From profiles to function in epigenomics

Stefan H. Stricker°, Anna Köferle and Stephan Beck**

***UCL Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6BT**

° Institute of Stem Cell Research, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, Neuherberg, 85764, Germany and Biomedical Center, Ludwig-Maximilian-Universität, Grosshaderner Strasse 9, Planegg-Martinsried, 82152, Germany

Correspondence to SB

e-mail: s.beck@ucl.ac.uk

Abstract

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

Introduction

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

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. A number of ground-breaking studies have revealed that alterations to chromatin, the nuclear complex of macromolecules consisting 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 **Box1**). For the purpose 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).

Catalogs of chromatin marks and features obtained from cells and tissues at different stages of development and disease states have since become an extremely useful resource. Epigenomic profiling was the key to discover 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 previously not allowed the interrogation of individual or combination of marks to test for direct functional effects. The majority of research consequently 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" [7]. Today epigenetic research is at a turning point. New approaches, benefiting from the remarkable developments in genome editing, enable us to move forward and finally elucidate which

individual chromatin marks or features play causal roles in processes such as gene regulation, cellular memory, cellular differentiation and disease aetiology [8].

The term "function" or "functional" means different things to different people and, in our view it is at times incorrectly used 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 **Figure 2**. Inferred function is usually based on correlation of aggregated marks or features with observed effects, e.g. gene activity states or phenotypes, but cannot establish whether marks play truly causal roles. In 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 will emphasise the level of function that can or has been demonstrated using different experimental approaches. We will discuss what can be learned from comparative chromatin profiling, and how associations of chromatin marks with phenotypes can identify candidate regions for functional testing. While fully appreciating the importance of plant epigenomics and associated resources [9-13], this review will focus on mammalian chromatin. We will briefly revisit insights gained from early knockout studies but mainly concentrate on recent epigenome editing approaches [14] (sometimes also referred to as epigenetic engineering [15]), which test causality directly. A short overview of methods for epigenome editing will be provided. However, for a more detailed discussion of technology aspects, we refer the reader to excellent recent reviews on genome and epigenome editing [16-18]. Finally, we will speculate how these approaches could be used to efficiently deduce causal function from profiles in the future. While we recognize the importance of differential expression and binding of transcription factors, nucleosome positioning and chromatin remodelers to gene regulation (thoroughly reviewed previously)[19-21], we will focus mostly 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 online links for details) 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 we now have over and above the genome by several orders of magnitude.

Projects and Data types

Starting in 2003, ENCODE (Encyclopedia of DNA Elements) was the first international project employing large-scale epigenomic profiling to identify regulatory elements in the human genome. ENCODE pioneered many of the required technologies (e.g. 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 to the effort. ENCODE became a member of the International Human Epigenome consortium (IHEC), a project launched in 2010, which aims to generate 1,000 reference epigenomes in primary tissues and cell types and has become the umbrella organisation under which national and international epigenome efforts are jointly coordinated. IHEC currently has 9 members, ENCODE (USA), Roadmap Epigenomics (USA), BLUEPRINT (EU), DEEP (Germany), Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC; Canada), Core Research for Evolutional Science and Technology (CREST; Japan) and the national epigenome projects from Korea, Singapore and Hong Kong (see online links for details). Within the context of this review, the key IHEC achievements to date are the introduction and 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, over 7,000 data sets from over 350 tissues and cell types were available including 198 complete and 847 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 in the following sections, integration of

epigenomic features with genetic variation (e.g. from the Catalog of published genome-wide association studies (GWAS Catalogue)), gene expression (e.g. from the Genotype–Tissue Expression (GTEx) project) and other data is tremendously useful to pinpoint candidate variants for functional analysis. However, one limitation hampering fast exploitation of these resources for functional and other analyses today 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 both in numbers and complexity. **Figure 3** illustrates this complexity of profiles and **Box 2** provides details on the individual marks and features profiled to date. These can be divided into the following categories: DNA modifications, which includes C5-methylcytosine (5mC), the first and best studied epigenomic modification discovered in 1948 [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 on the DNA level but this number is likely to increase in the future and analysis of their chemical and biological functions is subject of intense ongoing research [26, 27]. Histone modifications represent by far the largest category among profiled chromatin marks. With currently 12 known chemical modifications, which can occur at over 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 about 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 (e.g. 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 function. Interestingly, common trait-associated single-nucleotide polymorphisms (SNPs) identified through GWAS approaches frequently lie outside coding regions but fall into DNase1-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, due to technological limitations.

While more than 100 different RNA modifications are known [33], and new ones are continuously being discovered [34-36], 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 to enhancers and promoters 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 play functional roles [38].

Data integration and interpretation

A large variety of different marks and features have been profiled to date, 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. But without additional context, it remains difficult to predict from these descriptive data which of the large number of marks, features and profiles are most indicative for causal and quantitative effects (**Figure 2)**. Consequently, next-generation approaches for the integration and interpretation of chromatin features attract high interest. IHEC, for example, has identified a subset of nine required profiles and assays (**Figure 3**) for the generation of so-called reference epigenomes in order to bundle obtained epigenomic data and to maximize its potential to infer function. The rapidly growing number of reference epigenomes registered in EpiRR (see online links) 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 applied to 111 human epigenomes so far [39]. Typically, such integrated data sets consist of 20-50 genome-wide profiles making up a multi-dimensional data matrix as illustrated in **Figure 4**. 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: detection of low-quality data, inference of missing data, and, as a consequence, a more accurate and complete annotation and interpretation of epigenomes.

