https://doi.org/10.1038/s41576-023-00648-z

Pioneer factors — key regulators of chromatin and gene expression

Martha L. Bulyk, Jacques Drouin, Melissa M. Harrison, Jussi Taipale & Kenneth S. Zaret

Pioneer factors are a group of transcription factors with important roles in gene regulation during development. Their unique ability to bind to compacted chromatin, promoting its remodelling for gene expression, sets them apart from other regulatory proteins and makes them essential players in cellular differentiation, fate determination and reprogramming. Here, we have tasked five experts with discussing our current understanding of pioneer factors and their functions to showcase why exploring these proteins offers unique insights into gene regulation and cellular identity.

How would you define a pioneer factor?

Melissa M. Harrison. Development is regulated by the initiation of novel gene expression programmes, which are driven by transcription factors (TFs) binding to distinct *cis*-regulatory modules. Nucleosomes are refractory to binding of many TFs, and binding is further limited in compacted chromatin. By contrast, active *cis*-regulatory modules are largely devoid of nucleosomes and, therefore, accessible to TFs. It was unclear how developmentally regulated *cis*-regulatory modules initially gained accessibility. Pioneer factors provided a powerful framework for resolving this conflict with the initial demonstration that FOXA (also known as HNF3) could bind closed, nucleosome-occupied chromatin, promote accessibility and facilitate the later binding of additional, tissue-specific TFs[1]. Since the initial studies, it has become clear that multiple disparate factors share these defining characteristics of a pioneer factor[2]. Although the distinct biochemical mechanisms may differ, pioneer factors share the capacity to define *cis*-regulatory modules to drive gene expression programmes.

Martha L. Bulyk. I consider a pioneer factor to be a sequence-specific TF that can bind DNA recognition sites embedded in nucleosomes, which leads to the opening of chromatin by various molecular mechanisms that depend on the particular pioneer factor. In my group, we refer to nucleosomal binding as 'pioneer binding' and the subsequent opening of chromatin as 'pioneer activity'.

Ken S. Zaret. Most broadly, I would define a pioneer factor as a TF that can target a gene or enhancer in closed chromatin and initiate chromatin opening, as seen during cell fate changes[1]. A more refined definition is that a pioneer factor can target a DNA site on a nucleosome and elicit changes to the underlying nucleosome or neighbouring nucleosomes that allow other factors to enter the chromatin and impart a new functional identity to the local domain. Protein cross-linking studies and cryo-electron microscopy structures are revealing that pioneer factors have essential interactions with core histone proteins in the nucleosome, firmly establishing nucleosome interactions as being central to function. Pioneer factors can displace linker histone or distort DNA, leading to exposure of an underlying nucleosome in chromatin. The initial nucleosome perturbations by pioneer factors are transient but can be locked in and expanded by enabling the action of nucleosome remodellers[3], as initially seen for GAGA factor and the NURF complex[4].

Jacques Drouin. Pioneer factors have the unique ability to access their target DNA sequences within so-called closed chromatin and to initiate the process of chromatin opening[5]. Beyond this description of their unique properties, what are their unique features that provide them with this capacity? Closed chromatin is condensed and, hence, contains mostly nucleosomal DNA. Unsurprisingly, most pioneer factors have the ability to interact with nucleosomal DNA; however, it remains unclear to what extent this ability is necessary and sufficient for pioneer action[6], or whether it is permissive in conjunction with other features, such as the ability to interact with heterochromatin proteins. Although interaction with heterochromatin was observed for several pioneer factors, its importance for pioneer action remains to be defined.

Jussi Taipale. A pioneer factor is a TF that can, when present in a sufficiently high concentration, access and bind to nucleosomal DNA in cells without needing accessory factors or other TFs. As the name says, pioneer factors bind first, paving the way for other TFs to bind to nearby sites. So my suggested definition is operational, not simply qualitative and biochemical. By this definition, factors that are never expressed at a high enough level in vivo to access their nucleosome-obstructed binding motifs would not qualify as pioneer factors. However, in pathological conditions such as cancer, a non-pioneer factor could gain pioneer activity due to mutation or overexpression. By contrast, many pioneer factors commonly act in a non-pioneer capacity when their expression level is insufficient to effectively access their binding motifs in nucleosomal DNA, but sufficient to bind to already open chromatin (see, for example, ref. 7).

Note that my definition differs from the definition based on whether a TF binds to all its motifs in cells; the latter definition would include many TFs that contain arrays of C2H2 zinc fingers such as KRAB repressors, CTCF and REST. Because consecutive zinc fingers block access to continuous stretches of the major groove, steric hindrance prevents most C2H2 zinc finger TFs from accessing nucleosomal DNA. Hence, although many multi-zinc finger TFs bind to most of their high-affinity motifs, they most likely do so by binding to DNA when a nucleosome is not present, rather than by specifically recognizing their motifs in nucleosomal DNA. In addition, the tendency of these factors to bind to the genome alone makes them more similar to 'lone rangers' than pioneers.

