Drug Discovery Today: Technologies
Vol. 7, No. 1 2010
Editors-in-Chief
Kelvin Lam – Harvard University, USA
Henk Timmerman – Vrije Universiteit, The Netherlands
Mechanistic pharmacology, new developments

Epigenetic control of the immune system: histone demethylation as a target for drug discovery

Jose M. Lora[1,3,*], David M. Wilson[1], Kevin Lee[1], Christopher G.C. Larminie[2]

[1]Immuno Epigenetics Discovery Performance Unit – EpiNova, GlaxoSmithKline, Stevenage, United Kingdom
[2]Computational Biology Department, GlaxoSmithKline, Stevenage, United Kingdom

In recent years significant progress has been made in our understanding of epigenetic control of a wide range of cellular processes. This has come about both through the concerted effort of the research community and through the development of technologies essential to the area. The importance of the epigenetic control of the immune system is becoming increasingly clear, and therefore epigenetics presents itself as an attractive, and potentially ground-breaking, entry point to tackle immune-mediated conditions. The advances in our understanding are in part due to the development of next generation sequencing technologies and chromatin immunoprecipitation. When combined, these approaches have allowed studies at the chromatin level to understand cellular responses to cell-external cues on a genome scale. This has contributed significantly to improved understanding of chromatin, its regulation through histone post-translational modification (PTM), and the enzymes and proteins involved in 'writing', 'reading' and 'erasing' these histone PTMs. In this review we focus specifically on the progress made in understanding the mechanisms involved in modulating histone methylation in the context of immunity and discuss the potential, and the challenges, presented by this exciting area for drug discovery.

Section editors:
Terence Kenakin – Glaxo Smith Kline, Systems Research 3-2104.2B, Research Triangle Park, NC 27709, USA.
Alan Wise – Edinburgh, UK

Introduction

Host defense in metazoans crucially depends on the differentiation of the cell lineages that constitute the immune system. These lineages can be broadly divided into lymphoid and myeloid subsets, which differentiate from hematopoietic progenitors in specialized tissues such as the bone marrow and the thymus. Cells from the lymphoid lineage give rise to T and B cells, and will take part in the so-called adaptive immune response, whereas myeloid cells (macrophages, dendritic cells and granulocytes) constitute the innate immune response [1,2]. During the past decade it has become well established that the transcriptional programs underlying these commitment and differentiation processes are subject to epigenetic control. Although it is becoming increasingly clear that the distinction between adaptive and innate immunities is an oversimplification, it has provided a useful paradigm to increase our understanding of immune function, and we will maintain it across this review for simplicity.

Upon encounter with a pathogen, animals mount an inflammatory response which will isolate, destroy and

*Corresponding author.: J.M. Lora (jose.lora@constellationpharma.com), (Christopher.G.Larminie@GSK.com)
[3] Current address: Constellation Pharmaceuticals, 215 First St., Cambridge, MA 02142, USA.

1740-6749/$ © 2010 Published by Elsevier Ltd. DOI: 10.1016/j.ddtec.2010.10.001

eliminate the invading agent. Furthermore, certain attacks will be 'remembered' by specialized memory cells that will quickly respond to subsequent encounters with the same pathogen. For this process to be effective, the immune system displays two fundamental properties: First, distinction between self and foreign and, second, resolution of the inflammatory process once the pathogen has been neutralized. However, environmental, genetic and epigenetic factors can result in the deregulation of this exquisitely controlled process and result in autoimmunity (failure to distinguish self from foreign) or chronic inflammation (failure to resolve an ongoing inflammatory response).

Eukaryotic DNA is densely packaged as chromatin, a protein/DNA complex. Histone proteins H1, H2A, H2B, H3 and H4 are the main protein constituents of chromatin. Nucleosomes are the fundamental units of chromatin and comprise two turns of the DNA around a histone hetero-octamer (2xH2A, 2xH2B, 2xH3 and 2xH4) core. Chromatin can be in two major conformational states, relaxed (euchromatin, transcriptionally active) and condensed (heterochromatin, transcriptionally inactive). The histone amino-terminal tails protrude from the nucleosome, and can be modified in multiple ways including; acetylation, methylation, phosphorylation, ubiquitination, sumoylation and citrullination. These modifications are fundamental for the control of the chromatin conformation. Of cardinal importance are acetylation (catalyzed by histone acetyl transferases, or HATs), which promotes the euchromatic state, and methylation (mediated by histone methyl transferases, or HMTs), which is associated with both conformations in a context-dependent manner. Thus, methylation on lysines 4 or 36 of histone 3 (designated as H3K4 and H3K36, respectively) is found in actively transcribed genes, whereas methylation on lysines 9 or 27 (H3K9 and H3K27, respectively) is associated with transcriptional repression. A further layer of complexity is defined by the fact that methylation can be found with different 'valencies', mono-, di- and tri-methylation (me1, me2 and me3, respectively). The epigenetic information contained in these post-translational modifications is transduced (or 'read') by proteins containing motifs capable of docking to these marks, such acetyl-lysine-binding bromodomains, and lysine-methyl-binding chromo-, tudor- and plant homeotic (PH)-domains. In addition to the 'writers' (transferases) and 'readers', the dynamic nature of the system is achieved by enzymes able to 'erase' acetyl marks (histone deacetylases, or HDACs) and methyl marks (histone demethylases, or HDMs). The result of the contributions of these enzymatic and effector activities is a highly dynamic system of paramount importance to control the response of eukaryotic cells to environmental cues, resulting in differentiation, activation and proliferation. For excellent reviews on chromatin biology and epigenetics, see Refs. [3,4].