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 e.g. ChromHMM [41] or Segway [42, 43] which have been trained on a variety of datasets, resulting in chromatin state annotation with inferred function (including "promoter", "enhancer", "insulator", "transcribed" or "repressed") at a particular genomic locus. Several of such inferred enhancers 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]. Based on these and many more specialized tools such as Epigram, epiGRAPH, Epilogos, eFORGE, Epigwas, ChromNet and the Epigenetic Clock (see online links for details), 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 (termed P2F) presented here. Using integrative analysis as illustrated in **Figure 4**, the authors predicted the cell type and regulatory element (enhancer) in and through which a genetic variant 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 annotation generated from profiling 127 human cell types to predict the regulatory nature of the target region and the cell type in which this region would most likely 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 risk-allele and non risk-allele carriers to assess changes in gene expression. Finally, they restored 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- 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, first attempts following this strategy using a general pipeline are extremely encouraging but existing experimental and computational limitations as well as currently unknown challenges will have to be overcome as the field moves forward [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 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 orchestrated chromatin changes that occur during development and disease. Data integration enables these candidate sites to be reduced to a subset with inferred function (**Figure 2**). The experimental validation of their relevance, however, remains difficult. To some extent, genetic approaches have successfully provided evidence for the importance of chromatin marks. Here we will discuss two widely used approaches: genetic manipulation of the DNA domains underlying an epigenomic feature; and genetic manipulation of the enzymes responsible for their establishment or removal.

Genetic manipulation of sites of chromatin marks or features

Individual epigenomic features can be removed through manipulation of the underlying DNA sequence. While it is possible to mutate or delete single bases harbouring DNA modifications, this approach is not applicable to histone modifications or larger epigenomic features. In many cases, entire genomic domains containing the feature of interest have to be excised instead. Gene targeting has been an exceptionally successful approach to functionally link several epigenetic mechanisms (DNA methylation, chromatin insulation, noncoding transcription) to genomic imprinting [49-52]. However, in most cases, genetic manipulation only provides indirect evidence for causality (**Figure 2**), since functional consequences could

be attributed just as well to the loss of the genomic DNA sequence rather than to the loss of the epigenomic feature.

Today, genetic manipulation is often still the only available option for conducting functional experiments (e.g. Fanuci et al.[53]). Improved methods using targetable nucleases [54] have made it easier to 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** and **Figure 3**, local chromatin architecture can be profiled and segmented into topologically associating domains (TADs), whose boundaries appear 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 reduced expression of some of the associated genes [60]. Likewise, 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].

While in these above examples a handful of candidate features within small and well defined regions (imprinted domains, individual enhancers or single topological domains) were manipulated, most epigenomic profiles contain many hundreds or thousands of candidate marks distributed across the entire genome. A strategy for how genetic manipulation can nevertheless be used to interrogate many epigenomic candidate sites at once was recently introduced [61]. Korkmaz et al. integrated different published chromatin profiles (e.g. H3K4me2, H3K27ac) and transcription factor binding sites to generate a candidate list of active enhancers bound by p53. To reveal which of these enhancers are necessary for a specific function of p53, namely 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 they 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

While generally successful in attributing causal function to genomic domains hosting epigenomic marks and features, the aforementioned approaches cannot establish how large the contribution of epigenetics to the observed effects is. A second experimental strategy using genetic targeting, namely deletion or mutation of chromatin modifying enzymes, overcomes this drawback and thus was most instrumental in demonstrating their participation in gene regulation. Epigenetic model systems applicable to early embryogenesis or embryonic stem (ES) cell based experiments (amongst others: genomic imprinting and retro-transposon silencing) were especially useful to attribute crucial roles to chromatin modifications, since those allowed to study the acute effect that loss of certain chromatin marks (e.g. H3K9 methylation and DNA methylation) had 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 [66], Dnmt3a and Dnmt3b [67], G9a [68], Suv39h1 and Suv39h2 [69], Hdac1 [70], Ezh2 [71], SetDB1 [72] or LSD1 [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. Loss of expression of single proteins can easily trigger lethality (even early [74, 75]). Consequently, it is difficult to deduce 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 possess non-histone targets as well [76, 77]. Therefore, resulting embryonic phenotypes cannot always be attributed to misregulation of histone marks alone.

Embryonic lethality arising from germline depletion of genes can be circumvented by conditional knockouts. This strategy has been so successfully applied to chromatin modifying enzymes that they can be discussed here only incompletely. The well-studied hematopoietic [78-82], muscular [83] and cardiac [84, 85] systems are typical examples which have been used to show the crucial roles played by DNA methylation [78-80], H4K20 methylation [83], H3K27 methylation [81, 82, 84] and histone acetylation [85] in somatic stem cell homeostasis, lineage specification or progression. It should, however, also be mentioned that detected

phenotypes in such studies are hardly driven by de-regulation of a multitude of genes as they are often surprisingly specific (e.g. affecting only certain lineage choices [78, 81] or cellular phenotypes [83]) and in some cases have even been rescued by normalizing expression levels of single genes [82-84].