What are the most important molecular and cellular functions of pioneer factors?

M.M.H. With a focus on how developmental gene expression programmes are regulated, the most important aspect of pioneer factors is not a single molecular mechanism but the overarching conceptual framework that they

bind and open closed chromatin. The ability of pioneer factors to define novel *cis*-regulatory modules allows pioneer factors to act at the top of gene regulatory networks. It is this capacity that enables pioneer factors to drive reprogramming, transdifferentiation and, when dysregulated, diseases such as cancer. This is achieved by binding to closed chromatin and promoting accessibility for additional TFs that regulate gene expression. Nonetheless, the molecular mechanisms may differ between factors. Proteins that function as pioneer factors have a diversity of different DNA binding domains, and biochemical and structural studies have demonstrated that they engage the nucleosome through distinct modalities[2,8]. Some pioneer factors may primarily promote accessibility by directly functioning to evict histones, whereas others function through the recruitment of ATP-dependent chromatin remodelling complexes.

The contributors

Martha L. Bulyk is a Professor of Medicine and Pathology at Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. Her research laboratory develops high-throughput experimental and computational methods for investigations of transcription factors and *cis*-regulatory elements, with a focus on mechanisms of their interactions and the impact of genetic variants on gene regulation.

Jacques Drouin is Director of the Laboratory of Molecular Genetics at the Institut de recherches cliniques de Montréal, Canada with academic positions at Université de Montréal and McGill University. His research focuses on transcriptional and epigenetic regulation of pituitary cell differentiation and function. He is a Fellow of the Royal Society of Canada and of the Canadian Academies of Sciences and of Health Sciences.

Melissa M. Harrison is a Professor of Biomolecular Chemistry at the University of Wisconsin School of Medicine and Public Health, USA. Her research is focused on transcriptional control of conserved developmental transitions. One major area of investigation is understanding the rapid and efficient genomic reprogramming that occurs during early development to establish a totipotent embryo. She was a Vallee Scholar and is a Romnes Faculty Fellow and Vilas Faculty Mid-Career Investigator.

Jussi Taipale is a professor at Cambridge University, UK, at Karolinska Institutet, Sweden and at University of Helsinki, Finland. His laboratory focuses on high-throughput biology — combining experimental and computational approaches to understand two systems-level questions that are presently poorly understood: the mechanisms that control growth of tissues and organisms, and the rules that specify how DNA sequence determines when and where genes are expressed.

Kenneth S. Zaret is Professor of Cell and Developmental Biology at the Perelman School of Medicine, University of Pennsylvania, USA. His research focuses on how to overcome chromatin barriers during cell fate changes in embryogenesis, regeneration and disease. Ken is the Director of the Institute for Regenerative Medicine at University of Pennsylvania and originated the concept of pioneer factors.

J.T. The main molecular function of pioneer factors is the ability to recognize their motifs in the context of chromatin and the nucleosome. Biologically, they are central in both opening new chromatin and maintaining existing open chromatin — functions that are central to gene expression and cell fate determination.

M.L.B. A critical function of a pioneer factor is that its nucleosomal binding enables subsequent binding by co-regulatory factors, leading to chromatin opening and allowing additional regulatory factors to (co-)bind and activate gene regulatory cascades[9]. In this way, pioneer factors serve as key initiators of regulatory cascades in development, differentiation and cellular reprogramming.

J.D. The most important function of pioneer factors is their action in the specification of new cell fates. This role is crucial as it circumvents mechanisms that are active in all cells to maintain the status quo of chromatin organization, whether active or inactive. By targeting regulatory sequences within the inactive part of the genome and deploying new enhancer repertoires, pioneer factors are unique in initiating the process of programming new cell identity. They thus act in development at crucial junctions to allow establishment of new cell lineages.

It is also noteworthy that beyond opening the chromatin structure at enhancers for activation of new gene expression programmes, some pioneer factors also implement epigenetic memory through DNA demethylation; following enhancer activation by the pioneer and its removal, enhancers thus remain accessible in the primed state[10].

K.S.Z. Single-molecule tracking studies have revealed that the non-specific nucleosome scanning feature of pioneer factors enables them to scan compacted forms of chromatin, whereas non-pioneer factors that do not bind nucleosomes scan more open domains of chromatin[11]. I view the compact chromatin scanning feature of pioneer factors as being central to their roles in enabling cell fate changes, where silent gene regulatory sequences must be accessed and activated for a new cell fate.

Numerous genetic studies have revealed that cell fate changes involving the closed chromatin targeting function of pioneer factors include zygotic genome activation, embryonic development and cellular reprogramming[2]. An unanticipated feature was a role for pioneer factors in cancer; pioneer factors enable oestrogen and androgen hormone receptors to bind chromatin in breast cancer and prostate cancer, respectively, and parts of pioneer factors can become fused to different activation domains, generating oncogenic fusion proteins[12]. These discoveries extend pioneer function into acute gene regulatory scenarios that are distinct from cell fate control and have led to attempts to block pioneering function to suppress cancer cell growth.

