Pioneer factors as master regulators of the epigenome and cell fate

Aurelio Balsalobre and Jacques Drouin ✉

Abstract | Pioneer factors are transcription factors with the unique ability to initiate opening of closed chromatin. The stability of cell identity relies on robust mechanisms that maintain the epigenome and chromatin accessibility to transcription factors. Pioneer factors counter these mechanisms to implement new cell fates through binding of DNA target sites in closed chromatin and introduction of active-chromatin histone modifications, primarily at enhancers. As master regulators of enhancer activation, pioneers are thus crucial for the implementation of correct cell fate decisions in development, and as such, they hold tremendous potential for therapy through cellular reprogramming. The power of pioneer factors to reshape the epigenome also presents an Achilles heel, as their misexpression has major pathological consequences, such as in cancer. In this Review, we discuss the emerging mechanisms of pioneer factor functions and their roles in cell fate specification, cellular reprogramming and cancer.

Pioneer factors
Transcription factors with the unique ability to bind DNA target sites within closed chromatin, typically regulatory elements such as enhancers with undetectable levels of active-chromatin modifications before pioneer action.

Epigenetic memory
The processes and mechanisms responsible for maintenance of chromatin structure and gene activity status, either as heterochromatin or as active chromatin. The core epigenetic-memory mark of inactive chromatin is DNA CpG methylation.

Laboratoire de génétique moléculaire, Institut de recherches cliniques de Montréal, Montreal, QC, Canada.

✉*e-mail: jacques.drouin@ircm.qc.ca*

https://doi.org/10.1038/s41580-022-00464-z

Pioneer factors are transcription factors that have the unique ability to open closed chromatin to activate gene expression[1–3]. For development of complex organisms, this is an important process as it allows various levels of transcription-factor access to different parts of the genome. Beyond DNA compaction, the organization of the genome into disparate chromatin domains creates an additional level of gene regulation through epigenetic mechanisms. In development and during the course of cell differentiation cascades, mechanisms of epigenetic modification of DNA and histones that unravel or mask domains of the genome through chromatin remodelling are crucial for tissue specification and organogenesis. Pioneer factors are master regulators of these epigenetic processes and hence are master regulators of development and cell differentiation.

Whereas numerous mechanisms are in place to ensure stability of the genome and its expression[4], epigenetic remodelling by pioneer factors is the means to override these mechanisms so as to alter cell fates during precisely orchestrated development. In this role, pioneers have the unique ability to recognize and bind their target DNA sequences within closed chromatin[5]; these sites are not accessible to most transcription factors. Pioneers then initiate chromatin reorganization and opening through a process that is only now becoming understood, particularly the initial events. Most targets of pioneer action studied so far are intergenic or intronic enhancers[6–12]; the importance of pioneer action at promoters is not so well defined. Similarly, the function of pioneers in repression of gene expression and the role of this repression in altering cell fate choices have not been much studied.

The first biochemical evidence in support of unique pioneer properties came with the discovery that the pioneer FOXA (also known as HNF3) can recognize and interact with its target DNA sequence wrapped around histones in nucleosomes[13]. This ability appeared to be in contrast to the ability of most transcription factors, which can bind their target sequence only in naked DNA. Since it was already known that active regulatory sequences such as enhancers and promoters are more accessible in chromatin than other sequences[14,15], these findings clearly separated transcription factors into those that access their target sequences only when the DNA is within open chromatin, and pioneers that can do so even in closed chromatin. In recent years, the investigation of pioneers and the scope of their activity revealed their crucial importance for cell differentiation in normal development[3], and also their power as tools for cellular reprogramming[16] and novel cell therapies[17]. The power of pioneers to remodel the epigenome and cell identity is not devoid of unwanted consequences; hence, it is not surprising that inappropriate expression of pioneers or the creation of chimeric pioneers through mutations or chromosomal translocations is implicated in cancer[18–21].

In this Review, we discuss current insights into the mechanisms of pioneer action, from the unique features of their chromatin interactions to initial chromatin remodelling, enhancer activation and implementation of epigenetic memory. We also discuss epigenetic memory in the context of transcriptional memory, which marks genes for subsequent reactivation. The functions of the currently known pioneer factors in cell fate decisions are