The focus of the present work will be to review specific aspects of the epigenetic control of immunity. In particular, we will discuss how dynamic changes in the methylation state of histones can mediate certain immune responses, how these processes can be analyzed at a genomic scale, and how they can be controlled with small-molecule inhibitors for the treatment of immune-mediated diseases.

Control of innate immunity by histone methylation/demethylation

It was not until 2004 that the first histone demethylase was described by Yang Shi's laboratory [5]. In this ground-breaking report, it was demonstrated that the H3K4me1 and me2 could indeed be reversed by an amino oxidase enzymatic activity present in Lysine Specific Demethylase 1 (LSD-1). Importantly, because the oxidative reaction catalyzed by LSD-1 requires a protonated nitrogen in the substrate, LSD-1 can only demethylate H3K4me1 and H3K4me2, but not H3K4me3. Reversion of the trimethylated state requires a different enzymatic mechanism, originally described by Yi Zhang and co-workers [6]. Indeed, FBXL11 was shown to harbor a Jmj-C domain able to demethylate histone 3 at lysine 36 (H3K36) by a mechanism involving a hydroxylation with Fe(II) and α-ketoglutarate as cofactors. Subsequently, several Jmj-C containing demethylases have been described with different substrate specificities, including H3K4, H3K9, H3K27 and H3K36. The enzymatic reaction catalyzed by the Jmj-C motif allows demethylation of trimethylated histones, as exemplified by several members of the JmjD2 subfamily.

The first description of dynamic changes in the methylation state of promoters controlling gene expression in immune cells came from work in human monocyte-derived dendritic cells [7]. Indeed, in unstimulated cells certain inducible pro-inflammatory genes such as ELC, MDC and IL-12p40 were shown to bear H3K9 methyl marks in their promoter regions. Remarkably, upon TLR4 engagement by LPS stimulation, these methyl marks were erased with the same kinetics as RNA polymerase II (RNA pol II) recruitment and gene transcription. An attractive interpretation of this early work invoked the existence of a demethylase activity involved in dendritic cell activation. All H3K9 demethylases described so far belong to the Jmj-C-containing family, although the putative demethylase(s) directly involved in LPS-induced demethylation in human dendritic cells remains unknown. These early observations fuelled interest in the role of methylation in the control of transcription during innate immune activation. Another Jmj-C protein, JmjD3, was shown to be a H3K27 demethylase tightly regulated in LPS-stimulated macrophages and controlled by NF-κB [8]. This study and others [9,10] demonstrated that JmjD3 was a *bona fide* H3K27me3/me2/me1 demethylase in biochemical and cellular assays, and that was quickly upregulated in macrophages after LPS stimulation. This kinetics of activation was

unique for JmjD3, as all other demethylases tested failed to display the same pattern of expression. Work on human M2 macrophages has suggested that the maintenance of the differentiated phenotype is associated with H3K27me1/me2 demethylation, in a process that could be reproduced to some extent with JmjD3-specific siRNAs [11]. However, JmjD3-mediated regulation of gene transcription by H3K27me3 demethylation in macrophages remains controversial, and its role in macrophage activation remains to be clarified. A global analysis of binding of JmjD3 across the entire mouse genome showed no association between JmjD3 and H3K27 methyl marks, but in fact a remarkable overlap with transcription start sites of active genes characterized by H3K4me3 [10]. Moreover, macrophages isolated from JmjD3-deficient mice showed no major changes in the expression of JmjD3 target genes, but rather a subtle contribution to mRNA elongation [10].

Control of adaptive immunity by histone methylation/demethylation

Some of the earliest definitive indications of epigenetic control of immune responses came from work by Zhao and co-workers, demonstrating the fast recruitment of the SWI/SNF complex to chromatin after activation of mouse T cells [12]. A decade later, and thanks to the significant advances undergone by genomic technologies and chromatin immunoprecipitation (ChIP-sequencing, see below), it has become clear that antigen receptor ligation results in profound changes in the 'epigenetic landscape' of T cells [13,14]. Indeed, the activating marks H3K9/K14 diacetyl and H3K4me3 were consistently found in association with the promoter regions of active genes required for T cell differentiation and function, such as NFATC1, 2 and 3. Conversely, genes not expressed in the T cell lineage, such as NFATC4 and *NeuroD*, were marked by the repressive mark, H3K27me3 [13]. These observations have been extended to describe the epigenetic status of distinct T cell subsets. Effector T cells, such as Th1, Th2 and Th17 are characterized by H3K4 trimethylation in the promoter regions of their canonical cytokines IFN-γ, IL-4 and IL-17, respectively. Conversely, these promoters are marked by H3K27me3 in the subsets where these genes are not expressed [14]. The case for the master transcription factors that control specification of effector (Tbx21/Th1, Gata3/Th2, RORc/Th17) and regulatory T (Treg) cells (FoxP3) was shown to be more complex, and indeed one striking finding of this work was the fact that the pattern of histone modifications present on these 'master regulators' endow these subsets with a certain level of plasticity. Thus, Tbx21 was marked by H3K4me3 in Th1 cells, and by H3K4me3 and H3K27me3 in the other lineages. This co-existence of activating and repressive marks defines the so-called bivalent domains, which are found in the promoters of silent genes that are poised for expression. Experimental support for the concept of plasticity in T cells subsets predicted by bivalent domains in the Tbx21 locus in Tregs is shown by the detection of IFN-γ expression in purified Tregs that have been cultured under Th1-promoting conditions [14].