What are the most informative methods for studying pioneer factors?

M.L.B. A combination of in vitro and in vivo systems is most powerful for providing insights into mechanisms of pioneer binding and activity. In vivo analysis can provide valuable information on the pathways that are regulated, the effects on chromatin and occupancy by additional factors. However, specific features regulating a pioneer factor's binding to nucleosomes can be obscured by the context of the nuclear milieu. In vitro methods enable a reductionist approach to reveal a putative pioneer's intrinsic capability for binding nucleosomes, and to assess where and to what sequences a pioneer binds and how that binding may be modulated by cofactors. Such methods should assay full-length proteins where possible, as regions outside the DNA binding domain may influence pioneer binding. For example, one study found that a short helical region outside the DNA binding domain of the pioneer factor FOXA1 promotes interaction of FOXA1 with core histones[13]. Highly parallel, sequencing-based assays of large libraries of DNA sequences on nucleosomes, depending on how they are designed, can provide insights into the sequence features of pioneer binding, such as the positional preferences of where, within nucleosomes, pioneer factors bind preferentially[14–16]. Structural studies of TF-bound nucleosomes, although typically limited to a single nucleosomal DNA sequence, serve as a lens for understanding molecular mechanisms by which pioneer factors interact with their binding sites on nucleosomes, such as which of potentially multiple DNA binding subdomains they utilize in binding nucleosomal sites, and which parts of other proteins in the complex may be involved in mediating the binding interaction[14,17].

K.S.Z. The definitive test of whether a TF is a pioneer is to test whether it can bind to its target DNA motif on nucleosomes. Although this is most readily assessed by in vitro studies with recombinant molecules, an in vivo surrogate is to assess whether a newly induced factor targets pre-existing nucleosome sequences, as determined by mapping DNA fragments that are MNase-resistant or enriched for core histones by chromatin immunoprecipitation (ChIP) or CUT&RUN[2]. More frequently used tests assess whether a newly induced factor targets genomic sites that are resistant in ATAC-seq or DNase-seq assays. Non-pioneer factors that cannot saturate a nucleosome binding site in vitro can target closed chromatin sites when vastly over-expressed in vivo, as defined by expression far beyond what the factor would exhibit in a natural cell fate-changing context. More physiologically expressed, nucleosome-binding pioneer factors will target MNase-resistant, ATAC-resistant and DNase-resistant sequences far more frequently than non-pioneer factors. Yet, in all cases, pioneer factors also target open chromatin sites[18]. Notably, more than half of pioneer factors' targeted sites are in closed chromatin[19,20], which reflects nucleosome affinity that may be imparted by core histone interactions compensating for partial DNA motif recognition.

J.T. The field is truly interdisciplinary, and no single method can be used to define pioneer activity. As the precise molecular mechanisms of pioneer activity differ between different factors, the most informative methods in vitro have been nucleosomal binding assays that can identify binding specificities of TFs on the nucleosome, and the factors' positional and orientational preference[14,15]. In vivo, key experiments include induced expression of TFs that are not normally expressed in a cell type, and then studying how motifs on nucleosomes are recognized using genome-scale methods such as MNase-ChIP[21]. Recently, methods such as BANC-seq[22] have also started to bridge the gap between in vitro and in vivo studies.

M.M.H. To generate a comprehensive understanding of how pioneer factors drive developmental transitions, it is essential to study them in both a biochemical and a physiological context. Elegant biochemical and structural assays have elucidated how pioneer factors engage DNA within nucleosomes, how pioneer factors interact with the histones, how modifications to the histones can affect this interaction and how cofactors stabilize interactions[8,17,23]. Although these studies provide essential insights into how pioneer factors engage chromatin, similar to any experiment there are limitations to these assays as they are often performed on mononucleosomes generated with DNA sequences that promote stable nucleosome positioning. By contrast, nucleosomes are often poorly positioned at enhancers in vivo. Furthermore, the strengths of the simplified biochemical systems necessitate that they lack the complex set of possible cofactors and chromatin modifications that exist in a cell. To understand the biological significance, biochemical studies must be complemented with investigations of protein function in vivo at endogenous levels[24]. Diverse biological contexts may have different requirements. For example, in all organisms studied to date, pioneer factors are required for the initial expression of genes from the zygotic genome[25,26]. At this early stage of development, the chromatin is relatively naïve and largely devoid of modifications to the histones. Later in development, as the chromatin matures and diverse sets of cofactors are expressed, these same pioneer factors can occupy distinct locations and have different functionalities. Additionally, the levels of expression of the pioneer influence binding and activity[27]. Complementing rigorous biochemical studies with functional studies within an organism is essential to understanding the biological role of pioneer factors in defining the *cis*-regulatory modules required for driving developmental transitions.

Are TFs either pioneering versus non-pioneering, or is it more informative to consider TFs on a continuum of pioneering activities?

