Review

Epigenetic memory: A macrophage perspective

Colin Logie a, b, *, Hendrik G. Stunnenberg a, b

a Department of Molecular Biology, Faculties of Science and Medicine, Nijmegen Centre for Molecular Life Sciences, Radboud University, 6500 HB Nijmegen, The Netherlands
b NCMLS, FNWI, Radboud University Nijmegen, The Netherlands

A R T I C L E   I N F O

Article history:
Received 9 February 2016
Received in revised form 16 June 2016
Accepted 23 June 2016

Keywords:
Memory
Monocyte
Macrophage
Nucleosome
Transcription factor
Epigenetics

A B S T R A C T

The molecular basis of cellular memory is a fascinating topic that progressed with great strides during the last few decades. In the case of cells of the immune system, cellular memory likely extends beyond cell fate determination mechanisms, since immunity can tailor its responses to a potentially hostile environment that is a priori variable if not unpredictable. One particularly versatile innate immune system cell type is the macrophage. These phagocytes occur in all organs and tissues as resident cells or as differentiation products of recruited circulating blood monocytes. They come in many flavours determined by the tissue of residence and by external factors such as microbes. Recently, macrophage epigenome profiling has revealed thousands of chromosomal loci that are differentially active in macrophages, revealing chromosome elements that drive macrophage gene expression. The most dynamic epigenomic mark is nucleosomal histone acetylation. This mark is found at gene promoters and enhancers and correlates very well with gene expression changes. A second mark is H3K4me3, which sharply decorates the promoters of most protein coding genes that are (potentially) expressed. H3K4me3 at promoters is surrounded by its precursor H3K4me1. However, most often H3K4me1 occurs without H3K4me3 at enhancers where it appears together with histone acetylation, but can persist long after acetylation decreased. Hence, the biochemical signal H3K4me1 embodies appears to be a key to the plasticity of macrophage gene expression potential.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Contents

1. Introduction .......................................................................................................................... 360
2. From chromatin to chromosome domains ...................................................................... 360
3. Epigenetic mechanisms ..................................................................................................... 360
  3.1. Chromatin remodelling ............................................................................................... 360
  3.2. Chromatin colours ...................................................................................................... 361
  3.3. Transcription initiation control .................................................................................. 361
  3.4. Transcription control elements on chromosomes as sites of molecular memory storage ......................................................... 361
4. Macrophage epigenomes .................................................................................................. 362
5. Conclusions and perspective ........................................................................................... 363
Acknowledgements ............................................................................................................. 363
References ............................................................................................................................. 363

* Corresponding author at: Department of Molecular Biology, Faculties of Science and Medicine, Nijmegen Centre for Molecular Life Sciences, Radboud University, 6500 HB Nijmegen, The Netherlands.
E-mail address: c.logie@science.ru.nl (C. Logie).

http://dx.doi.org/10.1016/j.smim.2016.06.003
1044-5323/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1. Introduction

The development of an adult organism from a zygote remains one of the most wondrous biological phenomena. Over a century of scientific research indicates that the molecular basis of this extraordinary feat is coded by the respective organism’s DNA genome. The intricate ballet of cellular migrations that occurs during early embryogenesis produces functional cells in all organs within one body with very high fidelity. This implies that very high levels of determinism are ultimately encoded in the DNA sequence. Clearly, the genome harbours a detailed blueprint to make and maintain a body. The question we will address here is how chromosomes store information such as ‘memories’, with a focus on macrophages.

2. From chromatin to chromosome domains

How does a differentiating monocyte acquire and maintain its macrophage (Mf) identity at the hand of a DNA code common to every other cell? Part of the answer to this question lies in the 46 human chromosomes and their biochemical milieu, namely chromatin. The diploid human genome contains ∼6 × 10^9 bp. This represents about 2 m of DNA, which fit in a cell nucleus that is 10–20 μm in diameter. Except during cell division, when chromosomal DNA is highly compacted to enable sister chromatid segregation to the daughter cells, each chromosome occupies a distinct territory within the nucleoplasm and intermingles little with other chromosomes [1,2].

