10. Zhao, Y. & Zhou, D. X. Epigenomic modification and epigenetic regulation in rice. *J. Genet. Genomics* **39**, 307–315 (2012).
11. Ha, M. Understanding the chromatin remodeling code. *Plant Sci.* **211**, 137–145 (2013).
12. Schmitz, R. J. & Ecker, J. R. Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends Plant Sci.* **17**, 149–154 (2012).
13. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
14. Mendenhall, E. M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* **31**, 1133–1136 (2013).
This paper coined the term 'epigenome editing' by providing an excellent example proving its capacity.
15. Köferle, A., Stricker, S. H. & Beck, S. Brave new epigenomes: the dawn of epigenetic engineering. *Genome Med.* **7**, 59 (2015).
16. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5–15 (2016).
17. Keung, A. J., Joung, J. K., Khalil, A. S. & Collins, J. J. Chromatin regulation at the frontier of synthetic biology. *Nat. Rev. Genet.* **16**, 159–171 (2015).
18. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR–Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
19. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
20. Mueller-Planitz, F., Klinker, H. & Becker, P. B. Nucleosome sliding mechanisms: new twists in a looped history. *Nat. Struct. Mol. Biol.* **20**, 1026–1032 (2013).
21. Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).
22. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
23. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
24. Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* **175**, 315–332 (1948).
25. Wu, T. P. *et al.* DNA methylation on N^6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
26. Booth, M. J., Raiber, E. A. & Balasubramanian, S. Chemical methods for decoding cytosine modifications in DNA. *Chem. Rev.* **115**, 2240–2254 (2015).
27. Balasubramanian, S. Chemical biology on the genome. *Bioorg. Med. Chem.* **22**, 4356–4370 (2014).
28. Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016–1028 (2011).
29. Khare, S. P. *et al.* HIstome — a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.* **40**, D337–D342 (2012).
30. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18 (2011).
31. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15**, 703–708 (2014).
32. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
33. Li, S. & Mason, C. E. The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genom. Hum. Genet.* **15**, 127–150 (2014).
34. Dominissini, D. *et al.* The dynamic N^1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
35. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways — 2013 update. *Nucleic Acids Res.* **41**, D262–D267 (2013).
36. Cantara, W. A. *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, D195–D201 (2011).
37. Wang, X. & He, C. Dynamic RNA modifications in posttranscriptional regulation. *Mol. Cell* **56**, 5–12 (2014).
38. Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
A prime example of the use of genome-editing technology to show the specific functionality of domains containing complex epigenomic features.
39. Roadmap Epigenomics, C. *et al.* Integrative analysis of111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
One of several key publications of the Roadmap Epigenomics Consortium providing a public resource of epigenomic data.
40. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33**, 364–376 (2015).
41. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
This publication is an example of the possibilities complex epigenomic data holds for the inference of functionality.
42. Hoffman, M. M. *et al.* Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9**, 473–476 (2012).
43. Hoffman, M. M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41**, 827–841 (2013).
44. Whalen, S., Truty, R. M. & Pollard, K. S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.* **48**, 488–496 (2016).
45. Claussnitzer, M. *et al.* FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
This paper shows how GWAs and epigenomic data can be integrated to gain new insights into disease mechanisms.
46. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
47. Saeed, S. *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
48. Spisak, S. *et al.* CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. *Nat. Med.* **21**, 1357–1363 (2015).
49. Kurukuti, S. *et al.* CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc. Natl Acad. Sci. USA* **103**, 10684–10689 (2006).
50. Murrell, A., Heeson, S. & Reik, W. Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat. Genet.* **36**, 889–893 (2004).
Among the first papers to show the involvement of chromatin architecture in genomic imprinting.
51. Wutz, A. *et al.* Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* **389**, 745–749 (1997).
52. Latos, P. A. *et al. Airn* transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* **338**, 1469–1472 (2012).
This publication shows the functional relevance of non-coding transcription (and not the ncRNA itself) to an epigenetic model of gene silencing.

53. Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M. S. & Mhlanga, M. M. Chromosomal contact permits transcription between coregulated genes. *Cell* **155**, 606–620 (2013).
54. Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334 (2014).
55. Zhou, H. Y. *et al.* A *Sox2* distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* **28**, 2699–2711 (2014).
56. Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant *EVI1* and *GATA2* deregulation in leukemia. *Cell* **157**, 369–381 (2014).
57. Webster, D. E. *et al.* Enhancer-targeted genome editing selectively blocks innate resistance to oncokinase inhibition. *Genome Res.* **24**, 751–760 (2014).
58. Wang, S., Wen, F., Tessneer, K. L. & Gaffney, P. M. TALEN-mediated enhancer knockout influences *TNFAIP3* gene expression and mimics a molecular phenotype associated with systemic lupus erythematosus. *Genes Immun.* **17**, 165–170 (2016).
59. Canver, M. C. *et al. BCL11A* enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis. *Nature* **527**, 192–197 (2015).
60. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
61. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR–Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016). **Among the first publication to show the potential of CRISPR for functional screens of non-coding elements.**
62. Bulut-Karslioglu, A. *et al.* Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 (2014).
63. Martens, J. H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800–812 (2005).
64. Branco, M. R., Oda, M. & Reik, W. Safeguarding parental identity: Dnmt1 maintains imprints during epigenetic reprogramming in early embryogenesis. *Genes Dev.* **22**, 1567–1571 (2008).
65. Seidl, C. I., Stricker, S. H. & Barlow, D. P. The imprinted *Air* ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J.* **25**, 3565–3575 (2006).
66. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992). **Early example of the use of gene targeting to study the effect of epigenomic marks.**
67. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**, 247–257 (1999).
68. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791 (2002).
69. Peters, A. H. *et al.* Loss of the *Suv39h* histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
70. Lagger, G. *et al.* Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J.* **21**, 2672–2681 (2002).
71. O'Carroll, D. *et al.* The *Polycomb*-group gene *Ezh2* is required for early mouse development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).
72. Dodge, J. E., Kang, Y. K., Beppu, H., Lei, H. & Li, E. Histone H3-K9 methyltransferase ESET is essential for early development. *Mol. Cell. Biol.* **24**, 2478–2486 (2004).
73. Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41**, 125–129 (2009).
74. Bradley, A. *et al.* The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm. Genome* **23**, 580–586 (2012).
75. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res.* **43**, D726–D736 (2015).
76. Zhang, X., Wen, H. & Shi, X. Lysine methylation: beyond histones. *Acta Biochim. Biophys. Sin. (Shanghai)* **44**, 14–27 (2012).
77. Glozak, M. A., Sengupta, N., Zhang, X. & Seto, E. Acetylation and deacetylation of non-histone proteins. *Gene* **363**, 15–23 (2005).
78. Broske, A. M. *et al.* DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat. Genet.* **41**, 1207–1215 (2009).
79. Lee, P. P. *et al.* A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15**, 763–774 (2001).
80. Shaknovich, R. *et al.* DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* **118**, 3559–3569 (2011).
81. Su, I. H. *et al.* Polycomb group protein Ezh2 controls actin polymerization and cell signaling. *Cell* **121**, 425–436 (2005).
82. Su, I. H. *et al.* Ezh2 controls B cell development through histone H3 methylation and *Igh* rearrangement. *Nat. Immunol.* **4**, 124–131 (2003).
83. Boonsanay, V. *et al.* Regulation of skeletal muscle stem cell quiescence by Suv4-20h1-Dependent facultative heterochromatin formation. *Cell Stem Cell* **18**, 229–242 (2016).
84. Delgado-Olguin, P. *et al.* Epigenetic repression of cardiac progenitor gene expression by Ezh2 is required for postnatal cardiac homeostasis. *Nat. Genet.* **44**, 343–347 (2012).
85. Montgomery, R. L. *et al.* Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev.* **21**, 1790–1802 (2007).
86. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12–27 (2012).
87. Domcke, S. *et al.* Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015). **Domcke *et al.* use DNA-methylation-free ES cells to study transcription factor dependencies.**
88. Riising, E. M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol. Cell* **55**, 347–360 (2014).
89. Schmitz, S. U. *et al.* Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J.* **30**, 4586–4600 (2011).
90. Lewis, P. W. *et al.* Inhibition of PRC2 activity by a gain-of-Function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857–861 (2013). **This publication reveals that histone mutations found in paediatric gliomas are globally blocking Polycomb-repressive complex 2 (PRC2) activity.**
91. McKay, D. J. *et al.* Interrogating the function of metazoan histones using engineered gene clusters. *Dev. Cell* **32**, 373–386 (2015).
92. McKnight, J. N., Tsukiyama, T. & Bowman, G. D. Sequence-targeted nucleosome sliding *in vivo* by a hybrid Chd1 chromatin remodeler. *Genome Res.* **26**, 693–704 (2016).
93. Knight, S. C. *et al.* Dynamics of CRISPR–Cas9 genome interrogation in living cells. *Science* **350**, 823–826 (2015).
94. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).
95. Maeder, M. L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE–TET1 fusion proteins. *Nat. Biotechnol.* **31**, 1137–1142 (2013). **Maeder *et al.* show that targeted DNA demethylation using TALE proteins can influence transcription of endogenous genes.**
96. 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). **A key example of epigenome editing using acetyltransferase activity.**
97. Kearns, N. A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015). **A further key example of epigenome editing based on a histone de-methylase.**
98. Rivenbark, A. G. *et al.* Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* **7**, 350–360 (2012).
99. Siddique, A. N. *et al.* Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. *J. Mol. Biol.* **425**, 479–491 (2013).
100. Bernstein, D. L., Le Lay, J. E., Ruano, E. G. & Kaestner, K. H. TALE-mediated epigenetic suppression of *CDKN2A* increases replication in human fibroblasts. *J. Clin. Invest.* **125**, 1998–2006 (2015).
101. Minucci, S. & Pelicci, P. G. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat. Rev. Cancer* **6**, 38–51 (2006).
102. Papp, B. & Plath, K. Epigenetics of reprogramming to induced pluripotency. *Cell* **152**, 1324–1343 (2013).
103. Heller, E. A. *et al.* Locus-specific epigenetic remodeling controls addiction- and depression-related behaviors. *Nat. Neurosci.* **17**, 1720–1727 (2014).
104. Amabile, A. *et al.* Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. *Cell* **167**, 219–232 (2016).
105. Liu, X. S. *et al.* Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247 (2016).
106. Vojta, A. *et al.* Repurposing the CRISPR–Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44**, 5615–5628 (2016).
107. Kungulovski, G. *et al.* Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenet. Chromatin* **8**, 12 (2015).
108. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720–724 (2016).
109. Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J. & Khalil, A. S. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* **158**, 110–120 (2014).
110. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
111. Tschopp, P. & Duboule, D. A genetic approach to the transcriptional regulation of Hox gene clusters. *Annu. Rev. Genet.* **45**, 145–166 (2011).
112. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
113. Deng, W. *et al.* Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849–860 (2014).
114. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* **16**, 716–726 (2015).
115. Köferle, A. *et al. CORALINA: A Simple Method for the Generation of gRNA Libraries Covering Complete Genomes* (in the press). **[Au: is there an update on this paper?]**
116. Lane, A. B. *et al.* Enzymatically generated CRISPR libraries for genome labeling and screening. *Dev. Cell* **34**, 373–378 (2015).
117. Lanphier, E., Urnov, F., Haecker, S. E., Werner, M. & Smolenski, J. Don't edit the human germ line. *Nature* **519**, 410–411 (2015).
118. Wyatt, G. R. Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochem. J.* **48**, 581–584 (1951).
119. Scarano, E. The control of gene function in cell differentiation and in embryogenesis. *Adv. Cytopharmacol* **1**, 13–24 (1971).
120. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232 (1975).
121. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**, 9–25 (1975).
122. McGhee, J. D. & Ginder, G. D. Specific DNA methylation sites in the vicinity of the chicken β-globin genes. *Nature* **280**, 419–420 (1979).
123. Sutter, D., Westphal, M. & Doerfler, W. Patterns of integration of viral DNA sequences in the genomes of adenovirus type 12-transformed hamster cells. *Cell* **14**, 569–585 (1978).
124. Jahner, D. *et al. De novo* methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**, 623–628 (1982).
125. Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T. & Migeon, B. R. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proc. Natl Acad. Sci. USA* **81**, 2806–2810 (1984).
126. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
127. Busslinger, M., Hurst, J. & Flavell, R. A. DNA methylation and the regulation of globin gene expression. *Cell* **34**, 197–206 (1983). **Among the first studies to functionally test the relevance of an epigenomic mark.**

128. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41**, 178–186 (2009).
129. Metivier, R. *et al.* Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45–50 (2008).
130. Kangaspeska, S. *et al.* Transient cyclical methylation of promoter DNA. *Nature* **452**, 112–115 (2008).
131. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007).
132. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
133. Libertini, E. *et al.* Information recovery from low coverage whole-genome bisulfite sequencing. *Nat. Commun.* **7**, 11306 (2016).
134. Martin, T. C., Yet, I., Tsai, P. C. & Bell, J. T. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics* **16**, 131 (2015).
135. Maze, I., Noh, K. M., Soshnev, A. A. & Allis, C. D. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nat. Rev. Genet.* **15**, 259–271 (2014).
136. Pradeepa, M. M. *et al.* Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat. Genet.* **48**, 681–686 (2016).
137. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
138. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
139. Barr, C. L. & Misener, V. L. Decoding the non-coding genome: elucidating genetic risk outside the coding genome. *Genes Brain Behav.* **15**, 187–204 (2016).
140. Baker, M. Making sense of chromatin states. *Nat. Methods* **8**, 717–722 (2011).
141. Cano-Rodriguez, D. *et al.* Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nat. Commun.* **7**, 12284 (2016).
142. Choudhury, S. R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR–dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* http://dx.doi.org/10.18632/oncotarget.10234 (2016).
143. Xu, X. *et al.* A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* **2**, 16009 (2016).
144. Snowden, A. W., Gregory, P. D., Case, C. C. & Pabo, C. O. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression *in vivo*. *Curr. Biol.* **12**, 2159–2166 (2002).