The dynamic control of T cell function and development by the status of histone methylation suggests a crucial role for histone methyl transferases and demethylases in the adaptive immune system. Genetic deletion in mice of transferases mediating H3K4 (MLL1), H3K9 (G9a) and H3K27 (Ezh2) results in defects in the development or function of the adaptive immune system [15–17]. However, in all cases analyzed immune function is not overly abrogated but rather, the phenotypes observed are relatively subtle or confined to specific compartments, suggesting significant compensatory mechanisms. Thus, $MLL1^{+/-}$ mice develop a normal immune system, but memory Th2 cells isolated from these animals show impaired Th2 cytokine release after re-stimulation, through a mechanism that involves binding of MLL1 to the Gata3 promoter region and presumably H3K4 methylation [15]. Interestingly, Th2 function appears to have another control point governed by a demethylase [18]. Indeed, Mina, a Jmj-C-containing protein, is recruited to the *Il4* promoter in an NFAT-dependent manner, and in this way acts as an *Il4* repressor. Moreover, enforced expression of Mina in $CD4^+$ T cells resulted in blockade of *Il4* expression, but not *Il2* or *Ifng*. Conversely, morpholino-based Mina knockdown in $CD4^+$ T cells increases the expression of *Il4*, but not *Ifng*. Although Mina is a functional demethylase (as communicated in Ref. [18]), the role of its catalytic activity in the control of Th2 function remains under investigation. Th1 differentiation has also been reported to be controlled in part by a H3K27 demethylase, JmjD3 [19,20]. In this case, JmjD3 was shown to bind to a specific region of the T-box domain of T-bet, and to be required for the optimal expression of several Th1-specific genes, such as *Ifng*. Strikingly, a distinct but partially overlapping region of the T-box domain binds an H3K4 methyltransferase activity [19]. These observations further suggest the crucial role that the dynamic control of the methylation state of chromatin plays in the regulation of cell type specific gene expression in the immune system.

G9a and EZH2, methyl transferases that catalyze the deposition of repressive marks, play distinct roles in B cell development and function. B-cell-specific deletion of G9a resulted in a significant loss of H3K9me2, as expected. However, this defect did not result in any major disruption of B cell development, but instead in a modest impairment in B cell proliferation and differentiation into plasma cells [16]. By contrast, Ezh2 has a crucial function in B cell maturation, as demonstrated in work by Su and co-workers. Ezh2 was found to be highly expressed in developing B cells, especially in the pro-B cell stage, and its deletion in the B cell compartment was characterized by a severe defect in immunoglobulin heavy chain rearrangement [17].

The importance of histone modifications in the regulation of the immune system is further emphasized by the presence of human mutations in essential epigenetic elements that lead to immune disease. Perhaps one of the most striking and well documented cases is the linkage between point mutations in the PHD finger of Rag2 and severe combined immune deficiency (SCID) in humans and mice [21]. Moreover, null mutations in the bromodomain-containing protein Sp110 has been shown to result in immunodeficiency [22].

Chromatin modifying enzymes as drug targets for autoimmunity and inflammation

Taken together, the crucial role of epigenetics in the control of the immune response and the existence of genetic links between chromatin modifying proteins and immune disease warrants the search for small-molecule inhibitors targeting key epigenetic proteins as immunomodulatory drugs. There are excellent reviews covering histone methyl transferase chemistry and drug discovery [23], and hence here we will focus on histone demethylases as drug targets in the next section of the review. The emergence of deep sequencing and chromatin immunoprecipitation (ChIP:Seq) is of cardinal importance for the understanding of epigenetic changes and for the characterization of epigenetic drugs, and a perspective of this technology will also be provided.