Chromosomal DNA is packaged in arrays of nucleosomes. These are discs, ∼11 nm in diameter and 5.5 nm in height, that consist of an octamer of histone proteins (H3/H4)(H2A-H2B) and ∼147 bp of DNA [3]. The length of the linker DNA between any two nucleosomes varies from less than 10 to over 90 bp, with dominant nucleosome repeat lengths of 190–200 bp in human blood cells [4]. In fact, the nucleotide composition of DNA can help to rotationally position DNA on the nucleosome surface through phased TA, TT, AA and GC dinucleotides positioned every 10/11 bp or multiples thereof, suggesting that the DNA strands’ sequences themselves can influence nucleosomal DNA packaging [5–7]. Because nucleosomes can stack in various solenoidal structures, some of which are favoured by short or long linker lengths [8], arrays of nucleosomes can form higher order structures that are often modelled as one- or two-start solenoids of polynucleosomal fibres [9] that are ∼30 nm wide [10].

Some 20,000 protein coding genes are distributed along the length of the human chromosomes. Every gene has one or more promoters from which transcription of mRNAs by RNA polymerase II (RNAPII) can start [11]. Furthermore, distal elements called ‘enhancers’ modulate the activity of the promoters [12,13]. Different cell types activate gene expression differentially by utilizing different parts of the genome’s regulatory repertoire of cis-acting elements. The mechanisms underlying the differential use of the genome’s regulatory elements involve physical changes in chromatin that promote or inhibit gene expression [14–16].

From a regulatory perspective, the genome is partitioned in several thousand independently functioning ‘topologically associated domains’ (TADs). These have been defined using formaldehyde-fixed cellular chromosome restriction cleavage followed by ligation, allowing interacting chromosome segments to be detected as ligation products that are amenable to next-generation DNA sequencing [17]. TADs are relatively invariant, even across mammalian orders such as primates and rodents [18], although it is possible that TADs vary in conformation during ontogeny [19]. An early study reported ∼2000 TADs ranging in size from 80 kb to 10 Mb [18]. Considering that DNA sequences located within one TAD are crosslinked significantly more often than segments located in adjacent TADs, the current model is that genes within one TAD can be influenced by cis-acting regulatory elements within that TAD through DNA looping, while genes outside the TAD are much less likely to be contacted [20]. Much research indicates that a protein ring consisting of cohesin works together with a very tight DNA binding protein called CTCF to establish TAD boundaries, presumably by forming the basis of stable chromatin loops [21–24]. It is thought that they act as ‘boundaries’ because they can interrupt or prevent enhancer-promoter contacts [25]. Intriguingly, inter-chromosomal clustering of a select number of TADs marked by the repressive H3K27me3 histone modification has been documented recently in embryonic stem cells [26–28]. Furthermore, some TADs correspond to nuclear lamin associated domains [29]. The biophysical properties of topologically associated chromatin domains remain to be elucidated.

3. Epigenetic mechanisms

Currently, epigenetic mechanisms encompass post-transcriptional molecular systems, such as production of microRNAs, which can regulate mRNAs in trans through base pairing, leading to translation inhibition or mRNA degradation [30]. However, most epigenetic mechanisms described to date implicate protein-DNA interactions that affect gene expression at the level of mRNA synthesis, either by chemically modifying DNA bases, as is the case for 5-methyl cytosine, or by recruiting protein complexes involved in nucleosome modification and remodelling that modulate chromatin structure to promote or inhibit transcription initiation or elongation. The concept of epigenetics as formulated by Conrad Waddington in 1942 for ontogenesis [31], by Mary Lyon for X chromosome inactivation in female eutherians [32] and by Arthur Riggs [33] and Robin Holliday [34] for 5-methyl cytosine suggests that epigenetic mechanisms are involved in establishing heritable chromatin states [35]. Hence, epigenetics could be said to deal with mechanisms to establish durable decisions, while gene control may also describe more punctual fluctuations in transcription levels. Still, the distinction between transcription control and epigenetic control of gene expression is probably a semantic one when considering the question at the molecular level since nucleosomes are involved in both cases [36]. DNA-bound histones in the form of nucleosomes are ideally suited to perform molecular memory-related tasks because their residence time on DNA can be of the order of many hours [37–39].