M.L.B. TFs are very likely on a continuum of pioneering activities, with some exhibiting no pioneering ability at physiological levels. Some TF DNA binding domains may have a more favourable structure for binding DNA on nucleosomes, such as a short recognition α-helix[8]. Differences among members of a TF class may lead to some being strong versus weak nucleosome binders; for example, the basic helix–loop–helix (bHLH) DNA binding domain of c-Myc seems to be too rigid to bind nucleosomes on its own. By contrast, the basic helix of the bHLH pioneer factor Ascl1 is shorter, allowing for binding to nucleosomes. Overall, different pioneer factors use different DNA binding modes to bind nucleosomes and recognize either the same or partial (or degenerate) versions of the motifs that they recognize on naked DNA[28,29]. The different properties of TFs that confer the ability to bind nucleosomes remain unclear. One challenge in delineating such features is that disordered regions of pioneer factors, which are far less understood than their DNA binding domains and whose positions within nucleosomes are not well-characterized, can contribute to pioneer interactions with nucleosomes[17].

J.T. The pioneering activity itself exists as a continuum, whereby the concentration of the factor has a key impact. Most TFs bind to their 10–15-bp motifs with higher binding energy than a nucleosome binds to the same length of DNA. As the nucleosome binds to a longer sequence (147 bp), a high concentration of TFs and/or multiple binding sites are still needed to displace a nucleosome[30]. There are both quantitative and qualitative differences in the way individual TFs access nucleosomes. Pioneer factors can bind to nucleosomes using multiple different binding modes, including classical end–edge binding, periodic binding, dyad binding, oriented binding and binding to both DNA gyres. An individual factor can access nucleosomes using several of these modes. However, because of steric hindrance, some modes are unavailable to members of particular structural TF classes[8]. In those cases, the binding mode utilization can be seen as a qualitative rather than a quantitative feature.

K.S.Z. Many studies have shown that there is a range of affinity of TFs for nucleosomes in vitro and a range of targeting nucleosomes or closed chromatin in vivo. Furthermore, point mutations that alter pioneer factor interactions with core histones affect their ability to bind or open compacted chromatin[13,31], which clearly demonstrates that 'pioneering' is defined by nucleosome binding. On the basis of a comparative analysis of the ability of many TFs to bind nucleosomes and on structure studies, it seems that there are various ways in which pioneer factors bind to, interact with and modify the nucleosome. In the future, such assessments will be more useful to define and classify pioneering.

J.D. Some TFs are clearly incapable of pioneer action; thus, it seems justifiable to think of TFs in the binary context of pioneer versus non-pioneer factors. However, for TFs with pioneer action, both genomic recruitment strength, as assessed by ChIP followed by sequencing (ChIP–seq), and the action of pioneer factors appear as a continuum. Genomic sites, such as enhancers, that are

fully activated following pioneer action, tend to be sites with strong genomic recruitment. However, there are discontinuities in the spectrum of outcomes following pioneer recruitment, with some strong (relatively speaking) sites not being fully activated or, conversely, relatively weak recruitment sites being fully activated. These differences in response to pioneer-triggered recruitment likely reflect the contribution of the local environment (meaning enhancer structure and its chromatin organization, as well as cooperating TFs) to the process of chromatin opening. To start, not a single pioneer is known to be able to, on its own, conduct the process of enhancer activation from initiation to full activation. On the contrary, there are numerous examples where pioneer factors require the cooperation of non-pioneer TFs for full enhancer activation and cell fate determination[5]. Although this type of dependency might affect the ability of pioneer factors to fully open enhancer chromatin and lead to transcriptional activity, it remains possible that pioneer factors may suffice to initiate the process, leading to the primed enhancer status. The multiple steps involved in pioneer action may represent many interactions with other TFs and/or remodelling activities that could explain discontinuities in outcomes relative to pioneer recruitment strength. This is exemplified in structure–function studies of a few pioneer factors, in whom mutations or differential splicing do not affect genomic recruitment but, nonetheless, hamper cell specification activity and enhancer opening. For example, we recently showed loss of melanotrope transcriptome activation either by a natural isoform of the pioneer PAX7 or by carboxy-terminal deletions of the same pioneer: both alterations seem to block relevant enhancers at the primed state without preventing genomic recruitment[32].

M.M.H. Our focus on the biological outcome of pioneer factor function has led us away from considering a binary classification and towards a more nuanced view that includes protein-intrinsic biochemical properties as well as specific features of the cells in which the factor is expressed[27]. Within a specific cellular context, pioneer factors are interfacing with a distinctive chromatin environment and a unique suite of cofactors. It is evident from numerous reports that both cellular features have an impact on the ability of pioneer factors to access individual regions of closed chromatin and to promote accessibility. Although pioneering activity is influenced by cellular context, there are concrete examples of proteins that clearly function as pioneer factors. For example, in the early *Drosophila* embryo, Zelda is required to define hundreds of *cis*-regulatory modules that drive widespread genome activation. Although other proteins are required for chromatin accessibility at a small subset of loci, many studies have demonstrated that Zelda and the DNA motif it binds have a uniquely important role in promoting chromatin accessibility, facilitating TF binding and driving this developmental transition[26,33]. Moreover, the initial discovery of Zelda as a pioneer factor essential for activation of the zygotic genome has led to the identification of pioneer factor-mediated zygotic genome activation in the early embryos of all organisms studied to date, including zebrafish, frogs, mice and humans[25,26]. Together, this suggests that although it is important to recognize that pioneer factors may not fall into a category completely distinct from other TFs, they possess properties that set them apart from other factors and these properties are regulated by their cellular context.