It has now been well established that epigenetic proteins, including HDMs exist as part of large, often dynamic multiprotein complexes [24]. This is an area of scientific investigation that remains relatively uncharted but in a seminal paper Shi and co-workers have explored the Hela cell LSD-1 complex, identifying at least 15 accessory proteins associated with this demethylase. Shi *et al.* also demonstrated that several of these accessory proteins were essential for the enzymatic activity of the LSD-1 [25]. This observation parallels those made in the HMT field where for example it is now well established that Ezh2 requires association with several accessory proteins for optimal activity [26]. These findings suggest that intra-complex protein–protein interactions are essential for correct folding and the formation of the catalytic site. Consequently a reliance on more traditional screening approaches reliant on recombinant proteins which do not necessarily represent the normal physiological state of the target, may lead to spurious findings. Indeed, recent studies comparing the pharmacology of recombinant HDAC proteins and native HDAC complexes have revealed differences in the potency of a range of HDAC inhibitors (G. Drewes, personal communication). These observations have a remarkable similarity with those that have been made in the kinase drug discovery arena and lead one to conclude that the use of recombinant systems to study these enzymes and to search for novel modulators represents a potential limitation to HDM drug discovery and efforts should be made to minimise one's reliance on such artificial systems.

Histone demethylase drug discovery

Drug discovery progress across the Jumonji family of histone demethylases

The Jumonji family of HDMs currently comprises over 30 distinct members and is the largest enzyme class that modifies chromatin through the removal of a stable methyl mark (Fig. 1) [27]. Since their discovery, the family has received increasing attention as targets for drug discovery. However, several significant challenges remain to be fully resolved before this target class can be considered mainstream for drug discovery. In particular, tractability issues around protein biochemistry, assay development and the emerging nature of the medicinal chemistry still need to be addressed. Of particular note are the challenges faced around assay development that relate to the poor catalytic turnover of these enzymes in recombinant systems (typical $K_{cat} \sim 0.1\ s^{-1}$) [28]. This trait may be physiologically important, however, it may in part be related to the challenges involved in the expression and utilisation of full-length protein, and in developing assay formats that effectively mimic the endogenous cellular environment. For example, the development of biochemical assays that utilise nucleosomes as substrates and/or present the Jumonji enzyme in the context of an endogenous and competent epigenetic complex, may ultimately prove more relevant assay systems. Further complexities may also exist due to the instability and sensitivity of these enzymes to oxidative conditions which may in turn impair their recombinant catalytic activity [28]. Nevertheless, and despite these challenges, significant progress has been made regarding assay development for these enzymes. In particular, mass spectrometry based assays have been demonstrated to provide a tractable means of directly measuring substrate turnover [29]. AlphaScreen has also recently been demonstrated to be a tractable format for this enzyme family and is readily amenable to high throughput screening approaches [30].

The medicinal chemistry knowledge around the Jumonji family of HDMs is at an emerging stage with only a limited number of inhibitors being reported to-date (Fig. 2). As a protein family, they belong to the 2-oxoglutarate (OG; Fig. 2 compound **1**) utilising superfamily of oxygenases [31,32] which also includes well studied enzymes such as the Hypoxia-Inducible Factor-1 (HIF) prolyl hydroxylases (PHD 1–3; EGLN-1–3) [33,34]. The Jumonji subfamily recognise methylated lysine (and arginine) motifs on histones and other proteins [35]. However, despite their common substrate and similar architecture, significant differences exist in the residues that comprise the catalytic domain across the family. This is apparent even when there is good sequence homology within a subfamily. For example, in the conserved Jmj-C domain, the sequence identity of JmjD2a compared with family 2 group members JmjD2c and Jmj2De is 80 and 64%, respectively. Moreover, recent crystallography data have indicated that