3.1. Chromatin remodelling

Nucleosomal histones can recruit proteins that harbour histone binding domains [40–42]. When those interactions are promoted or disrupted by post-translational modification of histone residues such as lysines and arginines the function of the underlying DNA can change [43–48]. The protein domains that recognise modified histones are commonly called ‘readers’ of the histone modifications, and by analogy the enzymes that deposit and remove the modifications are called ‘writers’ and ‘erasers’. Crucially, next to covalent histone modification, several sub-types of SNF2-type ATP-dependent nucleosome remodelling enzymes are involved in inserting and extracting histone H2A variants into existing nucleosomes [49–51], sliding nucleosomes along DNA [52–54], and in the transfer of nucleosomes from one DNA segment to another [55]. Similar to histone code ‘writers’, some of the SNF2-family members reside in large multi-subunit protein complexes that not only bind nucleosomes to remodel them, but also harbour ‘reader’ domains that target their activity to specifically modified nucleosomes. A first complicating aspect of the writer/reader system of histone modifications is that ‘reader’ and ‘writer’ domains often co-occur in the same protein complex. A second complicating
aspect is that the ‘histone writers’ also modify non-histone factors. In fact, the lysine and arginine residues of readers and readers are themselves often subject to modification [56–59].

Biochemically, writers such as methylases, acetylases, kinase, ubiquitylases and other protein post-translational modification enzymes are antagonised by erasers such as demethylases, deacetylases, deubiquitylases, etc. … Furthermore, when one histone lysine is the target of both acetylation and methylation, as is the case for histone H3 lysines 9 and 27 (H3K9 and H3K27), it will take at least two enzymatic steps to move from the repressive methylated state to the activating acetylated state (or vice versa) as these lysine modifications are mutually exclusive. Thus, it is thought that the status quo between opposing nucleosome-modifying activities is subject to alterations by sequence-specific DNA binding transcription factors that recruit specific chromatin modifiers and remodelers to individual loci, tipping the balance towards more or less transcription [60].

3.2. Chromatin colours

At this point it is relevant to indicate that much of the genome is transcribed at a low level in every cell population. This finding has been replicated many times, most conclusively by next generation sequencing which demonstrated that indeed, most genes produce mRNAs, albeit at a low level. For instance, in monocytes and Mφs, ~25% of the ~20,000 protein coding genes are detectable at a very low level (>0.01 Reads Per Kilobase per Million sequence reads), while about 50% of protein coding genes achieve expression levels in excess of 1 RPKM [61]. Hence the majority of transcription units is potentially transcribed and it is therefore more a matter of how often transcription can take place rather than whether it takes place at all. A black and white, open or closed, view of transcriptional control is not consistent with the body of evidence. In fact, multiple ‘colours’ of chromatin have been described [62,63].

An example of a self-sustaining and even amplifying ‘chromatin colour’ was initially described for the constitutive heterochromatin mark H3K9me3, which is associated with gene repression [41]. The concept is that the reader of this modification can recruit the writer, forming a positive feedback loop that allows spreading of the mark along the length of the chromosome, as had been described in a process called position effect variegation whereby DNA loci translocated next to heterochromatin domains eventually are invaded by the heterochromatin mark and therefore durably silenced [64]. A more recent study provided support for a similar self-sustaining pathway for facultative heterochromatin mark H3K27me3; the H2A ubiquitylation activity within the H2K27me3 PRC1 ‘reader’ complex was shown to subsequently promote H3K27me3 deposition by the PRC2 ‘writer’ complex [65]. A third chromatin ‘colour’ is found at gene promoters, which are decorated with H3K4me3. This modification depends on H2BK120ub1 which is itself erased during transcription elongation by the de-ubiquitylation module of the SAGA H3/H2B acetyltransferase complex [66–70].