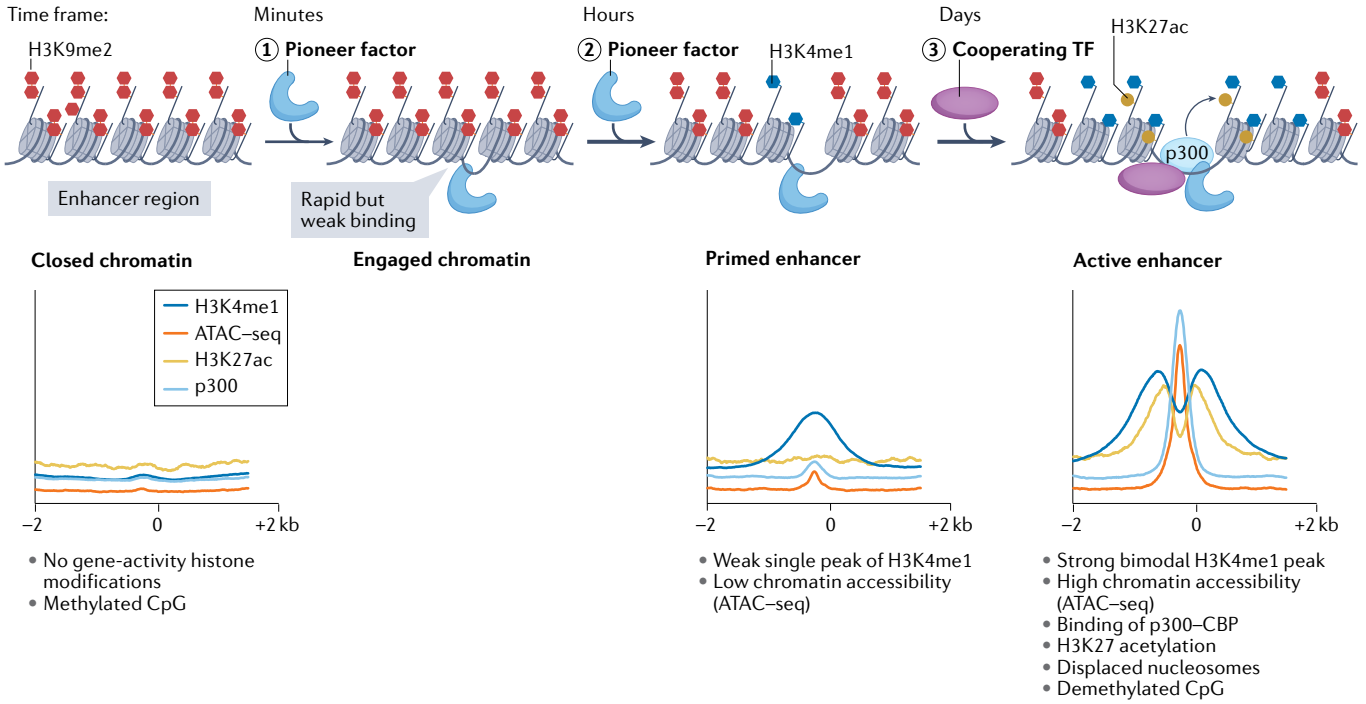

Fig. 1 | **The pioneer action.** Pioneer factors bind to sites in closed chromatin (BOX 1) that may harbour the facultative heterochromatin mark dimethylated histone H3 Lys9 (H3K9me2). Their initial binding is rapid but weak (step 1); pioneer–DNA interaction may displace DNA from the nucleosome, as shown for SOX2 and OCT4 (REFS[26,27]). The engaged chromatin sites then exhibit an increase in pioneer recruitment and stabilization of pioneer–chromatin interactions (step 2). The accompanying initiation of chromatin opening by pioneers is characterized by low levels of monomethylated histone H3 Lys4 (H3K4me1) peaking at enhancer cores and the emergence of low levels of chromatin accessibility (as measured by assay for transposase-accessible chromatin with high-throughput sequencing (ATAC–seq)). These enhancers are 'primed'. Step 3 of chromatin opening likely involves recruitment of cooperating transcription factors (TFs) and occurs within the timescale of DNA replication and cell division. The resulting 'active enhancer' state is characterized by a bimodal distribution of H3K4me1, reflecting nucleosome displacement at the centre of the enhancer together with high chromatin accessibility, recruitment of the general transcription co-activator p300–CREB-binding protein (CBP) and acetylated H3K27 (H3K27ac). In parallel with chromatin remodelling (not shown), mammalian enhancers opened by pioneer factors undergo DNA demethylation. The boxes present average profiles for the indicated marks centred at the pioneer binding site. The indicated time frames integrate data from different systems[6,7,11,43,44].

briefly described and discussed in the context of cell-type reprogramming. Finally, we discuss the implications of pioneer activities in cancer and development.

Chromatin opening by pioneer factors

The initial steps of pioneer action are the least understood. How do pioneers interact with chromatin? Is heterochromatin accessible to pioneers? What chromatin modification impairs or facilitates the remodelling process? What chromatin modification initiates the remodelling required to open chromatin? What other proteins are involved in pioneer action?

Chromatin features of pioneer targets

The DNA target sequences recognized by pioneers are likely to be wrapped around nucleosomal histones in closed chromatin. It is thus expected that pioneers have the ability to recognize DNA in nucleosomes within nuclear compartments that may contain heterochromatin.

Pioneer interaction with nucleosomal DNA. DNA footprinting experiments performed in developing tissues showed that the DNA-binding sites of the transcription factors FOXA and GATA are occupied before onset of target gene expression[22]. The presence of packed nucleosomes in the targeted chromatin domains could have been a barrier to their recruitment, but it was not. FOXA was shown to bind target sequences in nucleosomal DNA, and in vitro it appeared to be more potent than GATA in nucleosome binding[5]. Since most pioneer factors have the ability to bind nucleosomal DNA, this property appears to be typical of pioneers[23], including of the *Drosophila melanogaster* pioneer Zelda[24]. However, the extent to which nucleosomal DNA binding is unique to pioneers has recently been put into context through extensive in vitro binding studies[25].

Recent structural studies revealed the nature of interactions between the pioneer factors SOX2 and OCT4 (which are pluripotency-promoting factors) and GATA3 and their DNA targets present in nucleosomes[26–28]. These factors yielded stable complexes with their target DNA suitable for structural studies when the sequences were located near the DNA entry or exit points of nucleosomes, which, in some instances, distorts interactions between DNA and nucleosomal histones (FIG. 1, step 1). Hence, these interactions may in themselves trigger some relaxation of DNA–nucleosome interactions. Interestingly, the nucleosome-binding ability of SOX2–OCT4 heterodimers is greater than that of

Transcriptional memory
The processes and mechanisms by which the first round of gene transcription remodels the epigenome for quicker transcription upon subsequent reactivation.

DNA footprinting
Techniques to reveal the presence of DNA-bound proteins through their impairment of access to DNA.

Basic helix–loop–helix
(bHLH). Structure of a DNA-binding domain of a family of transcription factors.

Pioneer factor-resistant sites
Chromatin recruitment sites typically showing weak pioneer binding that does not produce any change in chromatin organization.

Melanotrope cell
Endocrine cell of the pituitary intermediate lobe that secretes the pro-opiomelanocortin-derived hormone α-melanotropin.

Constitutive heterochromatin
Part of the genome in which the chromatin is permanently 'closed' and transcriptionally inactive; typically marked by trimethylated histone H3Lys9 (H3K9me3).

Facultative heterochromatin
Part of the genome in which transcriptionally inactive chromatin is amenable to cell type-specific activation; enriched in dimethylated histone H3 Lys9 (H3K9me2).

Topologically associating domains
(TADs). Chromatin domains that are partially insulated from flanking domains by boundaries that contain convergent binding sites for CCCTC-binding factor (CTCF).

Assay for transposase-accessible chromatin with high-throughput sequencing
(ATAC–seq). Technique to map chromatin accessibility based on insertion frequencies of a transposable element within accessible DNA.

either protein alone, which is in line with the cooperative activity of these pioneers in cell reprogramming into pluripotent cells[27]. It is, however, noteworthy that for GATA3, the sites that are subject to chromatin opening in vivo tend to harbour the target sequence about two helical turns closer to the dyad axis[28]. This structure was not determined, but it is suggested to impair DNA–nucleosome interaction. It was proposed that most pioneers capable of nucleosome binding (such as FOXA, OCT4 and PU.1) interact on one side of the nucleosome with their target DNA sequence through a recognition α-helix domain[29] and that this initial contact might involve only a partial DNA motif as shown for OCT4, SOX2 and Krüppel-like factor 4 (KLF4)[27,30]. The basic helix–loop–helix (bHLH) pioneer factor ASCL1 may be a particular case of recognition on both sides of the nucleosome, but structural studies are needed to elucidate this interaction[29].

Heterochromatin histone modifications that impair or facilitate pioneer action. Another property observed with many pioneers is their recruitment to many target sites without an apparent effect. The typically weaker pioneer recruitment sites of this class were termed 'pioneer factor-resistant sites'[7] or 'low-level sampling sites' (denoting low-to-intermediate chromatin immunoprecipitation followed by sequencing signals that are above the genomic background but below the stringent peak calling threshold)[6]. For the pioneer PAX7, this subset of resistant sites has a similar distribution of cognate DNA sequence motifs as pioneered sites, yet PAX7 recruitment at these sites does not initiate any change in chromatin organization[7]. In this respect, PAX7 is an interesting case, because it has pioneer activity that specifies the pituitary intermediary lobe fate and melanotrope cell fate, but it also has crucial developmental activity in myogenic progenitors. Accordingly, subsets of muscle-specific PAX7 binding sites are known, yet these muscle-specific sites are not accessible, not even as 'resistant' recruitment sites, in pituitary cells[31]. This observation clearly highlights the important role of the chromatin environment, specifically of heterochromatin proteins, which likely define regions of heterochromatin that may or may not be permissive for recruitment of a particular pioneer. The nature of these heterochromatin proteins remains to be identified, whether they are permissive for pioneer recruitment or constitute a barrier to recruitment. Our knowledge of heterochromatin subtypes remains relatively vague.