What is known about the context dependency of pioneer factor activity?

M.M.H. Pioneer factor binding and function are dependent on both the genomic and the cellular contexts[20,24,27,34]. Despite their ability to target closed, nucleosomally occupied binding sites, pioneer factors do not bind to all genomic instances of their target motifs. Both chromatin structure and the location of cofactor-binding motifs can regulate the ability of pioneer factors to occupy specific loci. Because chromatin structure and cofactor availability are variable between cell types, pioneer factors also have cell type-specific patterns of binding and activity. In addition, tissue-specific features of the pioneer factor (that is, post-translational modifications, levels of expression or expression of distinct isoforms) may also contribute to tissue-specific pioneer factor activity. Overall, data from studies of multiple different pioneer factors clearly indicate that epigenetic and tissue-specific features regulate pioneer factor genomic occupancy as well as the ability of the factor to promote accessible chromatin.

K.S.Z. Pioneer factors clearly exhibit cell-context binding to their target sequences, showing that stable binding involves interactions with other proteins that were not previously bound to a site until the pioneer factor enabled such. By contrast, assessments of binding events that are below the threshold of statistically called peaks show that pioneer factors exhibit many more sub-threshold binding events than non-pioneer TFs[19,20,35], perhaps reflecting a scanning mode that is less dependent upon the cellular context. Pioneer factors have varying abilities to target different forms of silenced chromatin, marked by histone H3 trimethylated at K27 (H3K27me3), H3K9me3 or neither but compacted by linker histone[2], indicating how the chromatin environment modulates pioneering.

J.D. Most pioneer factors seem unable to target sites within constitutive heterochromatin, that is, heterochromatin that is enriched in the repressive modification H3K9me3. For example, in pituitary cells where the pioneer PAX7 specifies the melanotrope fate, the well-documented myogenic target sites of PAX7 are completely inaccessible (and within H3K9me3-rich chromatin), and the reverse is true of pituitary-specific sites in muscle cells. Constitutive heterochromatin is thus a barrier between very different cell fates[10]. Some pioneer factors, such as PAX7, preferentially target closed chromatin that is relatively enriched in the repressive modification H3K9me2, which is associated with facultative heterochromatin. Chromatin compaction involves the linker histone H1, and the level of H1 limits the strength and onset of pioneer action[36]. It is displaced after pioneer action, consistent with relaxation of chromatin structure.

Of note, many pioneer factors exhibit large (30–60%) subsets of genomic recruitment sites (as measured by ChIP–seq) where their presence does not seem to change chromatin organization. Some have called these subsets 'resistant'[10], and they may reflect an inability to initiate chromatin opening because of a barrier, or the ability of pioneer factors to interact with heterochromatin components. Weak chromatin interactions may serve to scan the genome[28], and in some instances the underlying binding sites have low DNA sequence conservation whereas in other instances the underlying DNA sequences have similar properties to subsets that are successfully activated by the pioneer[37]. Resistant sites are often not within H3K9me3-rich heterochromatin. Hence, other barriers to pioneer action must be involved.

M.L.B. Context dependence is a very interesting and not well-understood aspect of pioneer factors. Clearly, there is some form of nucleosomal context dependence on pioneer binding, as not all nucleosomes that contain

a particular DNA recognition sequence for a given pioneer are occupied by that factor; for example, looking at FOXA2 binding to potential regulatory regions across three cell types showed that only ~6–13% of sites matching a FOXA2 motif were occupied by FOXA2 (ref. 20). One form of nucleosomal sequence context dependence is that pioneer factors typically display some positional preference in terms of where, within nucleosomes, they bind: many TFs bind near the entry or exit sites of nucleosomes ('end binders'), some bind periodically, a few have been reported to be gyre spanning, whereas others bind at or near the dyad[15,38]. (For most TFs that have been reported to bind at the ends of nucleosomes, it remains unclear whether they actually engage nucleosomal DNA binding sites, or whether they are binding to free DNA that is transiently unwrapped from the histone core.) In vitro binding selections of protein binding to nucleosomes, based on the high-affinity, synthetic '601' DNA sequence[14,16] or those built on completely randomized DNA[15] (which, in practice, selects for DNA that has a favourable sequence composition for stable assembly into nucleosomes), have provided some information on how different factors can bind in the context of such nucleosomes; however, more studies are needed to understand how the genomic sequence context of nucleosomes influences pioneer binding[17].