Acknowledgements

S.H.S. was supported by DFG (STR 1385/1-1). A.K. was supported by a CRUK Ph.D. Fellowship. S.B. was supported by the EU-FP7 BLUEPRINT Project (282510) and the NIHR UCLH Biomedical Research Centre (BRC84/CN/SB/5984).

Competing interests statement

The authors declare no competing interests.

DATABASES

Protein Data Bank: http://www.rcsb.org/pdb/home/home.do

FURTHER INFORMATION

4D Nucleome: http://www.4dnucl.org
ChromNet: http://chromnet.cs.washington.edu
eFORGE: http://eforge.cs.ucl.ac.uk
ENCODE: https://www.genome.gov/10005107/encode-project
Epigenetic Clock: https://labs.genetics.ucla.edu/horvath/dnamage
Epigram: http://compbio.mit.edu/epilogos
epiGRAPH: http://epigraph.mpi-inf.mpg.de/WebGRAPH
Epigwas: https://www.broadinstitute.org/mpg/epigwas
Epilogos: http://compbio.mit.edu/epilogos
EpiRR: http://www.ebi.ac.uk/vg/epirr
GTEx: https://www.genome.gov/27543767/genotypetissue-expression-project-gtex
GWAS Catalogue: https://www.ebi.ac.uk/gwas
Hlstome: http://www.actrec.gov.in/histome
IHEC: http://ihec-epigenomes.org
modENCODE: http://www.modencode.org
Modomics: http://modomics.genesilico.pl
Resident EpiGeneSys artists: http://www.epigenesys.eu/en/science-and-you/art-and-science
RNA mDB: http://mods.rna.albany.edu
TargetFinder: https://github.com/shwhalen/targetfinder

SUPPLEMENTARY INFORMATION

See online article: S1 (movie)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Author biographies [Aus: would you like to include a homepage link?]

Stefan H. Stricker is a group leader at the German Research Center for Environmental Health and the Biomedical Center Munich, Germany. He obtained a diploma in biology at the Ludwig-Maximilian-Universität in Munich, Germany, in 2003. He then received a Böhringer Ingelheim Foundation scholarship, joined the Barlow laboratory at the Research Center for Molecular Medicine and received his Ph.D. in 2008 from the Universität Wien, Austria. A European Molecular Biology Organization (EMBO) long-term fellowship allowed him to conduct his postdoctoral projects in the Smith, Pollard and Beck laboratories at the Cambridge Stem Cell Institute, UK, and the University College London (UCL) Cancer Institute in London, UK. His research focuses on the mechanisms of epigenetic features that determine cellular properties.