While these observations of focused transcriptional consequences and specific phenotypes can in many cases be explained by 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 large enough to cause cellular phenotypic changes and that those might strongly depend on the cellular context. Consistent with this is the notion that homeostatic cells often remain relatively unaffected from pharmacological inhibition of chromatin modifying enzymes, while many cancer cells show an enhanced "epigenetic vulnerability" (for a concise review see Dawson et al.[86]). Another example is the finding that, in vitro, even global loss of canonical epigenetic marks does not necessarily result in major transcriptomic changes. The (almost) complete loss of DNA methylation [87], H3K27me3 [88] or an artificial induction of H3K4me3 marks [89] in ES cells, for instance, results only in minor transcriptional changes, despite affecting their differentiating progeny. Taken together these data indicate that a majority of epigenetic marks may not play 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 or transformation. An alternative (but not mutually exclusive) explanation would be that many chromatin marks in the epigenome have opposing causal roles, which are often "canceled out" when marks are globally altered, implicating more pronounced phenotypes when the manipulation of chromatin modifications is restricted to few individual loci. In summary, genetic manipulation of chromatin modifying enzymes has been crucial to implicate their causal involvement in many biological processes; however the functional involvement of individual chromatin marks can mostly still only be inferred **(Figure 2)**.

Site-specific epigenome editing

Recruitment of chromatin-modifying enzymes to specific loci

The genetic experiments described above indicate that only a fraction of marks detected in epigenomic profiles may play direct causal roles (**Figure 2**). Consequently, new experimental approaches able to test the causality of individual epigenetic marks directly are in high demand. Several new approaches are currently emerging including the exploitation of naturally occurring or engineered histone mutations [90, 91] or 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 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 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 target sequence using zinc finger or TALE architectures involves assembling multiple repetitive protein domains that each recognise a particular DNA base in the target sequence. In contrast, the CRISPR–Cas system can be targeted to a precise genomic location by specifying the base sequence of a part of a synthetic RNA known as the guide RNA (gRNA). Remarkably, Cas9 is able to target genomic sites, even when those 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 targeting 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 testing whether individual chromatin marks have causal effects on gene expression (as illustrated in **Figure 5a**). In addition to more general 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 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 transcriptional up-regulation of this gene [95]. Similarly, a dCas9–p300 histone acetyltransferase fusion has been used to activate transcription of *MyoD* and *Oct4* from proximal promoters and distal enhancers. In many cases, induction of mRNA production achieved with dCas9–p300 is stronger than that achieved by a classical trans-activator 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] while 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 up- and down-regulate mRNA levels, providing direct evidence that chromatin modifier can regulate transcription. Whether the observed effects are exclusively mediated via 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 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 this it will be necessary to study how the catalytic activity of the chromatin modifier at a particular site impacts 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 the latter could manifest itself in several ways. Some epigenomic features might be dominant in their effect (e.g. affecting polymerase activities directly), while others might be dependent on certain pre-requisites to reveal a functional involvement (e.g. transcriptional priming: poising the cellular response spectrum by forming a transcription factor binding platform for example) [19, 102]. Thus, in some cases (and quite similar to most other biological mechanisms) the function of an epigenomic feature could depend on cell type, culture condition or 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 level are biologically relevant. However, it is encouraging that several studies have already achieved changes in protein level through epigenomic editing [98] [97] [103]. Ultimately, however, it will be important to test directly whether engineered chromatin modifiers can influence cellular or organismal phenotypes. Some reports made such a connection already showing for example that 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 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 reduce to background levels in vitro [104, 105] indicating they are actively or passively lost, while in another report, they were found to persist [106]. Since the targeted sites differed (and in the latter case 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 have to be examined with care. Expression of the targetable chromatin modifier, engineered modifications, transcriptional and phenotypic changes should be monitored over time.

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 ("silencing", "activating") to comprehensively quantify the contribution of individual marks in defining endogenous transcriptional states. For this, it would be useful to establish a hierarchical order of these marks i.e. to elucidate which functional chromatin marks are primary triggers (influencing other epigenomic features) and which are usually occurring 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, noncoding 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, a large number of different types of engineered chromatin marks should be compared directly (**Figure 5b**). To allow this, epigenomic editing approaches should not solely concentrate on re-assessing the importance of well-studied chromatin marks and instead include some of those marks and features less comprehensivly analysed to date (depicted in **Box 2**). This requires, however, the swift development of a series of new engineering tools. In yeast, the first important steps in this direction have already been taken [107]. Here, over 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, some as inhibitors, allowing subsequent ranking based on effect size, i.e. changes in protein expression in response to targeted chromatin modification [107]. Studies in mammalian cells have not been as comprehensive so far but some recent publications followed a similar strategy already to compare the effects of a series of chromatin domains [108] and modifications [106] on candidate genes in mammalian cells [94].

Such approaches will be the basis to elucidate the hierarchical order of chromatin marks. Targeting more than one chromatin modification to the same locus will allow to elucidate which marks are causing others to change and which are functionally dominant, antagonistic, additive or 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, due 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 (see for example [38, 49, 109, 110]). Today, strategies are emerging that can alter domain boundaries without affecting the underlying genomic DNA sequence. While targetable CTCF proteins have not yet been reported to date, a zinc finger–Ldb1 fusion has been used already, to target the inactive β-globin locus in an erythroblast cell line lacking GATA1 [111]. Binding of Ldb1 induced the formation of a chromatin loop between the promoter and a locus control region and was sufficient to activate 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 could have translational value for the treatment of sickle-cell anemia [111]

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 pressing questions in epigenetics today is how much of the genome will be controllable in a similar way. The availability of a large spectrum of epigenome editing tools could in the future be used to screen for individual epigenomic marks which are either necessary or sufficient for specific cellular phenotypes (**Figure 5c**). 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 the library design. For example, designing gRNA libraries focusing on informative regions identified from integrative analysis of epigenomic profiles would greatly reduce the complexity of libraries and thus enhance the power of such epigenomic screens.