How do the dynamic properties of chromatin affect the actions of pioneer factors?

K.S.Z. Fluorescence recovery after photobleaching and single-molecule tracking studies of core histones have revealed that all chromatin is dynamic, in terms of nucleosome mobility. Yet the same studies reveal a range of dynamic mobility. For example, chromatin at the nuclear and nucleolar peripheries is far less dynamic than at many sites within the nucleus[11]. Early on, we found that pioneer factors were dynamic in chromatin but that their nucleosome affinity enabled non-specific scanning of closed chromatin, whereas non-pioneer factors were deficient in this activity[2,11]. In a new study, we directly compared the pioneer factors SOX2 and FOXA1 and the non-pioneer factor HNF4A regarding whether they target high-turnover nucleosomes in vivo[35]. The data showed that the non-pioneer factor HNF4 targets nucleosome regions that are highly dynamic, whereas most of the sites targeted by the pioneer factors turn over very slowly. During cell fate changes, the ability to scan silent genes in less dynamic chromatin seems crucial.

J.D. The dynamic nature of chromatin states only adds different 'flavours' to the poorly defined notion of 'closed chromatin'. For example, recent work[39] defined different subsets of heterochromatin states associated with the peripheral nucleoplasm or B compartment: these subdomains, either defined by physical measurements or through association with peripheral nuclear components such as lamins, now present with better opportunities to define the so-called closed chromatin states that may be permissive for pioneer recruitment and action.

Furthermore, the compartmentalization of chromatin domains into peripheral inactive compartment B versus central nucleoplasmic active compartment A places the context of pioneer action at discrete enhancer sequences within the context of global compartment switching observed for entire topologically associating domains (TADs) following pioneer action. This begs the, so far unanswered, question of the relationship between local pioneer action at enhancers and large domain (TAD) compartment switching.

What are the key future directions and unanswered questions in pioneer factor research

K.S.Z. How do different pioneer factors interact with the histone octamer, and how do the different interactions lead to local chromatin exposure and enable nucleosome remodeller function? How do different pioneer factors target different forms of silenced chromatin, or heterochromatin? Can the ability to target a particular type of heterochromatin be grafted from one pioneer factor to another? Answers to these questions will help us create hybrid factors that may be more efficient at overcoming the chromatin barrier to induce new cell fates.

M.L.B. Beyond positional preferences for where pioneer factors bind within nucleosomes, some other context feature(s) — either the presence of a cofactor(s) or, possibly, the nucleosomal context itself — must be regulating pioneer binding. Even the same genomic binding site for a given pioneer may be occupied differentially across cell types, as was seen for FOXA2, whose binding across three cell types was found to be mostly cell type-specific[20]. There are many potential sources of cell type-specific pioneer binding. In vitro binding to synthetic nucleosomal sequences based on the high-affinity '601' model sequence template showed that OCT4 alone bound more preferentially to the entry or exit site in one half of the nucleosomes, whereas OCT4 and SOX2 together bound approximately symmetrically across the dyad axis, in a strongly cooperative manner[14].

We are just beginning to scratch the surface of what controls pioneer binding to nucleosomes. In addition to cofactors, which may be expressed in a cell type-dependent manner, epigenomic features of nucleosomes can also differ across cell types and their roles in regulating pioneer binding have been essentially unexplored. In a recent study, the DNA binding domain of OCT4 was found to interact with the amino-terminal tail of H3, and, intriguingly, the post-translational modification H3 acetylated at K27 (H3K27ac) was found to increase OCT4 binding to more internal sites on the nucleosomes and to enhance OCT4–SOX2 cooperative binding to nucleosomes[17]. These results suggest that a broad set of histone post-translational modifications may be an additional axis in the multidimensional condition space modulating TF binding to nucleosomes. My group is working to understand which sequence and epigenomic context features of nucleosomes influence binding by various pioneer factors.

J.T. To me, the most interesting future directions are elucidating the sequence of events leading to opening of chromatin at different scales and understanding how TF–nucleosome interactions define local and megabase-scale chromatin states. I am also particularly interested in how pioneer factors can alter the higher-order chromatin state to make regions of chromatin accessible to the large macromolecular complexes involved in transcription (many TFs are relatively small, <5 nm in diameter, whereas the Mediator–RNA polymerase II complex is very large at >20 nm in diameter)[40].