Be that as it may, the role of heterochromatin as a barrier to pioneer recruitment is documented. The strongest barrier to pioneer recruitment is trimethylated histone H3 Lys9 (H3K9me3), which marks constitutive heterochromatin. Indeed, initial recruitment of pluripotency factors is prevented by this modification[32], and H3K9me3-marked constitutive heterochromatin is refractory to recruitment of other pioneers[7]. The importance of H3K9me3 as a barrier to cell differentiation and presumably pioneer factor activity is supported by experiments that inactivated an H3K9 methyltransferase, which resulted in perturbation of gene expression programmes[33]. Whereas H3K9me3 defines constitutive heterochromatin that is refractory to pioneer recruitment and activity, relative enrichment in the related dimethylated H3K9 (H3K9me2) is found at sites that are permissive for pioneer activity by PAX7 (REF.[7]). H3K9me2 is associated with facultative heterochromatin and it marks a subset of permissive closed chromatin (FIG. 1). It is noteworthy that another gene-repressing histone modification, H3K27me3, which is associated with the repressive activity of Polycomb group proteins, is not particularly enriched at pioneered sites, although it is present at some sites of pioneers other than PAX7 (REFS[7,34]). Some researchers have termed the pioneer-accessible chromatin 'naive' to indicate the absence of histone modifications associated with gene activity, but without reference to heterochromatin subtypes. The small number of heterochromatin protein markers limits the current ability to define chromatin environments that are either permissive or restrictive for pioneer recruitment and activity despite recent attempts to identify candidate proteins[35]. In BOX 1, we discuss the nature of the 'closed' chromatin that is permissive for pioneer action.

Large chromatin domains and nuclear compartments. Heterochromatin and euchromatin are not distributed randomly within the cell nucleus. Rather, heterochromatin domains are mostly physically associated with the nuclear lamina at the periphery of the nucleoplasm, whereas euchromatin domains occupy its centre[36]. These localizations are termed 'compartment A' and 'compartment B': upon activation, genes in heterochromatin-associated compartment B at the nuclear periphery appear to move into gene-active compartment A in parallel with changes in chromatin organization[37]. Whereas most studies of pioneer action have so far focused on remodelling of chromatin structure at enhancers (FIG. 1), pioneer action may involve entire genomic regions defined as topologically associating domains (TADs). Changes in genome topology as broadly revealed by a combination of chromatin conformation capture methods (for example, Hi-C) and assay for transposase-accessible chromatin with high-throughput sequencing (ATAC–seq) were shown to precede changes in gene expression during B cell reprograming by CCAAT/enhancer-binding protein-α (C/EBPα), OCT4, SOX2, KLF4 and MYC[38]. These changes involve increased interactions between loci within TADs and are manifested by the appearance of regions of increased chromatin accessibility ('ATAC–seq peaks') within loci subject to activation[38]. Similarly, in the pituitary gland, the appearance of accessible chromatin conformation in TADs containing hallmark genes of pituitary cell identity is dependent on PAX7 and its non-pioneer cooperating factor TPIT (also known as TBX19)[31]. The relationship between pioneer activity at specific regulatory elements and its large-scale effect on chromatin accessibility within TADs is not yet understood, but the implication of privileged interactions within TADs for transcription factor recruitment has been discussed[38]. Indeed, an enhancer displaced into a different TAD acquired properties of its new TAD location[39]. Recently, phase separation was found to be involved in spatial segregation

Box 1 | **Important unresolved issues**

The study of pioneer factors is still relatively new; hence, defining issues such as what pioneers are exactly and how or where they operate on the epigenome remain unresolved. Two such issues are discussed here.

What is 'closed' chromatin?

This Review and previous articles refer to the state of chromatin before pioneer action as 'closed' chromatin; but what is closed chromatin? The short answer is that it is a placeholder name for lack of more precise information. Closed chromatin is often defined by what it is not rather than by what it is: closed chromatin does not bear any mark of active chromatin such as DNA accessibility, monomethylated histone H3 Lys4 (H3K4me1), trimethylated H3K4 (H3K4me3) or acetylated H3K27 (H3K27ac). Is closed chromatin a form of heterochromatin? Constitutive heterochromatin is defined as large domains of compacted chromatin that are marked by H3K9me3, and is not a target of most pioneer factors studied so far, including pluripotency factors[7,32]. Sites in chromatin subjected to opening by pioneers are mostly found within facultative heterochromatin, which is loosely defined by its relative enrichment in H3K9me2 (REF.[7]). Beyond these chromatin modifications, it remains difficult to relate the closed-chromatin targets of pioneers to the broad domains that were historically defined as heterochromatin and euchromatin.

How different are pioneer factors from transcription factors?

Some researchers have argued that pioneers are not that different from most transcription factors and that a gradient of properties could demarcate the behaviour of so-called pioneers, for example, a greater affinity of pioneers for DNA target sites, for closed-chromatin proteins or for nucleosomes[135]. Others have emphasized interactions of pioneers with other transcription factors (dynamic-assisted loading) and the relative stabilization of pioneers through protein–protein interactions at specific genomic sites as parameters that define one factor as 'pioneering' and another as a 'cooperating' transcription factor[168]. Although these models may provide a formal or generalized description of pioneers, the unique property of pioneers observed in cells remains their ability to target sites in closed chromatin and initiate its opening. For this property to be accurately assessed, it is crucial to define the initial chromatin state of putative pioneered enhancers, as the action of a pioneer at a closed site must not be confounded with its action at primed sites (FIG. 2b). Failure to thus distinguish subsets of genomic target sites would result in ambiguous patterns of histone modifications particularly if data are analysed as average plots. For example, the description of sites pioneered by ASCL1 during reprogramming of fibroblasts to neurons as a trivalent mixture of gene-activating (H3K4me1 and H3K27ac) and gene-repressing (H3K9me3) histone modifications may be the result of analysing target site subsets that contain both closed and primed sites[129]. The unique properties of pioneers must therefore be determined at well-characterized closed-chromatin sites. That said, it is possible that different pioneers have different closed-chromatin substrates that remain to be defined. In summary, pioneer factors are unique because only these factors can initiate chromatin opening in a biological context, whereas other transcription factors are unable to do so.

of heterochromatin compartments from euchromatin compartments, which may involve specific heterochromatin proteins such as the heterochromatin protein 1 (HP1) factors; furthermore, phase separation of HP1α may be regulated through post-translational modifications such as HP1α phosphorylation[40]. It is thus possible that pioneers acting on a subset of target sites in heterochromatin and lamina-associated loci may trigger a change in the structure of heterochromatin proteins that ultimately results in compartment switching. At present, this model is speculative, but it defines the framework of pioneer recruitment to genomic sites and subsequent triggering of chromatin remodelling.

Initiation of pioneer action

The starting and end points of pioneer activity are clear: sites of closed chromatin with few distinguishing features become open and acquire chromatin modifications typical of active enhancers. The primed enhancer state appears to be an intermediate step in this process, but the earliest chromatin alterations following pioneer recruitment remain undefined.

Chromatin features at enhancers. Active enhancers have a distinctive chromatin profile that sets them apart from promoters, transcription start sites and other regulatory and structural chromatin domains[15]. Fully active enhancer elements are marked by monomethylated H3K4 (H3K4me1) and nucleosome depletion, the latter shaping the H3K4me1 distribution into a bimodal pattern of methylation at nucleosomes flanking both sides of the nucleosome-depleted region (FIG. 1). DNA in this region is more accessible than most of the genome on average, as revealed by various assays, including DNase I sensitivity, formaldehyde-assisted isolation of regulatory elements coupled with high-throughput sequencing (FAIRE–seq)[41] and, now most commonly used, ATAC–seq[42]. In addition, fully active enhancers are typically marked by recruitment of p300 and/or CREB-binding protein (CBP; also known as CREBBP), which are general transcription co-activators and have acetyltransferase activity and are responsible for H3K27 acetylation at nucleosomes flanking active enhancers[15]. An intermediate state of potentially active or 'primed' enhancers has low-level H3K4me1 without nucleosome depletion, and hence the H3K4me1 distribution pattern appears as a weak single peak, particularly in comparison with the bimodal distribution of H3K4me1 observed at fully active enhancers. Primed enhancers may have DNA accessibility signals that are much weaker than those of active enhancers as measured by ATAC–seq, which may or may not be revealed depending on the sensitivity of a particular data set.