As outlined in **Figure 5c** such screens will enable the identification of the subset of causal epigenetic marks amongst the many with inferred function. gRNA libraries representing loci identified in epigenomic profiles are introduced into cells expressing the relevant dCas9–modifier fusion. Those cells responding 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 behavior) and used to gain

information about the individual gRNAs those cells received. It is difficult to predict today which cellular phenotypes will be most susceptible for these approaches, however it is not unreasonable to start with those cellular phenotypes that are clearly epigenetic, easy and accurate to measure and ideally reversible (e.g. cell identity, cell cycle control, migration).

In principle, epigenetic screens are feasible already today. For example, instead of 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. On a small-scale such an epigenetic screen 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 de-methylation and abrogated transcription of associated genes[112], hence the authors then used this system to screen eight candidate enhancers to investigate the effect of H3K4 de-methylation on ESC self-renewal [112]. Scaling up such epigenome editing approaches to epigenome-wide screens has the potential to reveal many (and eventually all) epigenetic marks and features playing causal roles in a given cellular phenotype **(Figure 5c)**.

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 so far, and some of the profiling technologies used at the time have become outdated and/or evolved (e.g. epigenomics on the single cell level) [113] indicating a continued need for profiling efforts. Moreover, the toolset of efficient dCas9–chromatin modifier fusion proteins has to be expanded. There is, for example, to date no 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 [114, 115]. And finally, to make the most of the new molecular tools discussed in this review, reliable in vitro models or in vivo approaches that allow assessment (and selection) of induced phenotypic changes have to be developed [103]. Considering the rapid progression of CRISPR-based technology during recent years it is not inconceivable 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 showing the way towards functional analysis of epigenomic marks and features which remains to be one of the main challenges for epigenomics. Based on current evidence, the key innovation promising 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 out of the myriad of those with inferred function identified through epigenomic profiling and data integration. Returning to the analogy of an epigenetic landscape from which we started, epigenomic engineering promises to turn the static landscape depicted by Waddington into a dynamic environment as illustrated in the contemporary animation (**Figure 1B; Supplementary information S1 (movie))** created by the resident artists of EpiGeneSys (see online links), the European Network of Excellence for Epigenetics and Systems Biology. In the context of this review, the pulling of the strings represents the approaches discussed here to screen for chromatin marks that causally influence cellular fate in health and disease. Based on the tremendous progress made to date, it is perhaps not surprising that expectations are running high to translate any fledgling new insights already into novel medicines and treatments. While 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 similarly to those currently discussed for genome editing technology [116]. With these promising developments in mind, does this mean that epigenomic profiling is essentially completed and a thing of the past? Certainly not, since new marks and features are still being discovered and new or improved profiling technologies are still being developed. It will, however, be interesting to explore which type of profiling turns out to be most informative for which field of research and, in particular, for the discovery of causal functions hidden in chromatin.

Display items:

Box 1 | Summary of key early epigenetic experiments addressing function of DNA methylation.

Chromatin can be altered in a large variety of ways, but only a few chromatin features have been shown to functionally involved in gene expression. The chromatin mark to get first (and to date still most) attention is DNA methylation. Discovered in the late 1940s as a modification of cytosine bases [24, 117] it became early on a prime candidate for an epigenetic effector, because of its uneven distribution in the genome and its heritability [118-120]. First correlations between gene expression and DNA methylation were reported on a series of highly informative model loci (e.g. chicken and mammalian globin genes, the X-chromosome inactivation centre (XIC), genomic imprinting and virus, transgene or retrovirus silencing) [121-123]. However, it was only after the use of inhibition [124] or deletion of DNA methyltransferases [125], in vitro methylation of DNA [126] or genetic deletion of differentially methylated regions [51], that functional connections could be deducted. Because of epigenomics and transcriptomics we know today that the relationship between DNA methylation and gene expression is likely to be more complicated than initially suspected. While marks at certain positions correlate with silencing of some genes (e.g. in colon cancer) [127], others are rather uninformative or even occur at active genes [128-131].

Box 2 | Profile types and categories (including embedded Figure 3).

The number of epigenomic marks and features that can be profiled is inherently a moving target. Consequently, the profile types and 6 categories illustrated in **Figure 3** and explained here are incomplete and subject to change. The cartoon of chromatin depicts common marks and features that are further grouped into six boxed categories (see Glossary for detailed descriptions) listed on the right side. On the DNA level, modifications have been shown to occur at positions C5 or N3 on cytosines and at position N6 on adenines and to be catalysed either enzymatically by DNA methyltransferases (DNMTs) for 5mC or chemically for 3mC. The mechanism of modification for 6mA is still unknown. As part of an active demethylation pathway mediated by TET enzymes, 5mC can be further modified by stepwise oxidation to 5hmC, 5fC and 5caC, respectively. Because of their versatility, profiling of DNA modification can be configured for multiple readouts, including differentially methylated positions (DMPs) or regions (DMRs),