It is clear that to be a pioneer factor, a TF must be able to recognize its motif on nucleosomal DNA. Some factors such as SOX2 can destabilize the nucleosome by their binding energy[23], but keeping chromatin open in vivo is still a dynamic process requiring continuous activity of chromatin remodellers[41,42]. Our knowledge is thus at the level where we have example cases and understand some high-level processes, but at the same time we lack almost completely the ability to predict how sequence determines TF binding and the chromatin state. In other words, we have

conceptual understanding but lack predictive understanding. We do not even know the full complement of pioneer factors in any organism, let alone all their affinities to all DNA sequences in the presence and absence of nucleosomes. We also do not understand how heterochromatin features such as linker histones and HP1 affect TF binding. Much remains to be understood also at the conceptual level — our picture of cells is often that of the steady state, and very little is currently known about the sequence of events that is initiated after the binding of pioneer factors, and how the different binding modes could affect these processes. For example, it is plausible that TFs with different binding modes could have different functions, for example, the oriented mode could define the border between open and closed chromatin, and the end binding and dyad binding modes could help stabilize some nucleosomes at very specific genomic positions. In summary, current evidence indicates that there is a very rich interaction landscape between TFs and nucleosomes, much of which remains uncharacterized. I am very excited about the overall field and believe that many ground-breaking discoveries await us.

J.D. The foremost obvious question with regards to pioneer action is the nature of the chromatin environments that are either permissive or restrictive to their action. Another important question is whether all pioneer factors share similar properties with regards to their permissive chromatin environments. Could differences in this chromatin permissiveness be a basis for a hierarchy of pioneer factors defined by their differential access to different chromatin types? And could this define different steps in cascades of tissue specification in early development?

With regards to the mechanism of action, the key question is clearly to define the minimal process triggered by pioneer factors following their recruitment to closed chromatin sites. Some recent work[36] defined two steps in this process, with the first involving biochemical actions whereas the second step requires the passage through cell division and its role in nuclear compartment switching. In this model, what would be the minimal requirement to trigger the process?

As pioneer factors implement epigenetic memory, what is required to maintain this memory, and what is the chromatin status of enhancers in the memory state? Furthermore, what could be the mechanisms to reverse this memory state or the action of pioneer factors? It is well-documented that in addition to opening new enhancer repertoires, pioneer factors participate in repression of alternate cell fates: we know little of the pioneer-associated mechanisms for repression of alternate cell fates.

M.M.H. There remain many exciting, unanswered questions regarding pioneer factors. These include identifying the barriers to both genome occupancy and activity. It will be important to address how these barriers regulate the tissue-specific functionality of pioneer factors and how dysregulation of pioneer factors lead to disease. Many studies have focused on the unique aspects that enable pioneer factors to engage closed chromatin, but it is also critical to determine how this binding results in a change in chromatin state. Mechanisms may differ for individual factors; but given that not all pioneer factor-bound regions become accessible, this represents another important aspect of regulation. Additionally, recent breakthroughs in live-cell imaging have made it apparent that TF binding to chromatin is relatively transient, and that many factors are non-uniformly distributed within the nucleus. Future research into the domains that regulate binding dynamics and subnuclear distribution will help shed light on the functional importance of the subnuclear distribution of pioneer factors. Ultimately, these studies will have implications in understanding how pioneer factors regulate both the local chromatin environment and the broader structure of the genome within the 3D space of the nucleus to regulate gene expression networks.

Martha L. Bulyk [1,2] ✉, **Jacques Drouin** [3,4,5] ✉, **Melissa M. Harrison** [6] ✉, **Jussi Taipale** [7,8,9] ✉ & **Kenneth S. Zaret** [10,11] ✉

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [2]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [3]Laboratoire de génétique moléculaire, Institut de recherches cliniques de Montréal (IRCM), Montreal, Quebec, Canada. [4]Département de Biochimie, Université de Montréal, Montreal, Quebec, Canada. [5]Department of Biochemistry, McGill University, Montreal, Quebec, Canada. [6]Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI, USA. [7]Department of Biochemistry, University of Cambridge, Cambridge, UK. [8]Applied Tumour Genomics Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [9]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [10]Institute for Regenerative Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [11]Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
✉e-mail: mlbulyk@genetics.med.harvard.edu; jacques.drouin@ircm.qc.ca; mharrison3@wisc.edu; ajt208@cam.ac.uk; zaret@pennmedicine.upenn.edu

Published online: 22 September 2023

References

1. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
2. Zaret, K. S. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **54**, 367–385 (2020).
3. Frederick, M. A. et al. A pioneer factor locally opens compacted chromatin to enable targeted ATP-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* **30**, 31–37 (2023).
4. Tsukiyama, T., Becker, P. B. & Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525–532 (1994).
5. Balsalobre, A. & Drouin, J. Pioneer factors as master regulators of the epigenome and cell fate. *Nat. Rev. Mol. Cell Biol.* **23**, 449–464 (2022).
6. Roberts, G. A. et al. Dissecting OCT4 defines the role of nucleosome binding in pluripotency. *Nat. Cell Biol.* **23**, 834–845 (2021).
7. Blassberg, R. et al. Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. *Nat. Cell Biol.* **24**, 633–644 (2022).
8. Fernandez Garcia, M. et al. Structural features of transcription factors associating with nucleosome binding. *Mol. Cell* **75**, 921–932.e6 (2019).
9. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
10. Mayran, A. et al. Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat. Genet.* **50**, 259–269 (2018).
11. Lerner, J. et al. Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin. *Mol. Cell* **79**, 677–688.e6 (2020).
12. Jozwik, K. M. & Carroll, J. S. Pioneer factors in hormone-dependent cancers. *Nat. Rev. Cancer* **12**, 381–385 (2012).
13. Iwafuchi, M. et al. Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.* **52**, 418–427 (2020).
14. Michael, A. K. et al. Mechanisms of OCT4–SOX2 motif readout on nucleosomes. *Science* **368**, 1460–1465 (2020).
15. Zhu, F. et al. The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
16. Yu, X. & Buck, M. J. Defining TP53 pioneering capabilities with competitive nucleosome binding assays. *Genome Res.* **29**, 107–115 (2019).
17. Sinha, K. K., Bilokapic, S., Du, Y., Malik, D. & Halic, M. Histone modifications regulate pioneer transcription factor cooperativity. *Nature* **619**, 378–384 (2023).