The temporal sequence of pioneer activities. Extensive characterization of enhancer chromatin modifications throughout the genome before and after introduction of pioneer factors, together with the use of pioneer factor-mediated gene induction systems, showed that pioneer activity will result in the remodelling of subsets of enhancers towards the primed or fully active status[6,7]. Time course studies of pioneer activity further suggested a sequential process of chromatin remodelling at target enhancers (FIG. 1), and various pioneer-driven cell-type reprogramming schemes consistently exhibit such temporal properties[11,43,44]. Initial recruitment level of the pioneer PAX7 at its pioneered sites is weak but rapid (within 30 min), and during the following 12–24 h, the recruitment increases to its long-term, stable levels in parallel with the appearance of weak ATAC–seq signals and low-level H3K4me1 (REF.[7]). These initial alterations are consistent with a priming step of chromatin opening. Over the few next days, priming is followed by further increases in DNA accessibility (ATAC–seq), H3K4me1 and nucleosome depletion, and the appearance of a bimodal distribution pattern of H3K4me1 and recruitment of p300, which is followed by target gene activation (FIG. 1, step 3). This last step of chromatin remodelling occurs within the timescale of DNA replication and cell division; the role of replication in this step of chromatin remodelling remains

Active enhancers
Enhancers marked by nucleosome depletion, DNA accessibility, bimodal distribution of the active-chromatin modifications monomethylated histone H3 Lys4 (H3K4me1) and acetylated histone H3 Lys27 (H3K27ac) and the presence of the co-activator p300.

Primed enhancer
An enhancer with low levels of the epigenetic mark of activity monomethylated histone H3 Lys4 (H3K4me1) and DNA accessibility and no nucleosome depletion.

MNase–seq
A technique to visualize nucleosome positioning that uses partial DNA digestion with micrococcal nuclease 1 and high-throughput sequencing.

Nuclear receptors
Transcription factors that are activated upon binding of cognate ligands and translocate into the nucleus.

Mediator
A protein complex of about 30 proteins that integrates the inputs of enhancer-bound transcription factors to activate RNA polymerase II (Pol II) at promoters.

to be investigated and contrasted with the chromatin condensation that occurs at mitosis.

In view of the low levels of DNA accessibility and H3K4me1 in the transient state towards full enhancer activation, this transient state may be considered similar to stably primed enhancers. Be that as it may, this state represents the earliest evidence of chromatin alteration at enhancers targeted for full activation and appears to define, together with the ability to recognize target sites in closed chromatin, the truly unique properties of pioneers. Similarly, the use of micrococcal nuclease digestion with high-throughput sequencing (MNase–seq) to investigate nucleosome status at liver-specific enhancers revealed a slightly perturbed organization of 'accessible nucleosome configuration' at enhancers occupied by FOXA[12]. The monomethylation of H3K4 at pioneered enhancers may involve the myeloid/lymphoid or mixed-lineage leukaemia (MLL) histone lysine methyltransferase complex (also known as KMT2A), which methylates H3K4 as shown for MLL3 together with FOXA1 in breast cancer cells[45]. Despite its apparent importance in initiation of pioneer action, this step remains the least characterized, but obviously one that is crucial for understanding the unique aspect of chromatin remodelling by pioneers.

Cooperation with transcription factors

Whereas pioneer factors appear to be uniquely required to conduct the initial steps of chromatin opening described earlier herein, opening of enhancer chromatin for activity requires also chromatin remodelling complexes and enzymes that generally contribute to transcription activation[46] (FIG. 1). Chromatin opening at enhancers is evidenced by increased DNA accessibility and large gains in ATAC–seq signals, which correlate with nucleosome depletion at the centre of enhancer sequences, recruitment of the co-activator p300 or CBP and acetylation of H3K27. This process requires the SWI/SNF remodelling complex, and for the pioneer OCT4, it may involve direct interaction with the SWI/SNF ATPase component BRG1 (REF.[47]). Similarly, BRG1 is required[48,49] for function of the pioneers GATA3 (REF.[48]) and ISL1 (REF.[49]), as well as for the function of nuclear receptors[50]. Many transcription factors interact directly with p300–CBP[51], and SWI/SNF and p300 interact to modulate H3K27 acetylation[52]. However, it remains to be established whether it is DNA accessibility and/or the SWI/SNF complex that allows recruitment of p300 or whether H3K4 monomethylation leads to recruitment of p300–CBP and acetylation of H3K27.

The importance of cooperation and often synergism between transcription factors for transcription activation is well documented and typically relies on direct protein–protein interactions between the cooperating transcription factors[53]. Similarly, pioneers cooperate with other transcription factors for transcription activation once they have initiated chromatin opening (FIG. 2a), as was first shown for the endoderm-specifying pioneers FOXA and GATA4 (REFS[5,22]). Each factor may exhibit pioneer activity at subsets of sites where the other factor appears to fulfil the function of a cooperating factor. FOXA is the pioneer with which many transcription factors cooperate in different contexts (FIG. 2a): it cooperates with hepatocyte nuclear factor 1α (HNF1α)[54] or HNF4α[55] at liver-specific genes, with GATA6 in specifying the pancreatic fate[56] and in many hormone-dependent tissues and cancers, it is a pioneer for recruitment and action of nuclear receptors such as the oestrogen, androgen and glucocorticoid receptors[57–60]. The cooperation between FOXA and nuclear receptors results in either new or enhanced recruitment of nuclear receptors; this property led to the description of this cooperation as 'assisted loading'[61]. Others have coined the cooperation between pioneer and non-pioneer factors as 'settler factors', which require initial chromatin opening for action[62]. A controversy persists whether nuclear receptors themselves have pioneer activity[63–65].

It is the detailed analysis of mechanisms at sites of chromatin opening initiated by pioneers that defined the function of cooperating non-pioneers. The identification of steps in the pioneering process that are dependent on a pioneer and/or a cooperating non-pioneer allows a clearer definition of the unique properties that define pioneers[31]. PAX7 binds and initiates chromatin remodelling as revealed by low-level ATAC-seq signals, but requires the non-pioneer TPIT for complete chromatin opening and cell differentiation[31] (FIG. 2a). Similarly, the *D. melanogaster* pioneer Grainy head is sufficient for chromatin opening at epithelial enhancers, but it is not sufficient for transcription activation[66,67]. The essential role of the cooperating non-pioneer can be viewed as a pioneer-dependent trigger that is needed for recruitment of the remodelling complexes that open chromatin for enhancer activation. The role of a trigger transcription factor is well defined for pairs of cooperating transcription factors where one is a pioneer and the other is clearly a non-pioneer, but pairs comprising two pioneers exist, such as FOXA and GATA, or C/EBPα and PU.1.