differentially variable positions (DVPs), partially methylated domains (PMDs) and blocks of comethylation (COMETs) [132, 133]. On the histone level, 12 enzymatically catalysed modifications (see **Figure 3**) have been shown to occur at over 130 post transcriptional modification sites (PMTs) at the tails of the four core histones (H2A, H2B, H3 and H4) and some 30 histone variants [134]. Recently, further modifications (H3K64ac and H3K122ac) were also observed in the globular domain of H3 and shown by epigenome editing to define a new class of enhancers [135] adding to the evidence that our current knowledge of epigenetic modifications is far from complete. Despite their extraordinary variety, histone modifications are profiled essentially by a single assay and readout (ChIP-seq). On the feature level, we distinguish three categories of structurally different features for which multiple profiling assays (see Glossary) have been developed depending on the complexity of the readout. For profiling nucleosome occupancy, DNase1 footprinting has been the assay of choice in the past to identify cell-type-specific regulatory elements but ATAC-seq [136] is now becoming increasingly popular due to its simpler work flow and its ability to work with substantially fewer cells. Different implementations of chromatin conformation capture assays (see Glossary) are being used to connect enhancers to the promoters 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 HiC [137] allows segmentation of the epigenome into a variety of chromatin domains (e.g. Topologically associated domains (TADs)). On the transcriptome level, a multitude of noncoding RNAs have been mapped, which in some cases can regulate gene expression [138], while profiling of RNA modifications (for example occurring at positions C5 or N3 on cytosines and at positions N1 and N6 on adenines) has not yet been systemically analyzed. In the upper left panel, base positions where methylation has been found to date in DNA or RNA are marked in orange. 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 9 core marks (shown in the lower left panel) have been consistently profiled are those also referred to as IHEC reference epigenomes.

Figure 1 | Evolving views on the Waddington landscape.

(A) Epigenetic landscape as depicted by C.H. Waddington. In this analogy epigenetics influences cellular fate during development analogous to gravitational forces on a defined landscape. (B) Contemporary version of the Waddington landscape depicting epigenome editing (see also **Supplementary information S1 (movie)**). Epigenetic and epigenomic manipulation promises to dynamically change the landscape and thus, cellular phenotypes.

Figure 2 | Schematic indicating which experimental approaches are targeting which level of function. While epigenomic profiling alone results in descriptive information, integration of multiple layers of information allows to infer function. Genetic manipulation can reveal relevance of epigenetic features, however only indirectly. Currently used methods for epigenetic manipulation of single marks or features can reveal causality, future approaches, will enable us to identify novel functional marks in an epigenomic scale (epigenetic screens).

Figure 3 | Complexity of epigenomic profiling (integrated in Box 2).

Figure 4 | Schematic of multi-dimensional profile integration.

Integration is achieved in two steps. First, missing data are imputed using profiles from the same (vertical dashed box) and/or closely related samples (horizontal dashed box). Currently, histone modification and transcription factor (TF) binding profiles are mostly used for imputation but there is no reason why other profiles cannot be used as well. Second, additional, non-epigenomic data (not discussed here but equally important) can be added as appropriate and the entire data set per sample aggregated and segmented into chromatin states, ranging from 2 (as illustrated here) to >50 states, depending on context and complexity [139] . While the majority of current epigenomic data has been derived from healthy samples, data from diseased samples and further integration with muliti-omics data and pathways can be expected to follow soon e.g. as part of the recently established H2020 MultipleMS and SYSCID Consortia focusing on multiple sclerosis and chronic inflammatory diseases, respectively. For more details on the current strategy please refer to the FTO example given in the section on '*Data integration and interpretation'*.

Figure 5 | Strategies for epigenome editing.

(A) A chromatin modifier (or its minimal catalytic domain) is fused to a targetable DNA binding domain (here showing 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 investigating how editing of a mark at a particular site affects the expression of associated genes. Read-outs can be at the level of RNA, protein or phenotype as illustrated. (B) Schematic illustrating how targeted epigenome editing can be used to build a hierarchy of functional marks. A range of different chromatin modifiers 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. (C) Using epigenome editing allows identifying chromatin modifications impacting cellular phenotypes. Prior knowledge about the genes involved in regulating particular phenotypes and the location of regulatory elements can be used to design pooled gRNA libraries enabling targeted screens. If particular gRNAs target the chromatin modifier to functional sites, phenotypic changes are induced, allowing selection and determination of gRNA sequences.

Table 1 | Epigenetic modifiers used in epigenomic editing so far, their present use and effect.