Hence, nucleosomal histones can carry modifications that favour repression (H3K9me3, H3K27me3) or activation (H3K4me1, H3K4me3, acetylation) and these recruit ‘readers’ and ‘writers’ that induce a particular chromatin state at a given locus, participating in cellular signal transduction towards the genome.

3.3. Transcription initiation control

We currently know that mRNA transcription initiation involves histone modifiers, readers and remodelling complexes, general transcription factors, and RNAPII interacting factors, which together encompass more than 300 nuclear proteins [71]. Much research indicates that transcription initiation sites are highly defined. Moreover, alternative initiation site usage for a gene often correlate with cell lineage and environmental stimuli [11]. Together with general RNA polymerase II transcription factors, the chromatin remodelling enzymes coordinate nucleosome-free regions and participate in the chromatin transactions that take place as transcription initiation proceeds towards elongation. The nucleosome-free regions at transcription start sites range from 70 to 500 bp in size and usually harbour flanking +1 and −1 nucleosomes that are decorated with the H2A variant H2AZ and harbour H3K4me3 [72–74]. Notably, H3K4me3 can also be detected at active enhancers [75,76] while H2AZ is found at most accessible DNA regions of the genome [77].

In humans, CpG islands constitute ~70% of all gene promoter sites. CpG islands attract H3K4me3 deposition by binding to CFP1, which is an integral subunit of H3K4 trimethylase complexes [78–80]. On the other hand, H3K4me1, which decorates most promoter-flanking nucleosome arrays and many (potentially active) enhancers, is the product of a second set of H3K4 methylase complexes, namely MLL3 and MLL4 harbouring complexes [80–82].

Recent profiling of the genome-wide occupancy of eleven mouse SNF2-type nucleosome remodelling enzymes in embryonic stem cells indicates that not less than six of these localise to the nucleosome-free regions of most transcription start sites that display H3K4me3 on their flanking nucleosomes, although in three cases occupancy was not much correlated with actual gene expression levels [74]. Ep400, which does correlate with transcription activity, modulates H2A–H2AZ variant substitution and is part of a very large protein complex that also harbours the Tip60/Kat5 histone H4/H2A acetyltransferase [83–86]. Chd4, whose occupancy does not correlate directly with transcription levels, is a subunit of the NURD complex that also contains a pair of histone deacetylases (HDAC1 and HDAC2) alongside multiple ancillary subunits [56,87–89]. It remains to be determined in what sequence all these SNF2 ATPase come into play at gene promoters.

The ENCODE project performed chromatin immuno precipitation of sequence-specific transcription factors and showed that they constantly scan nucleosome free regions [90]. At nucleosome free regions, activation and repression domains of the transcription factors recruit or activate a diverse set of enzymes. Commonly encountered targets are subunits of the evolutionarily conserved multi-subunit general transcription factors called ‘the mediator’ and the P-TEFb elongation factor [91]. Like the TFIH helicase which is involved in RNAPII initiation, mediator and P-Tefb harbour specialised cyclin–cyclin dependent kinase protein pairs whose primary function is to phosphorylate the −50 C-terminal domain heptapeptide repeats of the large subunit of RNAPII [92–97]. A large body of literature indicates that transcription factors are able to recruit histone modifying and remodelling activities [98,99]. Understanding the exact mechanisms by which transcription factors bound to promoters and enhancers stimulate transcription remains an active field of investigation [100,101].

3.4. Transcription control elements on chromosomes as sites of molecular memory storage

To establish long-term memories in the brain, neurons need to remodel their connections with other neurons. It has been shown that this process requires transcription and relies on histone methylases that deposit H3K4me3 at gene promoters [102–104]. Hence, also in the brain, epigenetic enzyme activity is key in the storage of memories, presumably because nucleosome-based transcription control enables fixation of a transiently perceived molecular signal via altered gene expression in the following hours/days. More generally, it may be that the complex and highly coordinated set of protein–protein interactions that was first described to regulate gene expression potential in lower eukaryotes [105–108] can result in some level of self-sustaining topological
and chemical bookmarking of the human genome’s promoters and enhancers.