This complexity begs the following question: what is the unique contribution of a trigger or cooperating transcription factor and why is the pioneer not sufficient on its own to conduct full chromatin opening? Is it only a question of numbers: a minimal number of transcription factors must be present to stably recruit coregulatory proteins or complexes such as p300–CBP, SWI/SNF, MLL or Mediator to proceed to step 3 of chromatin opening (FIG. 1, step 3). This model is probable in view of the wide diversity of transcription factors associated with pioneer cooperation (TABLE 1). The dual role of transcription factors as either a pioneer or a cooperating transcription factor is illustrated in FIG. 2b: a pioneer can either prime or fully activate enhancers, or can function as a cooperating transcription factor to activate enhancers that are already primed. The most likely model to account for these context-dependent activities invokes contributions from other, unrecognized transcription factors.

In the pituitary gland, not only are melanotrope-specific enhancers dependent on both PAX7 and TPIT for chromatin opening but so are entire TADs that encompass melanotrope-specific loci, such as the *PCSK2* and the *DRD2* TADs[31]. The unique property of the pioneer PAX7 is thus its ability to bind cognate

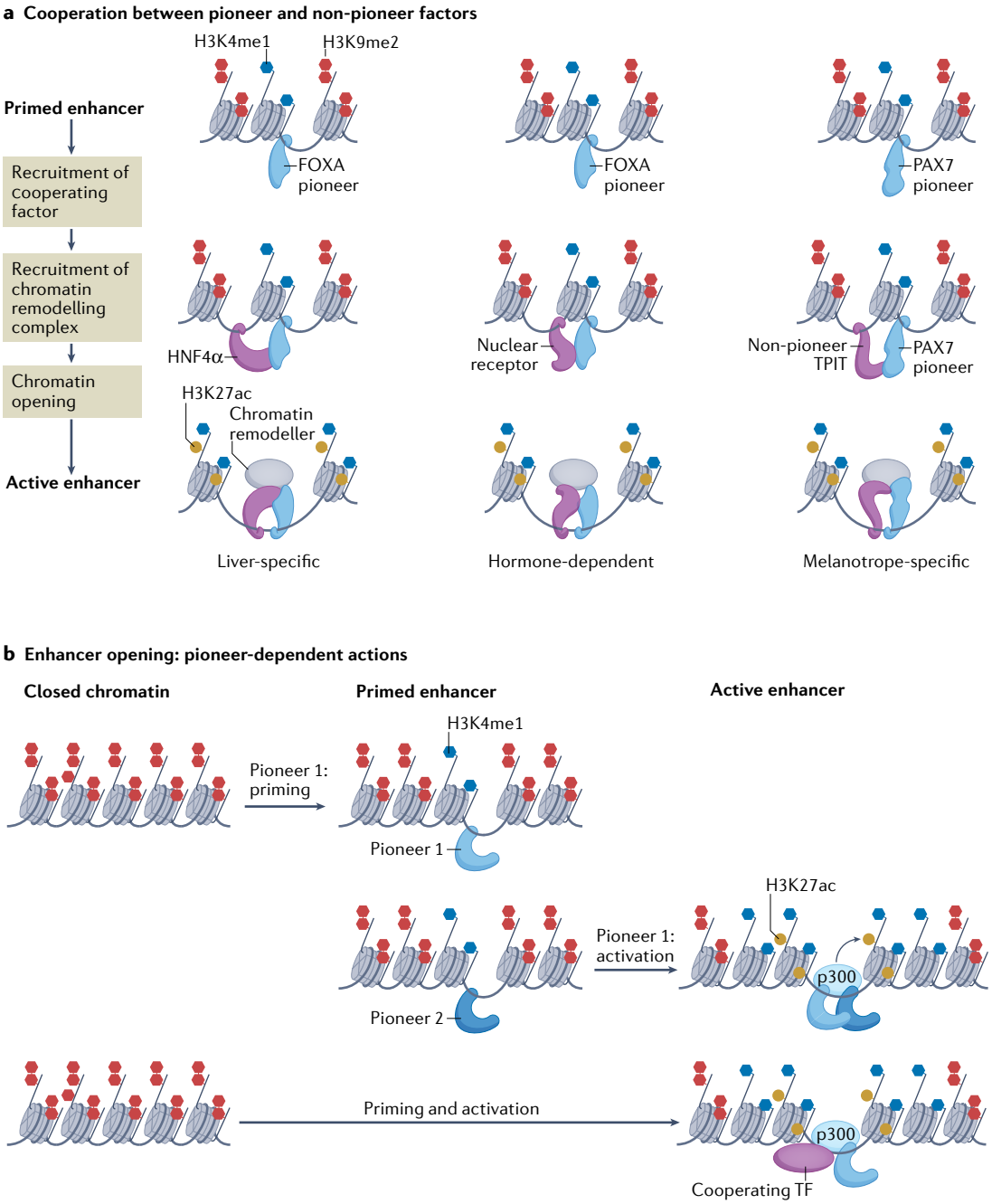

Fig. 2 | **Cooperation between pioneer and non-pioneer transcription factors.** Pioneer factors require other transcription factors to fully activate enhancers. **a** | In some cases, the same pioneer may cooperate with different transcription factors (TFs); for example, FOXA cooperates with hepatocyte nuclear factor 4α (HNF4α) to activate liver-specific genes, and FOXA cooperates with hormone-dependent nuclear receptors to implement hormone-responsive regulatory gene networks. However, it is the cooperation between pioneer and non-pioneer factors that provided data to discriminate between the unique chromatin-opening properties of pioneers and those of non-pioneers. For example, the pioneer factor PAX7 can access sites in facultative heterochromatin and initiate chromatin remodelling, but complete activation of target enhancers requires the non-pioneer TPIT. This interaction is associated with nucleosome depletion, presumably through recruitment of chromatin remodelling complexes such as SWI/SNF, and acetylated histone H3 Lys27 (H3K27ac). These different transcription factors interact directly with each other and with their cognate DNA sequences present within the enhancer. Target sequences for two such transcription factors are typically within 100 bp of each other but can also be many hundred base pairs from each other. **b** | At some enhancers, a pioneer factor (pioneer 1) may complete only the first step of enhancer opening, from closed chromatin to the primed state (top); a pioneer may appear to trigger complete enhancer activation if the cooperating non-pioneer transcription factor is already expressed in the cell (bottom). At enhancers that are already in a primed state, presumably through the action of another pioneer (pioneer 2), pioneer 1 may fulfil the role of a cooperating transcription factor and lead to complete enhancer activation (middle). H3K4me1, monomethylated histone H3 Lys4; H3K9me2, dimethylated histone H3 Lys9.