18. Delás, M. J. et al. Developmental cell fate choice in neural tube progenitors employs two distinct *cis*-regulatory strategies. *Dev. Cell* **58**, 3–17.e8 (2023).
19. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
20. Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).
21. Skene, P. J. & Henikoff, S. A simple method for generating high-resolution maps of genome-wide protein binding. *eLife* **4**, e09225 (2015).
22. Neikes, H. K. et al. Quantification of absolute transcription factor binding affinities in the native chromatin context using BANC-seq. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-01715-w (2023).
23. Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669–672 (2020).
24. Larson, E. D., Marsh, A. J. & Harrison, M. M. Pioneering the developmental frontier. *Mol. Cell* **81**, 1640–1650 (2021).
25. Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition revisited. *Development* **146**, dev161471 (2019).
26. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 (2019).
27. Gibson, T. J. & Harrison, M. Protein-intrinsic properties and epigenetic effects regulate pioneer-factor binding and function. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.18.533281 (2023).
28. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
29. Meers, M. P., Janssens, D. H. & Henikoff, S. Pioneer factor–nucleosome binding events during differentiation are motif encoded. *Mol. Cell* **75**, 562–575.e5 (2019).
30. Morgunova, E. & Taipale, J. Structural insights into the interaction between transcription factors and the nucleosome. *Curr. Opin. Struct. Biol.* **71**, 171–179 (2021).
31. Donovan, B. T. et al. Basic helix–loop–helix pioneer factors interact with the histone octamer to invade nucleosomes and generate nucleosome-depleted regions. *Mol. Cell* https://doi.org/10.1016/j.molcel.2023.03.006 (2023).
32. Bascunana, V. et al. Chromatin opening ability of pioneer factor Pax7 depends on unique isoform and C-terminal domain. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkad520 (2023).
33. Brennan, K. J. et al. Chromatin accessibility in the *Drosophila* embryo is determined by transcription factor pioneering and enhancer activation. *Dev. Cell* https://doi.org/10.1016/j.devcel.2023.07.007 (2023).
34. Hansen, J. L., Loell, K. J. & Cohen, B. A. A test of the pioneer factor hypothesis using ectopic liver gene activation. *eLife* **11**, e73358 (2022).
35. Lerner, J., Katznelson, A., Zhang, J. & Zaret, K. S. Distinct chromatin scanning modes lead to targeting of compacted chromatin by pioneer factors FOXA1 and SOX2. *Cell Rep.* **42**, 12748 (2023).
36. Gouhier, A., Dumoulin-Gagnon, J., Lapointe-Roberge, V., Balsalobre, A. & Drouin, J. Pioneer factor Pax7 initiates two-step cell-cycle dependent chromatin opening. Preprint at *bioRxiv* https://doi.org/10.1101/2022.11.16.516735 (2022).
37. Pelletier, A. et al. Pax7 pioneer factor action requires both paired and homeo DNA binding domains. *Nucleic Acids Res.* **49**, 7424–7436 (2021).
38. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* **23**, 804–809 (2009).
39. Shah, P. P. et al. An atlas of lamina-associated chromatin across twelve human cell types reveals an intermediate chromatin subtype. *Genome Biol.* **24**, 16 (2023).
40. Maeshima, K. et al. The physical size of transcription factors is key to transcriptional regulation in chromatin domains. *J. Phys. Condens. Matter* **27**, 064116 (2015).
41. Schick, S. et al. Acute BAF perturbation causes immediate changes in chromatin accessibility. *Nat. Genet.* **53**, 269–278 (2021).
42. Iurlaro, M. et al. Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nat. Genet.* **53**, 279–287 (2021).

Acknowledgements

M.L.B. is supported by US National Institutes of Health (NIH) grant R01HG012246. Research in the laboratory of J.D. is supported by the Canadian Institutes of Health Research (CIHR) and Compute Canada. M.M.H. is funded by NIH R35GM136298 and R01NS111647, and is a Romnes Faculty Fellow and Vilas Faculty Mid-Career Investigator. K.S.Z. is supported by NIH grant R01 GM36477.

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.