A straightforward class of models posits that in any cell type, a status quo is attained at all genes in the genome. This status quo can be biased towards transcription activation or towards repression, depending on the local concentration of opposing activities that control rate-limiting steps in the process of transcription. Hence, cell type-specific expression of histone ‘writer’ or ‘reader’ complex subunits may well be an important key to cellular fate fixation as this would modulate the activity of subsets of sequence specific DNA binding transcription factors that are already present in the cell but miss the presence of a particular chromatin remodelling complex (isoform) to influence gene expression [109–112].

Two memory-related questions can be asked at this point. How do cells manage to hold on to their identity and avoid becoming an unrelated cell type? A second question is how cells such as macrophages remember environmental signals they perceived in the past [113]?

An emerging first answer to both questions is that the ~1600 human TF genes (and the splice-isosforms they can give rise to) together with the chromatin remodelling factors that write, erase and read nucleosomal histone modifications are enough to explain the determination and differentiation of all the human cell types [114–116]. However, since a cell not only consists of chromatin and transcription factors, much more comes into play to maintain cell identity. Hence, RNA processing, transport and translation as well as non-transcription related interactions have the potential to form stable regulatory networks that can help perpetuate a given memory. Secondly, cells are embedded in a predictable environment consisting of other cells and tissue components, forming a complex environment that they can sense and to which they contribute. Thirdly, temporal sequences of cellular events are likely built in the epigenetic programs, so as to permit homeostasis in the long term by providing a predictable physiological outcome such as tissue turnover or repair. Hence, once a choice is made, a cells’ transcription factor network and its epigenetic program lead down a predetermined path [31]. Elucidation of the programs associated with the multitude of cell fates is a task for the present times.

4. Macrophage epigenomes

Macrophages (Mfs) and related phagocytes exist in both pro-terostome and deuterostome animals. In mammals, embryonically determined microglia in the brain, Kupfer cells in the liver and tissue-resident Mfs and dendritic cells, as well as osteoclast in the bones are examples of phagocytic cell types derived from the mesodermal lineage that performs haematopoiesis, the generation of blood cells [117,118]. Haematopoiesis branches out to form many cell types from one multipotent stem cell [119]. Our understanding of haematopoiesis, and in particular the non-lymphoid lineages, has progressed greatly thanks to recent single cell transcriptome analyses and mouse transplantation experiments, revealing a nested system of decision-making processes [20,120–123]. Furthermore, epigenic characterisation revealed great plasticity in the epigenetic programs of Mfs that consist of >10,000 enhancers whose activity changes when mouse Mfs are transplanted from one organ to another [124]. This is in keeping with the proposition that Mfs are very receptive to their tissue environment [16,125–129].

For more than a decade we have known the signature of histone-encoded signals at gene promoters and enhancers [130,131]. Currently, within the international human epigenome consortium (IHEC) and associated projects such as the European BLUEPRINT project, purified cell populations from healthy donors are being profiled for epigenetic marks along their chromosomes [132,133]. Addition of nuclease DNA accessibility data [134–136] alongside the nucleosome-borne information revealed that many acetylated and methylated nucleosomes at promoters and enhancers are adjacent to non-nucleosomal, highly accessible DNA stretches. Altogether, this provides accurate epigenome maps. Epigenome maps represent important information resources because they reveal the genetic pathways that are active as well as those that can be activated through further transcription factor binding [137].

Such anticipation of possible future events is the basis for a type of cellular memory that is thought to be important for innate immune memory and medicine because environmental signals, and the resulting epigenomic marks, are potentially amenable to pharmacological intervention when innate immune system cells are too active or not active enough [138].

Until recently, the powerful DNA rearrangement-mediated memory features of lymphoid cells overshadowed the potential for adaptive memory phenomena – that do not involve DNA rearrangements – in myeloid cells. During inflammation, monocytes differentiate into macrophages (Mfs) in the absence of cell division. Recent research underscores the capacity of adult monocytes not only to adapt their behaviour upon exposure to microorganism-associated molecular patterns (MAMPs) but also to consequently give rise to Mfs with different functionalities [139]. In fact, even invertebrate organism Mfs may undergo developmental training through engulfment of apoptotic embryonic cells, which appears to empower subsequent infectious bacteria phagocytosis [140].