Table 1 | **Functions of pioneer factors in development**

Pioneer	Committed cell	Cell of origin	Cooperating transcription factor	DNA demethylation	Unreported properties	Refs
Pioneers with ascertained properties[a]						
FOXA1, FOXA2	Liver	Endoderm (h, m)	GATA4 C/EBPβ, HNF4α	Yes	–	5,12,55,105
	Pancreas	Endoderm (h, m)	GATA6	Yes	–	56,68
	Hormone dependent	Breast, prostate, etc. (h)	Nuclear receptors	Not reported	–	57–60
GATA3, GATA4	Liver	Endoderm (h, m)	FOXA, HNF1α	Not reported	–	5,48,54
PU.1	Macrophage	Myeloid progenitor (h, m)	C/EBPβ	Not reported	–	8,10,110,153
PU.1	DN3 T cell	DN1 T cell (m)	Not reported	Not reported	–	69
C/EBPα	Macrophage	Pre-B cell (m)	PU.1	Not reported	–	11
EBF1	Pro-B cell	Pre-pro-B cell (m)	PAX5	Yes	–	9,43
TCF1	DP T cell	Early thymic progenitor (m)	Not reported	Not reported	–	115
ASCL1	Neuron	Neurectoderm (m)	BRN2	Yes	–	44,89,129,154
MYOD	Myoblast	Mesoderm (m)	Not reported	Not reported	–	134,135,155
PAX7	Melanotrope	Pituitary progenitor (m)	TPIT	Yes	–	7,31,118
SOX2, OCT4, KLF4	iPS cell	Fibroblast (h, m)	MYC	Not reported	–	30,32,128
Zelda	Zygotic gene expression	Early embryo (dm)	–	Not relevant	–	24
GAF	Zygotic gene expression	Late embryo (dm)	–	Not relevant	–	96
Opa	Segmented expression	Late embryo (dm)	–	Not relevant	–	97,98
Grainy head (GRHL1, GRHL2, GRHL3)	Epithelial cell	Eye–antennal imaginal disc (dm)	–	Not relevant	–	66,67
Foxh1	Mesendoderm	Early gastrula (xl)	–	Not reported	–	106
Pioneers with partially established properties						
GR, AR, ER	Hormone dependent	Breast, prostate, etc. (h)	FOXA1	Not reported	Disagreement between authors	63–65
AP-1	Not relevant	Not relevant	GR	Not reported	Histone modifications, nucleosome binding	156
NEUROD1, NEUROD2	Neuron	Embryonic stem cell (m)	Not reported	Yes	Nucleosome binding	132
ISL1	Cardiomyocyte	Cardiac progenitor (m)	GATA4	Not reported	Histone modifications	49
STAT5	T_H9 cell	Naive spleen $CD4^+$ cell (h, m)	BATF	Not reported	Nucleosome binding	157
HOXA13	Distal limb	Limb mesoderm (m)	Not reported	Not reported	Histone modifications, nucleosome binding	158

AR, androgen receptor; ASCL1, Achaete–Scute homologue 1; BATF, basic leucine zipper transcriptional factor ATF-like; BRN2, brain-specific homeobox/POU domain protein 2; C/EBPβ, CCAAT/enhancer-binding protein-β; dm, *Drosophila melanogaster*; DN1, double negative 1; DN3, double negative 3; DP, double positive; EBF1, early B cell factor 1; ER, oestrogen receptor; FOXA, forkhead box protein A; FOXH1, forkhead box protein H1; GAF, transcription factor GAGA; GR, glucocorticoid receptor; GRHL, Grainy head-like protein homologue; h, human; HNF1, hepatocyte nuclear factor; iPS cell, induced pluripotent stem cell; KLF4, Krüppel-like factor 4; m, mouse; MYOD1, myoblast determination protein 1; NEUROD, neurogenic differentiation factor; OCT4, octamer-binding protein 4; Opa, Odd paired; STAT5, signal transducer and activator of transcription 5; TCF1, T cell-specific transcription factor 1; T_H9 cell, T helper 9 cell; TPIT, T-box transcription factor, pituitary; xl, *Xenopus laevis*.
[a]Ascertained pioneer properties include binding to closed chromatin, the appearance of histone modifications associated with gene activity and of DNA accessibility, interaction with nucleosomes and a biological role shown by loss of function and/or gain of function.

sites in facultative heterochromatin and to initiate subtle remodelling (FIG. 1, steps 1 and 2), but subsequent steps in the pioneering process require cooperation with TPIT (FIG. 2), including stable implementation of chromatin modifications associated with gene activity (FIG. 1, step 3) and TAD DNA accessibility. Similar relationships appear to be involved in the pioneer action of FOXA2 with GATA4 (REF.[6]) and other cooperating pioneers such as C/EBPα and PU.1 (REFS[68,69]), or with pluripotency factors[38].

Maintenance of the epigenome

The local structure of chromatin defines its potential and purpose, simplistically viewed as active or repressed chromatin. Maintenance of these chromatin states is crucial for the stability of gene expression programmes and cell identity, and therefore there are mechanisms that maintain chromatin states following the passage of the replication fork. It appears that at the gene-repressed heterochromatin, the pre-existing, histone-modified