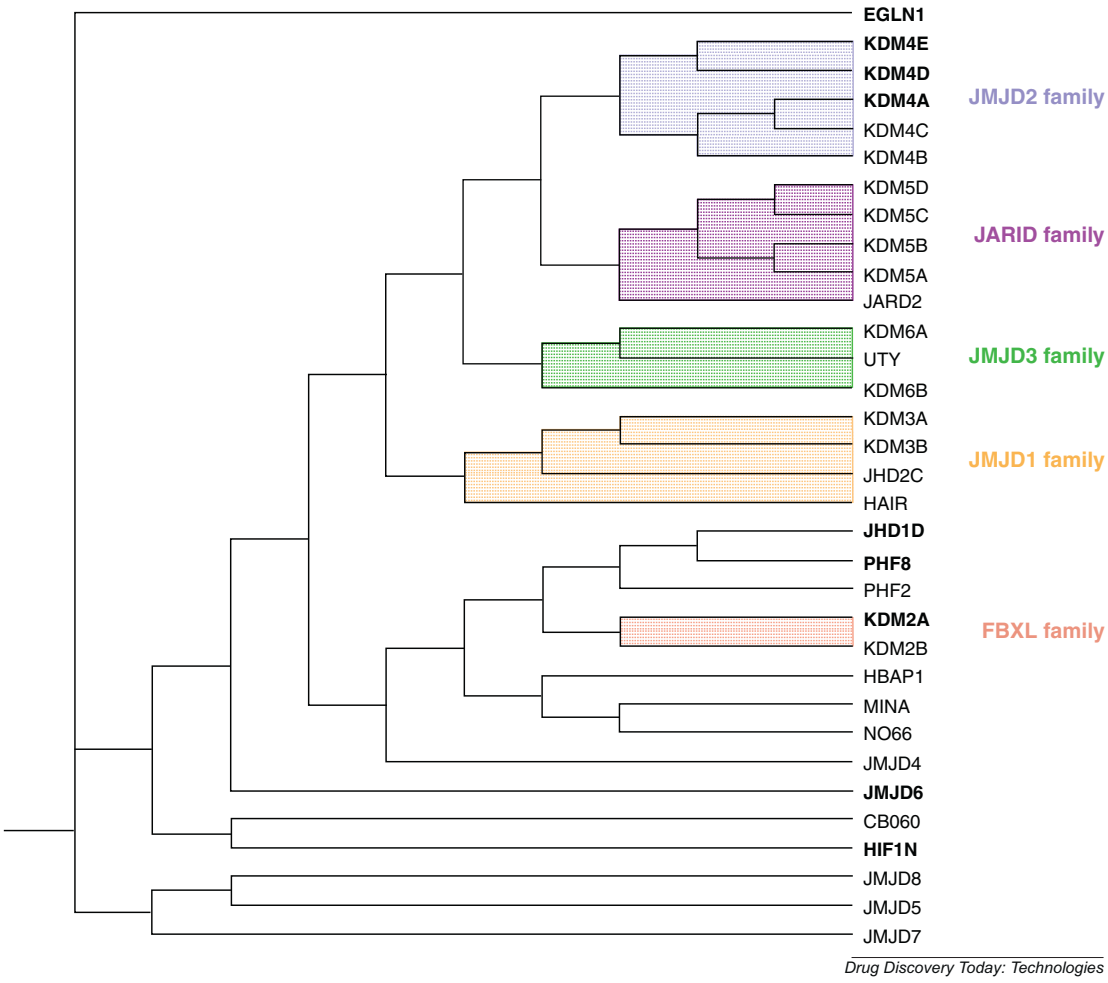

Figure 1. The human Jmj-C domain containing proteins; EGLN-1 (PHD2) from the wider OG utilising family shown for reference.

the binding mode of the co-factor within the conserved catalytic domain can vary significantly across different Jumonji family members [36]. Collectively this emerging body of structural information suggests that small-molecule drug design, targeting the Jmj-C domain, has promising potential for delivering selective inhibitors to probe the relevance of HDM activity on cellular phenotypes.

The co-factor mimetic, N-oxalylglycine (**2**; NOG) is catalytically incompetent and has been widely exploited in structural studies as a surrogate for OG. Using NOG as a start point, the Schofield group have reported on a series of analogues (Fig. 2; **4** and **5**) that have been demonstrated to inhibit JmjD2a and JmjD2e enzymes with low micromolar levels of activity whilst exhibiting good selectivity over PHD2, a more distant relative of the OG utilising family [37]. Encouragingly, this approach enables the concept of designing inhibitors based upon 'fragments' that compete for the co-factor binding site but which can be further elaborated to exploit the wider regions of the Jmj-C domain including the substrate binding pocket to deliver potency and selectivity. The potential of fragment-based approaches is further illustrated by the potent activity of pyridine 2,4-dicarboxylic acid (PDA; **3**) at JmjD2x enzymes [38]. The high ligand efficiency of such a co-factor mimetic could readily form the basis for inhibitor design approaches driven by structural insight. However, the clear challenge for such an approach will be to design small molecules that possess cell penetration given the apparent need for a carboxyl group in the co-factor binding pocket.

The recently reported public domain high throughput screens against JmjD2e (NCGC; Fig. 3) [39] and JmjD2c (Broad Institute) have in both cases identified potentially tractable small-molecule inhibitors to support lead optimisation efforts. Currently there are no reports on how these molecules have been further optimised against their respective targets and it is unclear how selective these hits are against other OG utilising enzymes. However, the evidence from these initial screens is encouraging and supports the significant potential of diversity approaches to identify novel start-points for small-molecule drug discovery targeting the Jumonji family.

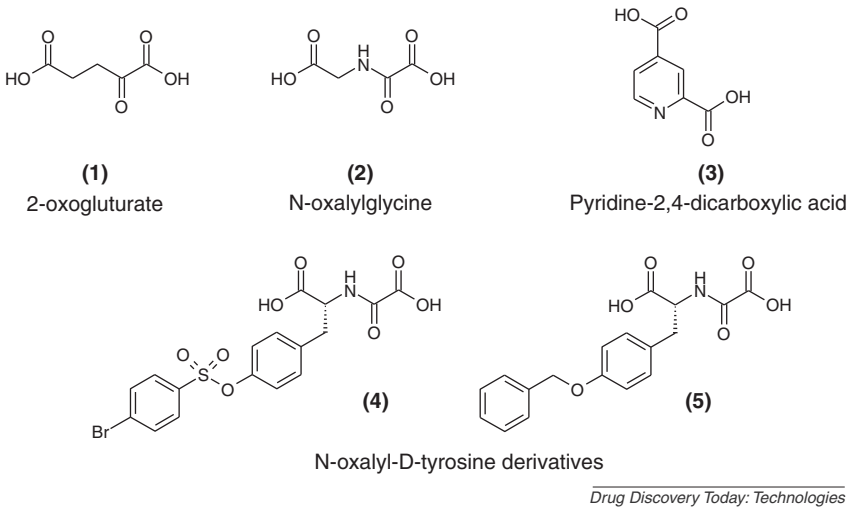

Figure 2. Known ligands or inhibitors of Jumonji family HDMs.