Besides autonomic and tissue patterned cell fate determination mechanisms, innate immune cells such as Mfs must tailor their response to a potentially hostile environment that is a priori variable if not unpredictable due to the plethora of commensal and pathogenic microbes and viruses they may encounter. To address this issue, we integrated the transcriptomes and the epigenomes of three populations of Mfs that were derived in vitro from healthy adult human monocytes. The monocyte to Mf differentiation conditions consisted of 6 days of cultivation in 10% pooled human serum, and differed only in the fact that the monocytes were exposed to beta-glucan (BG), a fungal cell wall component, or to lipopolysaccharide (LPS), a bacterial cell wall component, during the first 24h of in vitro culture. The third populations served as unexposed control. At the sixth day of this protocol, Mfs derived from BG-treated monocytes are more pro-inflammatory and better phagocytes than the control cells [139]. Hence they are said to be trained, a concept with applications in vaccination [141]. On the other hand, Mfs derived from monocytes treated with LPS secrete lower levels of IL6 and TNF than the control Mfs. They are said to be tolerant, a concept that has been linked to sepsis–induced immune paralysis [138]. In both instances, the Mfs studied on day 6 appear to ‘remember’ their precursor cell’s encounters with microbes during the first day of the experiment.

In total, 3.8 10^8 bp (0.12% of the human genome) undergo nucleosome-borne epigenomic dynamics following exposure to MAMPs [61]. As output, 12% of the transcriptome was different (4-fold change, p < 0.05) between the three Mf subtypes. Therefore, a sizable proportion of the human chromatin and Mf gene expression program is affected upon detection of microorganisms. Indeed, trained, control and tolerant Mfs display transcriptomic differences at all investigated levels, going from their interactions potential with the extracellular matrix to signal receptors and metabolite transporters at their plasma membrane, through organelle and cytoplasm constituents and the signalling pathways they harbour all the way down to RNA processing and transcription factor constellations, altogether resulting in the establishment of reproducible Mf fate trajectories.

Training due to BG treatment induces H3K27 acetylation at more than 3000 enhancer loci, generating a large exclusive signature. The genes targeted by these enhancers are generally also active in the other Mfs, however, indicating a positive modulating role on
gene expression and rarely all or nothing gene activation events. Likewise, LPS-induced nucleosome acetylation increases at some 500 loci. For the endotoxin tolerant Mfs obtained by pre-treatment of the monocyte precursor cells with LPS, these changes often involve sustained expression of genes that were already active in monocytes, but now through newly established enhancers that are unique to the tolerant Mfs [61].

Besides H3K27 acetylation, which reports both enhancer and promoter activity, H3K4me3 marks >20,000 promoter loci, decorating the promoters of all the genes whose expression is detected at the RNA level in the monocyte-derived Mfs. Contrary to the sharp H3K4me3 peaks encompassing less than 3 nucleosomes (median 350 bp), H3K4me1, a precursor of the promoter H3K4me3 mark, is distributed over larger regions (median 3.5 kb). Interestingly, H3K4me1 marking, persists long after H3K27ac signals diminished [61]. As suggested for by Gioacchino Natoli [142], H3K4me1 embodies mid- to long-term epigenetic activity, perhaps by facilitating ulterior (re)activation of a transcription start site or an enhancer, a state referred to as latency. Nevertheless, this mark can be transient [15].

Are there key orchestrators of the Mf epigenome? Yes, in total more than 200 TFs, 100 kinases and 20 epigenetic enzymes were differentially expressed after monocyte to Mf differentiation [61]. For instance, the KDM6B/JMJD3 lysine de-methylase is rather uniquely expressed in monocytes and the E-box binding C2H2 zinc finger TFIIF SNAIL1 of monocytes is largely replaced by its SNAIL3 paralog in Mfs. The NKβ regulator IRAK3 is epigenetically and transcriptionally upregulated in endotoxin paralysed Mfs, and G-coupled receptors, such as the Adora receptors, and many cAMP signal transduction factors are expressed differentially by BG trained Mfs relative to control and paralyzed Mfs [61]. Furthermore, the trained cell’s metabolism is remodelled: including increased glycolytic capacity, cytoplasmic malate and isocitrate dehydrogenases and cholesterol metabolizing enzymes [61,143]. These and other metabolic changes are themselves likely to impact the dynamics of chromatin modifications through altered availability of co-factors for the writers and erasers of histone modifications [82,144–146].