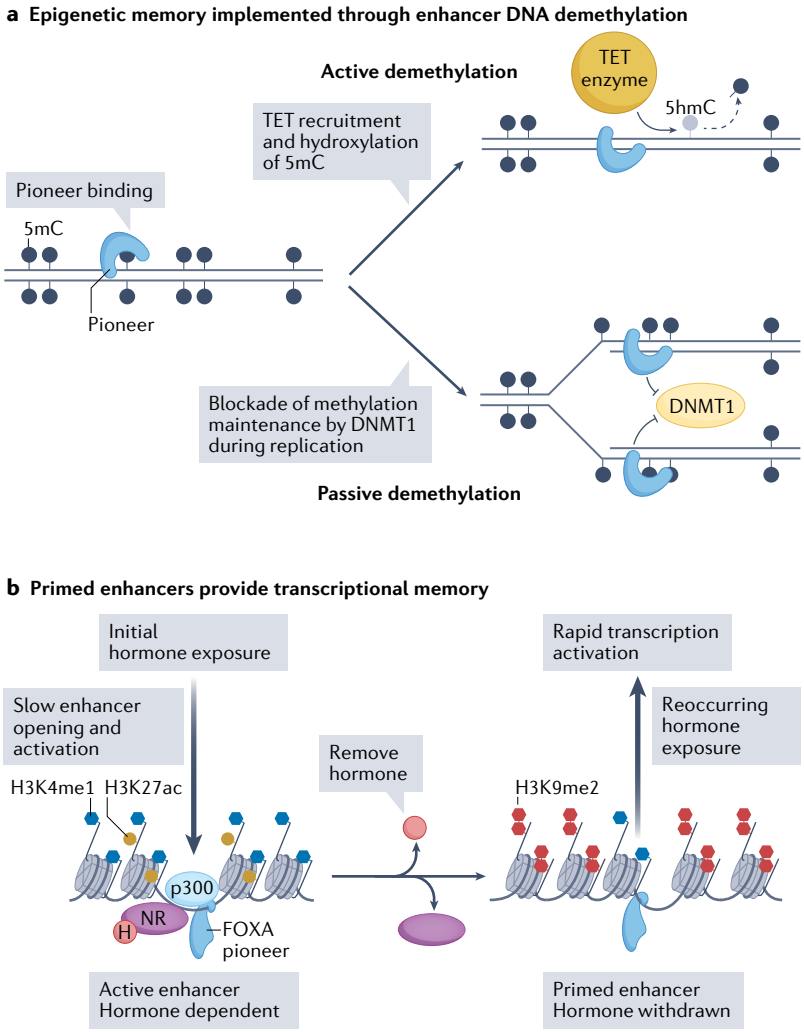

Fig. 3 | **Primed enhancers as a mechanism of epigenetic and transcriptional memories. a** | Epigenetic memory is implemented through pioneer-dependent enhancer DNA demethylation. As most genomic DNA, inactive enhancers in mammals have heavily methylated DNA. Upon pioneer activation, enhancer DNA becomes hypomethylated, and maintenance of this epigenetic state ensures long-term memory of active-chromatin organization at these enhancers. Enhancer DNA demethylation is achieved either through an active mechanism that relies on pioneer-dependent recruitment of the TET methylcytosine dioxygenases, which oxidize 5-methylcytosines (5mC) to 5-hydroxymethylcytosine (5hmC), thereby leading to demethylation (top), or through a passive mechanism that involves inhibition of methylation maintenance by DNA (cytosine-5)-methyltransferase 1 (DNMT1) (bottom). This maintenance function involves methylation of hemimethylated CpG at newly replicated DNA, and its pioneer-dependent blockade results in loss of methylation through dilution at successive cell divisions. **b** | Transcription-induced epigenetic memory is maintained at primed enhancers. Primed enhancers need a cooperating transcription factor such as a hormone (H)-dependent nuclear receptor (NR) for complete activation. When this cooperating transcription factor is no longer active following hormone withdrawal, the enhancers lose the gene activity-associated mark of acetylated histone H3 Lys27 (H3K27ac) and revert to a status that is similar to that of primed enhancers. Such transcription-induced memory at primed enhancers can facilitate their transcriptional reactivation in response to signals. H3K4me1, monomethylated histone H3 Lys4; H3K9me2, dimethylated histone H3 Lys9.

nucleosomes are reintroduced following replication, whereas at active chromatin, histone modifications associated with gene activity are introduced de novo[70]. In addition to reintroduction of nucleosomes with gene-repressing modifications (such as H3K9me2 and H3K9me3) following replication, enzymes responsible for the introduction and maintenance of H3K9me2 and H3K9me3 are present and required to maintain and expand this chromatin state. Hence, proteins that define active versus inactive chromatin states include enzymes that maintain the epigenetic state. Genomic DNA is also marked for this binary distinction of active versus inactive through DNA methylation, at least in species that have DNA CpG methylation. The repressive effect of DNA methylation on promoters was shown in the previous century by direct DNA methylation in vitro[71]. Many promoters are associated with G+C-rich regions known as 'CpG islands', and gene expression inversely correlates with the level of cytosine methylation at these regulatory sequences[72]. The idea that promoter methylation marks genes for silencing in a given cell is supported by comparing cells of very different lineages. However, more recent analyses of the relationship between DNA methylation and gene expression indicate that for cells of closely related lineages, such as different endocrine cells of the pancreas[73] or pituitary gland[7], it is not promoter access and methylation status that are crucial for expression but rather the methylation status and activity of cell-specific enhancers. Accordingly, tissue-specific enhancer demethylation accompanies gene induction and cell differentiation[74,75]. Thus, the enhancer subsets that define cell identity are hypomethylated in corresponding differentiated cells, whereas they are hypermethylated in related but different cells.

Pioneers establish epigenetic memory

The activation of enhancer repertoires that drive cell identity is dependent on pioneers that initiate chromatin opening at these enhancers. DNA demethylation through pioneer function at enhancers is a pioneer property that provides a basis for epigenetic memory at the DNA level; methylation dependency on pioneers was shown for FOXA and GATA4 (REFS[6,76]), PAX7 (REF.[7]), C/EBPα, KLF4 and TFCP2L1 (REF.[77]), EBF1 (REF.[9]) and neurogenic differentiation factor 2 (NEUROD2)[78]. Since DNA methylation status is maintained during replication, enhancer hypomethylation should remain following removal of the pioneer, and enhancer accessibility should be maintained (possibly in the primed state). This relationship was shown for the non-pioneers TPIT and the signal-inducible transcription factor STAT3 in cells where removal of PAX7 results in decreased DNA accessibility but maintained access to the non-pioneers TPIT and STAT3 (REF.[7]). Thus, the one-shot action of pioneers sets the stage for long-term implementation of new enhancer repertoires through enhancer DNA demethylation[75].

Demethylation of enhancer DNA by pioneers may involve an active mechanism based on recruitment of DNA demethylases of the TET family[79], as shown for FOXA[76,80,81], C/EBPα, KLF4 and TFCP2L1 (REF.[77]) and NEUROD2 (REF.[78]) (FIG. 3a). Alternatively, pioneers may cause enhancer DNA demethylation through a passive mechanism[82] of interference with methylation maintenance by DNA (cytosine-5)-methyltransferase 1 (DNMT1) and its regulatory protein UHRF1 (REF.[82]). Blockade of DNA methylation by DNMT1–UHRF1

is indeed the mechanism underlying global DNA demethylation that occurs in preimplantation embryos[83]. The DNMT1–UHRF system operates during replication to maintain CpG methylation: UHRF1 recognizes newly synthesized, hemimethylated DNA and recruits DNMT1 to methylate cytosines on the newly synthesized DNA strand[84]. This mechanism would therefore require passage through replication to establish new patterns of enhancer methylation. FOXA-induced DNA demethylation appears dependent on DNA replication in cultured cells[6], but liver FOXA-dependent enhancer demethylation requires the TET enzymes[76]. This apparent discrepancy between the passive and active mechanisms of demethylation will require further investigation to clarify the contributions of the two pathways.

Whereas the molecular basis of epigenetic memory based on DNA demethylation is broadly supported in mammals, as discussed above, the nature of epigenetic memory in *D. melanogaster*, which does not have significant cytosine methylation, is intriguing. Nonetheless, epigenetic or transcriptional memory may exist in fruit flies through the activity of the pioneer Zelda[85]. This memory of the timing of transcription activation is dependent on the number of Zelda-binding sites and the concentration of Zelda, but it is not dependent on mitotic bookmarking (BOX 2) or the presence of Zelda at mitosis[85]. The underlying mechanism remains unclear.

Transcription-induced epigenetic memory

A more specific form of epigenetic memory is observed at enhancers that are opened and activated through pioneer action and signal-dependent transcription factors, following the withdrawal of the signals. Examples of such transcription-induced epigenetic memory are found in innate immunity cells, in memory T cells and in B cells, and at nuclear receptor-dependent enhancers (FIG. 3b). For example, in T cells, chromatin accessibility at certain enhancers is greatly increased following their initial stimulation, is maintained at lower levels thereafter and is rapidly increased upon restimulation[86]. Such a memory function is found also in different innate immunity cells, including cells of the myeloid lineage (reviewed in REF.[87]). Following primary activation of innate immunity, promoters marked by H3K4me3 and enhancers marked by H3K4me1 retain some level of those modifications. Similarly to primed enhancers, the basal memory state of enhancers appears to be associated with single peaks and low levels of H3K4me1, whereas the active state of those enhancers correlates with acquisition of acetylated H3K27. The requirements for maintenance of an enhancer memory state (low level of H3K4me1 and low accessibility) following the removal of the activation signals remains vague.

It is unclear whether maintenance of the enhancer memory status requires the presence of DNA-bound transcription factors, although it was suggested (for example, in memory T cells) that the activity of ETS1 is required for maintenance of accessibility[86]. In these studies, it is not always feasible to separate the contribution of cooperating transcription factors from the contribution of pioneers, particularly with regard to chromatin accessibility data, so it is often impossible to exclude the possibility of enhancer occupancy in the memory state by an undocumented DNA-binding factor. For example, at enhancers activated by nuclear receptors, removal of the ligand (and therefore of the nuclear receptor) leads to decreased chromatin accessibility and to an enhancer state that resembles the primed state (FIG. 3b). In many instances, these enhancers are still occupied by the pioneer FOXA; however, the continued presence of FOXA does not appear to be required for maintenance of DNA hypomethylation at these enhancers[76].