REPRESSION			GENE ACTIVATION					
LSD1	LSD1	32 repressive histone effector or recruiter domains	TET1	TET1	TET1	p300	PRDM9	chromatin modifying enzyme
histone H3K4 demethylase	histone H3K4 demethylase	HDAC8 (Xl): histone deacetylase, NUE: histone	DNA demethylase	DNA demethyla se	DNA demethylase	HAT	K4 methylase	Function
full length	full length	HDAC8 (Xl): AA 1-325, NUE (Ct): FL, SET8 (Tg): AA 1590-1893, KYP: AA 1-	CD (amino acids 1418-2136)	CD	CD	CD (amino acids 1,048 - 1,664)	CD	Full length or Catalytic domain (CD)
dCas9	TALE	TALE	TALE	dCa9	dCa9	dCas9, ZF, TALE	dCas9, ZF	targeted via
enhancers	enhancers	promoters	promoters	promoters	promoters	promoters and enhancers	promoters	targeted to
up to 85% H3K4me2 loss, >90% loss of H3K27ac	65 % loss of H3K4me2 and 60% loss of H3K27ac (relative to TALE alone and	KYP: increased H3K9me1 (ca 1.4 fold), SET8: increased	DNA demethylation, 10 - 80 % decrease in methylation levels	DNA demethyla tion variable	DNA demethylatio n, 10 - 50 % decrease in	increased H3K27ac (up to 10 fold increase relative to D1399	Increase of H3K4me3 (up to 60%)	observed modification and effect size
>90% loss of mRNA	up to 50 % decrease in RNA level	up to 50-75 % decrease in RNA level	increase in transcription ~2-1,000 fold relative to off target	increase in transcripti on up to 10 fold	increase in transcription up to 2.5 fold	increase in transcription (10 - 10,000 fold increase in RNA	Upregulation og EpCAM (up to 8-fold)	effect on transcription and size of effect
ES cell morphology changes	NA	NA	NA	reduction in cell proliferati	reduction in cell proliferation	NA	NA	effect at the level of phenotype
Oct4 distal enhancer, 8 candidates enhancers thought to regulate pluripotency in	candidate enhancer in SCL (stem cell leukaemia) locus and 40 additional candidate enhancers, effect on	Grn2, NeuroG2	HBB, KLF4, RHOXF2	RANKL, MAGEB2, MMP2	BRCA1	promoters: IL1RN, MYOD, OCT4, β-globin (HBE, HBG), ICAM1 enhancers:	Promoters: ICAM1, RASSF1a, EpCAM, PLOD2	locus tested
mouse ES cells	K562	primary neurons, Neuro2a cells,	K562, HEK293, HeLa	HeLa, 293T	HeLa, MCF7	HEK293T	HEK293, A549	Model system/Cell lines
97	14	108	95	142	141	96	140	References

DNMT3a	DNA methyltransferase	DNMT3a: AA 602-912	dCas9	Promoters	increased DNA me	40-50% downregulation	NA	IL6ST, BACH2	HEK293	104
DNMT3a-DNMT3L	DNA methyltran sferase	Dnmt3L: C terminal domain, Dnmt3a	ZFP, TALE	CpG island	increased DNA me	40-60% downregul ation,	increased proliferatio	VEGF-A, p16	SKOV3 cells, HeLa cells,	99
DNMT3a	DNA methyltra nsferase	DNMT3a: AA 598-908	ZFP	Promoter	increased DNA me	60% downregu lation (RNA)	protein (up to 80%)	MASPIN, SOX2	SUM159 cells, MCF7	98
G9a	HMT	AA:829-1210	ZFP	Promoters	increased H3K9 methylati on (up to	40% loss of mRNA	NA	VEGF	HEK293	143
SUV39H1	HMT	full length and shorter constructs	ZFP	Promoters	increased H3K9 methylation (up to 2.8 fold)	40% loss of mRNA	NA	VEGF	HEK293	143

Supplementary movie file 1 | Contemporary version of the Waddington landscape (see also **Figure 1**). Epigenetic and epigenomic manipulation promises to dynamically change the landscape and thus, cellular phenotypes.

GLOSSARY TERMS:

COMET:

Blocks of comethylation identified by methylome segmentation

Pooled screens:

Approaches in which cells receiving the screening library (e.g. gRNA pools) are grown and selected together for a phenotypic change.

DNase1-hypersensitive sites:

Regions of chromatin that is sensitive to digestion with DNase1, indicating these are accessible and free of nucleosomes.

FAIRE-seq:

Formaldehyde Assisted Isolation of Regulatory Elements followed by sequencing is a technique that uses the solubility of open chromatin in the aqueous phase during phenol-chloroform extraction to identify sites of open chromatin.

ATAC-seq:

Assay for transposase-accessible chromatin sequencing is 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.

ChIP-seq:

Chromatin immunoprecipitation followed by sequencing. A method for mapping the distribution of histone modifications or chromatin-associated proteins or transcription factors along the genome. DNA and protein are cross-linked 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.

GWAS:

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

SNP:

Single nucleotide polymorphism is a single base-pair difference in the DNA sequence of individual members of a species.

TAD:

Topologically associated domains are regions of chromatin in which loci frequently interact with each other, based usually based on evidence from Chromosome conformation capture techniques. Loci located in different TADs do not come into contact frequently.

Chromosome conformation capture:

A group of techniques (3C, 4C, 5C, HiC, ChiAPet) that are used to map physical interactions between segments of DNA in three-dimensional space.

HiC:

Experimental method to map contacts formed between segments of DNA in 3 dimensional space on a genome-wide scale.

TALE:

Transcription activator-like effectors are DNA-binding proteins that have a modular architecture with each module (~34 amino acids) recognising a single nucleotide in a DNA sequence and therefore can be engineered to bind to a DNA sequence of choice.

Zinc-finger:

Modular DNA binding protein that can be engineered to bind to a sequence of choice

CRISPR–Cas9:

Clustered regularly interspaced short palindromic repeats–CRISPR-associated protein 9 are components of a bacterial defense system against viruses.

Cas9:

Cas9 stands for CRISPR-associated protein 9. It is useful for genome engineering because it can be guided to a particular site in the genome where it makes a double-strand break into DNA.

dCas9:

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

gRNA:

The gRNA is an artificial fusion of CRISPR (clustered regularly interspaced short palindromic repeat) RNA (crRNA) and transactivating crRNA (tracrRNA) used to target the Cas9 protein to a target site in the genome.

FACS:

Fluorescence-activated cell sorting is an experimental method to measure a fluorescence-based signal (from a reporter or antibody staining) emitted from individual cells of a population that can also be used to isolate single cells of interest.