5. Conclusions and perspective

We have known the ~20,000 human protein coding genes for more than 15 years, but are just now starting to discover all the regulatory DNA sequences that make up enhancers and promoters that control their expression. It is well established that enhancer and promoter usage is a defining property of cell lineages and sub-types [147]. Hence, comprehensive functional annotation of the epigenomically active loci will greatly impact diagnostic and prognostic tool development because novel fundamental knowledge about the inner workings of human cells will undoubtedly be revealed [148]. Within cellular chromatin, we can use generic technologies to systematically map accessible regions [136,149], nucleosome-borne epigenetic signals [130,150], post-replication 5-cytosine methylation [151,152] as well as loop interactions between enhancers and promoters [17]. Developing these new molecular biology resources through annotation in well-curated databases will extend our knowledge considerably. This can then be integrated with the human TF repertoire recognition motifs [153], and with knowledge on the chromatin remodelling enzymes that are recruited by TFs, to provide models of molecular signal transduction pathways where the DNA elements have importance on par with that of proteins. Very recently, a European COST action has been awarded to push these issues forward in a globally coordinated open context (www.cost.eu/COST_Actions/ca/CA15205).

Legend: Although the protein–DNA interactions that guide gene expression are extraordinarily intricate, when viewed from a DNA point of view, the number of possible interactions is small. The transactions involve distinct cis-acting DNA elements (promoters, enhancers, boundaries) that engage in higher order loops that are controlled by DNA-bound proteins. Operationally, a silencer may be a negatively acting enhancer, but it may also be a conditional boundary that can interrupt the looping interaction between a promoter and an enhancer. Similarly, decommissioning of a boundary may lead to gene activation, as this would permit novel promoter-enhancer loops to be formed [100].

Finally, an important corollary of epigenome profiling is that once DNA loci that function in regulation of specific cell behaviour have been catalogued (See vocabulary box), the genetic variation in these regions within the human population can be investigated too [149]. Since the majority of human variation resides outside the protein coding DNA we may then start to understand human biology much more deeply than we currently do.

Acknowledgements

We apologize to the many excellent researchers whose relevant work we did not cite due to space constraint and we thank J.H. Martens for useful comments on the manuscript. This work was funded by the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 282510–BLUEPRINT.

References


Transcription start site (TSS): Site where DNA transcription begins.

Transcription factor binding motif: DNA sequence motif recognized by proteins.

Histone modification: Nucleosome-bound molecular information consists of histone amino acid post translational modifications, of which more than 60 have been described to date.

DNA Accessibility: Obtained through observed accessibility to experimental probes. Is a region more likely to be bound by TFs than by nucleosomes. Nucleosome depleted regions, a.k.a. nucleosome free regions, occur at active TSSs and enhancers.

Promoter: DNA sequence for productive transcription initiation. The coordinates for a promoter must encompass at least one TSS, but extend to encompass adjoining DNA sequence and nucleosome-encoded molecular information.

Enhancer: Operational sequence that encompasses DNA- and nucleosome-borne information and which is capable of influencing juxtaposed promoter activity. Enhancers are devoid of functional TSS for mRNAs, although enhancers may be transcribed [154].

Differentially modified DNA: Defined by the extent of cytosine or other nucleobase post-replication covalent modification.

Loops to: topological notation for binary or higher order interaction between defined DNA elements.

Topologically associated domains (TADs): All the DNA located within a TAD is more likely to interact with other DNA sequences within the same TAD than with DNA located outside this TAD.

Boundary: DNA sequence that delimits one side of a TAD.