Anna Köferle is a Cancer Research UK (CRUK)-funded Ph.D. student in Prof. Stephan Beck's laboratory, University College London, UK, where she is developing a CRISPR-based epigenetic screening method. She obtained an undergraduate degree in Molecular and Cellular Biochemistry from the University of Oxford, UK, in 2012, where her final year project in Prof. Peter Cook's laboratory focused on engineering a gene loop formed by the gene *SAMD4A* using genetic manipulation with a zinc finger nuclease in order to test how changes in nuclear architecture might influence the expression of this well-studied gene. She has a general interest in how transcription is regulated in cells, in particular, the relative contributions that transcription factors, chromatin environment and nuclear architecture make to the regulation of initiation of transcription at the promoter.

Stephan Beck is Professor of Medical Genomics at the University College London (UCL) Cancer Institute, UK. Using experimental and computational approaches, his laboratory has broad interests in the genomics and epigenomics of phenotypic plasticity in health and disease. He received his Ph.D. in 1985 from the University of Konstanz, Germany, where he studied DNA structure. After appointments at the Medical Research Council (MRC) Laboratory of Molecular Biology in Cambridge, UK, Millipore Corporation in Boston, USA, and the Imperial Cancer Research Fund in London, UK, he joined the Wellcome Trust Sanger Institute, Hinxton, UK, in 1996. During his tenure as Head of Human Sequencing, he played a leading role in the sequencing and analysis of the human, mouse and zebrafish genomes. He is a founding member the Human Epigenome Project and the UK Personal Genome Project.

Figure permissions

Figures 1a,b

[Prod: credit lines for PowerPoint slides are needed]

Key points

- Many epigenetic approaches contribute to our understanding of gene regulation and cell identity. By pursuing these epigenetic approaches, different levels of functionality can be proven for epigenomic features.
- Epigenomic profiling is providing a descriptive view of the chromatin landscape, and data integration enables us to 'infer' functionality from complex data sets.
- Genetic manipulations are crucial to show the relevance of chromatin-modifying enzymes (and genomic domains that harbour epigenomic features); however, they can sometimes fall short of distinguishing the effect of an epigenomic feature from other induced changes.
- Epigenome editing provides a new possibility to test the functionality of epigenomic features directly. Numerous recent publications have indicated a ubiquitous applicability of epigenome editing and its potential in manipulating gene expression.
- Once current limitations are overcome, epigenetic screens — using large-scale epigenome editing approaches across the epigenome — will allow functional epigenomic features to be distinguished from non-functional counterparts.

Subject categories

Biological sciences / Genetics / Genomics / Epigenomics [URI /631/208/212/177]
Biological sciences / Molecular biology / Epigenetics [URI /631/337/176]
Biological sciences / Genetics / Gene regulation [URI /631/208/200]
Biological sciences / Genetics / Epigenetics / DNA methylation [URI /631/208/176/1988]
Biological sciences / Molecular biology / Chromatin / Histone post-translational modifications [URI /631/337/100/2285]
Biological sciences / Biological techniques / Epigenetics analysis / Chromatin analysis [URI /631/1647/2210/2211]
Biological sciences / Biological techniques / Genetic engineering [URI /631/1647/1511]
Biological sciences / Genetics / CRISPR-Cas systems [URI /631/208/4041]

Supplementary information

Supplementary movie file 1. Contemporary version of the Waddington landscape (see also FIG. 1). Epigenetic and epigenomic manipulation promises to dynamically change the landscape and thus cellular phenotypes. This movie is from 'EpiGeneScapes' by Paul Liam Harrison (2015). EpigeneScapes is a body of work that includes animations, drawings, sculptures, interactive installations and a series of mixed media prints derived from the C. H. Waddington concept of the epigenetic landscape. The animated model was developed in collaboration with Mhairi Towler and Link Li and was undertaken as part of Paul Liam Harrison's role as Associate Artist with EpigeneSys European Network of Excellence. http://www.epigenesys.eu/en/science-and-you/art-and-science. Courtesy of P. L. Harrison, University of Dundee, UK, and EpiGeneSys.

Brightcove ID: 5184448185001

ToC

000 **From profiles to function in epigenomics**

Stefan H. Stricker, Anna Köferle and Stephan Beck

A wealth of data is emerging from diverse studies of epigenomics, including genome-scale profiles of DNA methylation, histone modifications and higher-order chromatin features. In this Review, the authors discuss how, despite all this information, many challenges remain for inferring and proving the physiological and pathological functions of chromatin states. They describe the degrees of 'functionality' that are revealed by different experimental approaches, the value of integrative strategies and visions for the future.