References :

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-54 (2003).
2. Waddington, C.H. The epigenotype. 1942. *Int J Epidemiol* **41**, 10-3 (2012).
3. Waddington, C.H. The strategy of the genes; a discussion of some aspects of theoretical biology (Allen & Unwin, London,, 1957).
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**, 396-8 (2007).
8. Henikoff, S. & Greally, J.M. Epigenetics, cellular memory and gene regulation. *Curr Biol* **26**, R644-8 (2016).
9. Taudt, A., Colome-Tatche, M. & Johannes, F. Genetic sources of population epigenomic variation. *Nat Rev Genet* **17**, 319-32 (2016).
10. Zhao, Y. & Zhou, D.X. Epigenomic modification and epigenetic regulation in rice. *J Genet Genomics* **39**, 307-15 (2012).
11. Ha, M. Understanding the chromatin remodeling code. *Plant Sci* **211**, 137-45 (2013).

12. Schmitz, R.J. & Ecker, J.R. Epigenetic and epigenomic variation in Arabidopsis thaliana. *Trends Plant Sci* **17**, 149-54 (2012).
13. Law, J.A. & Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**, 204-20 (2010).
14. Mendenhall, E.M. et al. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* **31**, 1133-6 (2013).
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**, 159-71 (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**, 613-26 (2012).
20. Mueller-Planitz, F., Klinker, H. & Becker, P.B. Nucleosome sliding mechanisms: new twists in a looped history. *Nat Struct Mol Biol* **20**, 1026-32 (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**, 315-32 (1948).
25. Wu, T.P. et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532**, 329-33 (2016).
26. Booth, M.J., Raiber, E.A. & Balasubramanian, S. Chemical methods for decoding cytosine modifications in DNA. *Chem Rev* **115**, 2240-54 (2015).
27. Balasubramanian, S. Chemical biology on the genome. *Bioorg Med Chem* **22**, 4356-70 (2014).
28. Tan, M. et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-28 (2011).
29. Khare, S.P. et al. HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res* **40**, D337-42 (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**, 703-8 (2014).
32. Maurano, M.T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
33. Li, S. & Mason, C.E. The pivotal regulatory landscape of RNA modifications. *Annu Rev Genomics Hum Genet* **15**, 127-50 (2014).
34. Dominissini, D. et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441-6 (2016).
35. Machnicka, M.A. et al. MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res* **41**, D262-7 (2013).
36. Cantara, W.A. et al. The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* **39**, D195-201 (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**, 1012-25 (2015).
39. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
40. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* **33**, 364-76 (2015).
41. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
42. Hoffman, M.M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9**, 473-6 (2012).
43. Hoffman, M.M. et al. Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* **41**, 827-41 (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**, 488-96 (2016).
45. Claussnitzer, M. et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* **373**, 895-907 (2015).
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**, 381-6 (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**, 1357-63 (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 U S A* **103**, 10684-9 (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**, 889-93 (2004).
51. Wutz, A. et al. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745-9 (1997).
52. Latos, P.A. et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338**, 1469-72 (2012).
53. Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M.S. & Mhlanga, M.M. Chromosomal contact permits transcription between coregulated genes. *Cell* **155**, 606-20 (2013).
54. Kim, H. & Kim, J.S. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* **15**, 321-34 (2014).
55. Zhou, H.Y. et al. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev* **28**, 2699-711 (2014).
56. Groschel, S. et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369-81 (2014).
57. Webster, D.E. et al. Enhancer-targeted genome editing selectively blocks innate resistance to oncokinase inhibition. *Genome Res* **24**, 751-60 (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**, 165-70 (2016).
59. Canver, M.C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-7 (2015).

60. Guo, Y. et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-10 (2015).
61. Korkmaz, G. et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* **34**, 192-8 (2016).
62. Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* **55**, 277-90 (2014).
63. Martens, J.H. et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *Embo J* **24**, 800-12 (2005).
64. Branco, M.R., Oda, M. & Reik, W. Safeguarding parental identity: Dnmt1 maintains imprints during epigenetic reprogramming in early embryogenesis. *Genes Dev* **22**, 1567-71 (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**, 3565-75 (2006).
66. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26 (1992).
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**, 247-57 (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**, 1779-91 (2002).
69. Peters, A.H. et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-37 (2001).
70. Lagger, G. et al. Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J* **21**, 2672-81 (2002).
71. O'Carroll, D. et al. The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol* **21**, 4330-6 (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**, 2478-86 (2004).
73. Wang, J. et al. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet* **41**, 125-9 (2009).
74. Bradley, A. et al. The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm Genome* **23**, 580-6 (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**, D726-36 (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**, 1207-15 (2009).
79. Lee, P.P. et al. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15**, 763-74 (2001).
80. Shaknovich, R. et al. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* **118**, 3559-69 (2011).
81. Su, I.H. et al. Polycomb group protein ezh2 controls actin polymerization and cell signaling. *Cell* **121**, 425-36 (2005).
82. Su, I.H. et al. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat Immunol* **4**, 124-31 (2003).