Beyond inhibitors that directly target the catalytic site, there is also the potential to inhibit lysine (or arginine) demethylation via disrupting the recognition of the histone peptide with the surface of the Jumonji protein. The interactions the histone peptide makes with the enzyme surface have been explored with JmjD2a which recognises both lysine 9 and 36 on histone H3 (H3K9/H3K36) [40–42]. In the case of H3K36, the histone peptide makes 11 polar interactions on the surface of JmjD2a which extend for 5 amino acids either side of lysine 36. However, given the lack of hydrophobic contacts and any obvious protein–protein binding 'hot-spots', the prospects for designing small-molecule inhibitors, that directly perturb the binding of the histone peptide with the surface of the Jumonji enzyme, are likely to be challenging given the current knowledge of these interactions.

Despite the current medicinal chemistry knowledge of the Jumonji family of HDMs being in its infancy, the long-term prospects for small-molecule drug discovery are encouraging. The relatively high density of crystallography data across the Jumonji family (structures solved for ~25% of the family) will facilitate progress in the area and enable selective inhibitors to be rapidly developed. In addition, the expanding target class knowledge around protein biochemistry and assay development is well positioned to provide the necessary foundations for effective drug discovery. Given the current interest in epigenetic mechanisms across academia and pharma, there is a clear need for medicinal chemistry to identify potent and selective probe inhibitors of the Jumonji HDMs. In this regard, initiatives from the Structural Genomics Consortium and Broad Institute, in partnership with pharma, have the potential to expedite knowledge in this emerging area and, crucially, identify small-molecule cell active probes that will enable the direct correlation of a demethylation event with a change in cellular phenotype. This current gap in the armoury of drug discovery scientists is likely to reduce significantly as medicinal chemistry responds to the challenges posed by the Jumonji family of HDMs.

Understanding chromatin biology at a genomic scale

ChIP:Seq: a brief overview of the methodology

Chromatin immunoprecipitation combined with next generation sequencing (ChIP:Seq) is a powerful technique to survey DNA-binding protein interactions with DNA on a genome scale. In essence the technique involves covalently linking DNA binding proteins to the chromatin they bind,

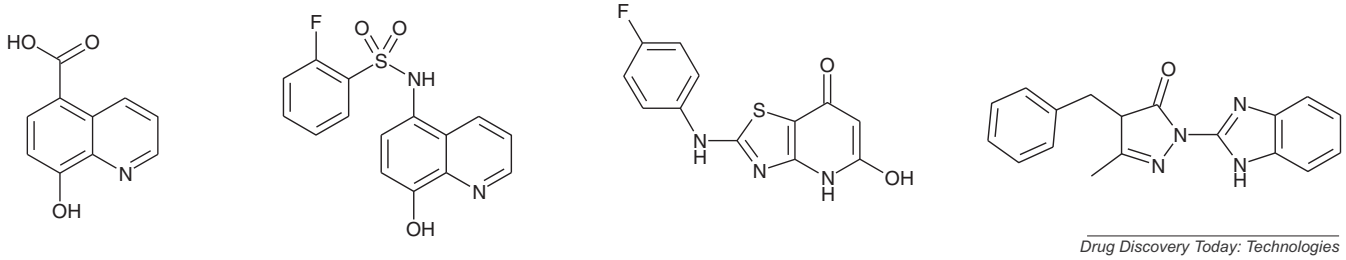

Figure 3. Hits identified by the NCGC screen of JmjD2e using a coupled assay format. Examples taken from confirmatory screen NCGC 2677.

sheering the DNA into small fragments, and immunoprecipitating the DNA:protein complexes using antibodies selective for the protein of interest. This results in the enrichment of DNA bound to the protein of interest. The co-immunoprecipitated DNA is sequenced using one of several massively parallel sequencing technologies and the resulting sequence reads are aligned to the reference genome giving rise to 'peaks' of enriched sequence reads corresponding to regions of the genome to which the protein of interest binds. The approach has massive potential to improve our understanding on a genome scale of the role of both transcription factors and chromatin-associated proteins in the regulation of gene expression. Because antibodies that selectively bind post-translationally modified histones are now available, it is possible to profile modified histones across the genome to understand how histone modifications correlate with the expression of the genes in which they occur [43]. This provides the potential to understand at the molecular level the global effect of modulating through small molecules the proteins that perform the histone 'reader', 'writer' and 'eraser' functions. Put simply, with ChIP:Seq one can assess the effect and therapeutic potential of targeting chromatin modifying enzymes such as the histone lysine demethylsases and histone deacetylases. In his review, Park [44] discusses both the advantages and issues associated with ChIP:Seq, whilst Roh *et al.* [13] and Wei *et al.* [14] report the results of genome scale analysis of histone modifications in T cells.