83. Boonsanay, V. et al. Regulation of Skeletal Muscle Stem Cell Quiescence by Suv4-20h1-Dependent Facultative Heterochromatin Formation. *Cell Stem Cell* **18**, 229-42 (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**, 343-7 (2012).
85. Montgomery, R.L. et al. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev* **21**, 1790-802 (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**, 575-9 (2015).
88. Riising, E.M. et al. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55**, 347-60 (2014).
89. Schmitz, S.U. et al. Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J* **30**, 4586-600 (2011).
90. Lewis, P.W. et al. Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science* (2013).
91. McKay, D.J. et al. Interrogating the function of metazoan histones using engineered gene clusters. *Dev Cell* **32**, 373-86 (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**, 823-6 (2015).
94. Chavez, A. et al. Comparison of Cas9 activators in multiple species. *Nat Methods* (2016).
95. Maeder, M.L. et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**, 1137-42 (2013).
96. Hilton, I.B. et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* (2015).
97. Kearns, N.A. et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* **12**, 401-3 (2015).
98. Rivenbark, A.G. et al. Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* **7**, 350-60 (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**, 479-91 (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**, 1324-43 (2013).
103. Heller, E.A. et al. Locus-specific epigenetic remodeling controls addiction- and depression-related behaviors. *Nat Neurosci* **17**, 1720-7 (2014).
104. Vojta, A. et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* (2016).
105. Kungulovski, G. et al. Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenetics Chromatin* **8**, 12 (2015).
106. Bintu, L. et al. Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720-4 (2016).

107. 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**, 110-20 (2014).
108. Konermann, S. et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-6 (2013).
109. Tschopp, P. & Duboule, D. A genetic approach to the transcriptional regulation of Hox gene clusters. *Annu Rev Genet* **45**, 145-66 (2011).
110. Splinter, E. et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20**, 2349-54 (2006).
111. Deng, W. et al. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849-60 (2014).
112. Kearns, N.A. et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* (2015).
113. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat Rev Genet* **16**, 716-26 (2015).
114. Köferle, A. et al. CORALINA: A simple method for the generation of gRNA libraries covering complete genomes (in revision).
115. Lane, A.B. et al. Enzymatically Generated CRISPR Libraries for Genome Labeling and Screening. *Dev Cell* **34**, 373-8 (2015).
116. Lanphier, E., Urnov, F., Haecker, S.E., Werner, M. & Smolenski, J. Don't edit the human germ line. *Nature* **519**, 410-1 (2015).
117. Wyatt, G.R. Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochem J* **48**, 581-4 (1951).
118. Scarano, E. The control of gene function in cell differentiation and in embryogenesis. *Adv Cytopharmacol* **1**, 13-24 (1971).
119. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226-32 (1975).
120. Riggs, A.D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14**, 9-25 (1975).
121. McGhee, J.D. & Ginder, G.D. Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature* **280**, 419-20 (1979).
122. 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**, 569-85 (1978).
123. Jahner, D. et al. De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* **298**, 623-8 (1982).
124. 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 U S A* **81**, 2806-10 (1984).
125. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362-5 (1993).
126. Busslinger, M., Hurst, J. & Flavell, R.A. DNA methylation and the regulation of globin gene expression. *Cell* **34**, 197-206 (1983).
127. 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**, 178-86 (2009).
128. Metivier, R. et al. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45-50 (2008).
129. Kangaspeska, S. et al. Transient cyclical methylation of promoter DNA. *Nature* **452**, 112-5 (2008).

130. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141-3 (2007).
131. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-22 (2009).
132. Libertini, E. et al. Information recovery from low coverage whole-genome bisulfite sequencing. *Nat Commun* **7**, 11306 (2016).
133. 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).
134. 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**, 259-71 (2014).
135. Pradeepa, M.M. et al. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet* **48**, 681-6 (2016).
136. 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 1-9 (2015).
137. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
138. 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).
139. Baker, M. Making sense of chromatin states. *Nat Methods* **8**, 717-22 (2011).
140. Cano-Rodriguez, D. et al. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nat Commun* **7**, 12284 (2016).
141. Choudhury, S.R., Cui, Y., Lubecka, K., Stefanska, B. & Irudayaraj, J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* (2016).
142. Xu, X. et al. A CRISPR-based approach for targeted DNA demethylation. *Cell Discov* **2**, 16009 (2016).
143. 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**, 2159-66 (2002).

Acknowledgements:

SHS was supported by DFG (STR 1385/1-1). AK was supported by a CRUK PhD Fellowship. SB was supported by the EU-FP7 BLUEPRINT Project (282510) and the NIHR UCLH Biomedical Research Centre (BRC84/CN/SB/5984).

FURTHER INFORMATION:

*HI*stome: http://www.actrec.gov.in/histome/

Modomics: http://modomics.genesilico.pl/

RNA mDB: http://mods.rna.albany.edu/

ENCODE: https://www.genome.gov/10005107/encode-project/

modENCODE: http://www.modencode.org/

Resident EpiGeneSys artists: http://www.epigenesys.eu/en/science-and-you/art-and-science

IHEC: http://ihec-epigenomes.org/

4D Nucleome: http://www.4dnucl.org/

GTEx: https://www.genome.gov/27543767/genotypetissue-expression-project-gtex/

GWAS Catalogue: https://www.ebi.ac.uk/gwas/

Epigram: http://compbio.mit.edu/epilogos/

epiGRAPH: http://epigraph.mpi-inf.mpg.de/WebGRAPH/

Epilogos: http://compbio.mit.edu/epilogos/

eFORGE: http://eforge.cs.ucl.ac.uk/

Epigwas: https://www.broadinstitute.org/mpg/epigwas/

ChromNet: http://chromnet.cs.washington.edu/

TargetFinder: https://github.com/shwhalen/targetfinder/

Epigenetic Clock: https://labs.genetics.ucla.edu/horvath/dnamage/

EpiRR: http://www.ebi.ac.uk/vg/epirr