ChIP:Seq: a diamond in the rough for drug discovery

Although ChIP:Seq is improving our knowledge of gene regulation through chromatin modification on a genome scale, there are several challenges with the technology from a computational/data analysis perspective in the drug discovery context. The data analysis workflow has three stages. The first involves downloading of the data, processing to an appropriate format and alignment of the resulting sequence reads to a reference genome. The second phase requires normalisation and quality control assessment of the data. The third phase involves visualisation of the data, identifying the peaks of enriched sequence reads, assessment of the peaks, and peak annotation. There are several important detail considerations at each of these phases which are beyond the scope of this review suffice it to say that there are choices to be made on methodologies, algorithms, filters and parameters to use at each stage, dependent on experimental design and aim. These issues are covered in the thorough review by Pepke *et al.* [45]. Here, we will focus on the challenges with ChIP:Seq specifically as they relate to drug discovery.

ChIP:Seq: an emerging and rapidly evolving technology

The rapid pace of both next generation sequencing and its application to ChIP:Seq has several downstream effects. First, distinct next generation sequencing platforms have been developed by several companies, these platforms are all constantly evolving to improve both their sequence read lengths, throughput and accuracy which ultimately requires continuous adaptation of the methods used to analyze the data. Second ChIP:Seq is only one of several uses of next generation sequencing platforms and has unique requirements for analysis tools. Currently this, combined with the costs of the experiments, results in a relatively limited market for any commercial software provider. These two factors have contributed to the public domain emerging as the most valuable source of cutting edge tools and algorithms to interpret ChIP:Seq data. The rapid pace of development does come at a cost and so the algorithms, scripts and software are often platform sensitive, and not always as robust or as easy to install as commercial software, nor is the associated documentation or support as comprehensive. This is not a criticism of the public domain as in fact there have been remarkable efforts to incorporate analysis workflows into packages such as BioConductor open source software for bioinformatics [46,47]. Furthermore, it is understandable because most funding is directed towards driving innovation (i.e. methods development) and not to establish robust and reliable data analysis services. Within Pharma, then, many groups will probably choose to incorporate the methods generated in the public domain into more reliable and robust workflows to fit their computing platforms.

ChIP:Seq analysis methods have been developed in response to the experiments against which they were targeted. Earlier ChIP:Seq experiments focused on understanding transcription factor binding which generally occurs across a small percentage of the genome at short, discrete regions. This gives rise to sharp, narrow and well defined peaks of sequence enrichment at these sites against the background. For peak finding algorithms, this has had an obvious impact and many peak identification algorithms work optimally with transcription factor binding data [45]. More recently, ChIP:Seq has been used to understand chromatin modification, utilising antibodies that specifically recognise methylated and acetylated forms of histone H3. Methylated and acetylated histones occur across a far greater proportion of the genome than transcription factors and some, such as H3K27 occur along larger stretches of chromatin giving rise to broad sequence enrichment peaks that are less well defined. Amongst the dozen or so peak identification algorithms available, only SICER has been developed to cope with such long range peaks [48], hampering analysis of some histone modification data.

Given the central role histone modification has on regulating downstream gene expression, even compounds selective against particular 'readers', 'writers' or 'erasers' are likely to have far reaching effects across the genome.

ChIP:Seq represents a potential technology to help unravel the molecular mechanism of action on a genome scale of targeting these proteins. We predict altered concentration of any selective inhibitors raised against these proteins will have quantifiable effects at individual gene loci as well as the range of genes they influence. This may prove pivotal towards both identifying a therapeutic window and identifying markers of both efficacy and safety to take forward. There are several technical challenges to overcome for this to be widely achievable. ChIP:Seq has been used largely to characterise protein:DNA interactions on a genome scale. To our knowledge, all experiments conducted to date have focused on identifying peaks. Commonly this is achieved by conducting experiments with two or more antibodies and then comparing the peaks generated on the reference genome. Thus peaks are described as being present or absent at a given position of the genome in relation to a control or the results from another antibody. In the drug discovery arena, we also want to quantify changes in peak size in response to drug and relate these to expression. This may require significantly deeper sequencing to obtain sufficient sensitivity and coverage at peaks of interest, will require analysis methods to quantify peak changes, potentially both in size and shape, and methods that can cope with biological replicates, all within a reasonable cost. Thus although the technology and the field have made tremendous progress, there are still challenges ahead to maximise ChIP:Seq's value as a tool for drug discovery.

References

1 Delves, P.J. and Roitt, I.M. (2000) The immune system, part I. *New Eng. J. Med.* 343, 37–49
2 Delves, P.J. and Roitt, I.M. (2000) The immune system, part II. *New Eng. J. Med.* 343, 108–117
3 Allis, C.D. *et al.* (2006) Overview and Concepts. In *Epigenetics* (Allis, C.D. *et al.* eds), pp. 23–61, CSHL Press
4 Kouzarides, T. (2007) Histone modifications and their function. *Cell* 128, 693–705
5 Shi, Y. *et al.* (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941–953
6 Tsukada, Y. *et al.* (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439, 811–816
7 Saccani, S. and Natoli, G. (2002) Dynamic changes in histone H3 Lys 9 methylation occurring at tightly regulated inducible inflammatory genes. *Genes Dev.* 16, 2219–2224
8 De Santa, F. *et al.* (2007) The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* 130, 1083–1094
9 Lan, F. *et al.* (2007) A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* 449, 689–695
10 De Santa, *et al.* (2009) Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. *EMBO J.* 28, 3341–3352
11 Ishii, M. *et al.* (2009) Epigenetic regulation of the alternatively activated macrophage phenotype. *Blood* 114, 3244–3255
12 Zhao, K. *et al.* (1998) Rapid and phosphoinositol-dependent binding of the SWI/SNF-like BAF complex to chromatin after T lymphocyte receptor signaling. *Cell* 95, 625–636
13 Roh, T-Y. *et al.* (2006) The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15782–15787
14 Wei, G. *et al.* (2009) Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating $CD4^+$ T cells. *Immunity* 30, 155–167
15 Yamashita, M. *et al.* (2006) Crucial role of MLL for the maintenance of memory T helper type 2 cell responses. *Immunity* 24, 611–622
16 Thomas, L.R. *et al.* (2008) Functional analysis of histone methyltransferase G9a in B and T lymphocytes. *J. Immunol.* 181, 485–493
17 Su, I. *et al.* (2003) Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat. Immunol.* 4, 124–131
18 Okamoto, *et al.* (2009) Mina, an *Il4* repressor, control T helper type 2 bias. *Nat. Immunol.* 10, 872–879
19 Miller, S.A. *et al.* (2008) Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. *Genes Dev.* 22, 2980–2993
20 Miller, S.A. and Weinmann, A.S. (2009) An essential interaction between T-box proteins and histone-modifying enzymes. *Epigenetics* 4, 85–88
21 Matthews, A.G. *et al.* (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation V(D)J recombination. *Nature* 450, 1106–1110
22 Roscioli, T. *et al.* (2006) Mutations in the gene encoding the PML nuclear body protein Sp110 are associated with immunodeficiency and hepatic veno-oclusive disease. *Nat. Genet.* 38, 620–622
23 Copeland, R.A. *et al.* (2009) Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug Discov.* 8, 724–732
24 McDonel, P. *et al.* (2009) Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int. J. Biochem. Cell Biol.* 41, 108–116
25 Shi, Y.J. *et al.* (2005) Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* 19, 857–864
26 Cao, R. and Zhang, Y. (2004) The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr. Opin. Genet. Dev.* 14, 155–164
27 Klose, R.J. *et al.* (2006) JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* 7, 715–727
28 Sakurai, S. *et al.* (2010) A miniaturisation screen for inhibitors of Jumonji histone demethylases. *Mol. BioSyst.* 6, 357–364
29 Tsukada, Y. *et al.* (2006) Histone demethylation by a family of JmjC-domain containing proteins. *Nature* 439, 811–816
30 Kawakura, A. *et al.* (2010) Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal. Biochem.* 404, 86–93
31 Purpero, V. *et al.* (2007) The diverse and pervasive chemistries of the α-keto acid dependent enzymes. *J. Biol. Inorg. Chem.* 12, 587–601
32 Loenarz, C. and Schofield, C.J. (2008) Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat. Chem. Biol.* 4, 152–156
33 Jelkmann, W. *et al.* (2007) Hydroxylation of hypoxia-inducible transcription factors and chemical compounds targeting the HIF-a hydroxylases. *Curr. Med. Chem.* 14, 1853–1862
34 Schofield, C.J. *et al.* (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. Inorg. Biochem.* 100, 644–669
35 Mukherji, M. *et al.* (2009) Identification of non-histone substrates for JmjD2A-C histone demethylases. *Biochem. Biophysical Res. Commun.* 390, 280–284
36 See PDB reference pdb2yul for the crystal structure of FBXL11
37 Schofield, C.J. *et al.* (2010) Selective inhibitors of the JmjD2 histone demethylases: combining nondenaturing mass spectrometric screening and crystallographic approaches. *J. Med. Chem.* 53, 1810–1818
38 Schofield, C.J. *et al.* (2008) Inhibitor scaffolds for 2-oxoglutarate-dependent histone lysine demethylases. *J. Med. Chem.* 51, 7053–7056
39 See PubChem(JmjD2e: AID 2147; AID 2677; AID 2187; AID 2188 and JmjD2c: AID 2099)
40 Zhang, G. *et al.* (2006) Structural insights into histone demethylation by JMJD2 family members. *Cell* 125, 691–702
41 Zhang, G. *et al.* (2007) Structural basis of the recognition of a methylated histone tail by JMJD2A. *Proc. Natl. Acad. Sci.* 104, 10818–10823

42 Schofield, C.J. *et al.* (2007) *Crystal Structures of Histone Demethylase JMJD2A Reveal Basis for Substrate Specificity*.
43 Barski, A. *et al.* (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837
44 Park, P. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10, 669–68038
45 Pepke, S. *et al.* (2009) Computation for ChIP-seq and RNA-seq studies. *Nat. Methods* 6 (Suppl), S22–32
46 Zhu, L.J. *et al.* (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* 11, 237
47 Gentleman, R.C. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80
48 Zang, C. *et al.* (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25, 1952–1958