As enhancer DNA demethylation is slower than primary enhancer activation and is a stable epigenetic modification, it is reasonable to associate the enhancer memory state with CpG hypomethylation within these enhancers[6,7,75,76,88,89]. Whereas enhancer DNA hypomethylation records the regulatory history of cells, it appears, at least in intestinal and blood cells, that mostly late-development enhancers (fetal rather than embryonic enhancers) retain a primed state through chromatin modifications; by contrast, the embryonic enhancers remain hypomethylated without primed-state chromatin modifications, a status that depends on the presence of Polycomb repressive complex 2 (REF.[75]). Thus, the idea that the primed state has an enhancer memory function is appealing. It is, however, not yet clear whether the primed state observed following de novo priming (FIG. 1) is the same as the memory or quiescent state of enhancers following withdrawal of an activating transcription factor (FIG. 3b): they are similar with regard to chromatin accessibility and H3K4me1 levels, but the former undergo mild CpG demethylation, whereas the latter have low CpG methylation[7,87]. It remains to be established whether and how the memory or quiescent enhancer status is linked to the extent of DNA demethylation.

Box 2 | **Mitotic bookmarking**

Some transcription factors, many of them pioneers, exhibit the property of mitotic bookmarking, meaning they are able to remain associated with their DNA targets during mitosis, whereas most transcription factors are ejected as the chromosomes condense. Mitotic bookmarking is thought to be a mechanism to re-establish the presence of transcription regulatory complexes following replication and mitosis[169]. Some bookmarking factors remain associated with a subset of their interphase targets, but others also have different recruitment sites at mitosis as shown for GATA1 (REF.[170]) and FOXA1 (REF.[171]). A large number of mitotic recruitment sites that do not correspond with interphase sites do not exhibit DNA sequence motifs cognate with the bookmarking factor. Interestingly, the pluripotency factors SOX2 and OCT4 also have bookmarking activity[172] that is not sequence specific[172], whereas the nuclear receptor bookmarking factor oestrogen-related receptor-β (ESRRB)[169] mostly retains sequence-specific interactions at mitosis[169]. The role of the broad nonspecific bookmarking by factors such as SOX2 is still not very clear. For the bookmarking factor ESRRB, mitotic binding is associated with maintenance of nucleosome position, which may be a mechanism to ensure re-establishment of transcriptional competence following mitosis[173]. Bookmarking by the transcription factor BRN2, which cooperates with the pioneer ASCL1 towards neural specification, mostly occurs through electrostatic interactions at sites that are different from its interphase sites, and these interactions promote rapid reinitiation of gene expression[174] following cell division[174]. It thus remains to be seen whether the bookmarking activity exhibited by pioneers at subsets of their sequence-specific genomic sites represents a continuation of their original chromatin opening activity or whether, as now appears more likely, this bookmarking activity is altogether different.

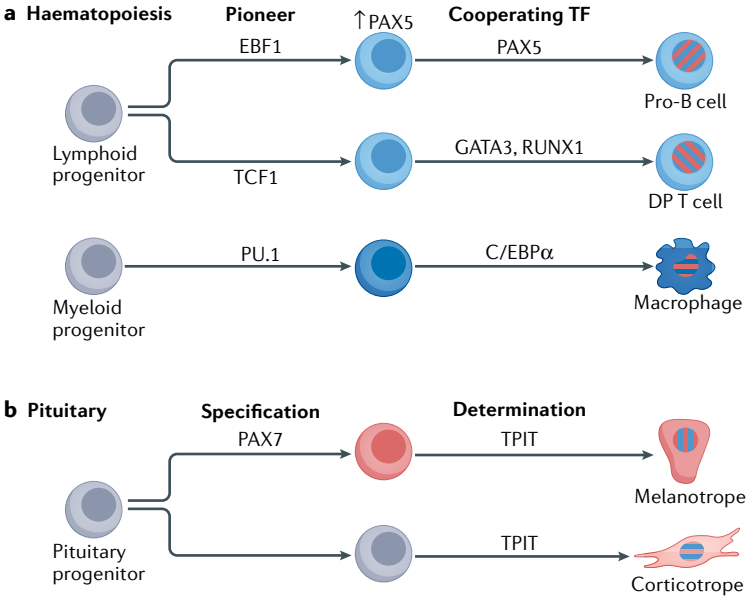

Fig. 4 | **Pioneer and non-pioneer transcription factors cooperate for specification and determination of cell fates. a** | Haematopoietic stem cells give rise to lymphoid and myeloid progenitors, which are further specified towards alternative cell fates through the activity of different pioneers. The pioneer EBF1 specifies the progenitor cells towards the B cell lineages in cooperation with PAX5, which is itself induced by EBF1, whereas TCF1 (also known as TCF7) specifies the T cell lineages in cooperation with GATA3 or RUNX1. The macrophage lineage is specified by PU.1 in cooperation with CCAAT/enhancer-binding protein-α (C/EBPα). **b** | In the pituitary gland, two lineages that express the same hormone precursor, pro-opiomelanocortin (POMC), arise in developmentally distinct tissues, the intermediate pituitary lobe and the anterior pituitary lobe; POMC-expressing cells in the two lobes process POMC into different hormones (not shown). The same transcription factor (TF), TPIT, determines terminal cell differentiation in both lineages, but prior expression of the pioneer PAX7 specifies the intermediate pituitary fate through opening of a unique repertoire of enhancers, which control melanotrope-specific gene expression. DP, double positive.

Pioneer–RNA polymerase II interactions

In addition to these enhancer-dependent mechanisms of epigenetic memory, there are examples of a specific type of epigenetic memory, which is implemented at promoters and is associated with increased levels of H3K4me3 and/or H3K4me2. In yeast, where most regulation occurs in the promoter-proximal region, transcriptional memory (for example, as observed after inositol starvation) is associated with increased H3K4me2 levels at target promoters[90] through occupancy by the transcription factor Sfl1. The priming of promoters by increased H3K4me3 levels was associated with enhanced recruitment of promoter-proximal RNA polymerase II (Pol II)[90], which would thus be ready for initiation of transcription upon activation.

Some pioneers may have evolved a specialized function that fulfils similar promoter priming through their direct interaction with Pol II, and this ability is associated with enhanced transcription activation[91]. The vast majority of pioneered loci are intergenic or intronic enhancers; only a few promoters were documented to be pioneer dependent such as the *Oacyl* promoter, which is opened by PAX7 (REF.[92]). However, recent evidence suggests that some pioneers may have a unique role in transcription regulation through association with Pol II. Indeed, this was first shown for the *Caenorhabditis elegans* FOXA homologue PHA-4, which binds promoters and recruits Pol II[93]. PHA-4 appears required for Pol II pausing in early development and hence for controlled release of Pol II later for synchronized initiation of gene expression.

It was recently reported that in mammalian cells, FOXA3 has a unique ability to interact with Pol II, in contrast to the related FOXA1 and FOXA2 (REF.[91]). The three FOXA factors bind similar subsets of pioneered enhancers[91], but only FOXA1 and FOXA2 are required for initiation of liver development[94]. By the reprogramming of mouse embryonic fibroblasts into hepatocyte-like cells, FOXA3 was shown to translocate from distal regulatory elements to the transcription start site of liver-specific genes. FOXA3 directly interacts with Pol II, and its recruitment to the transcription start site is required for significant target-gene expression and hepatocyte reprogramming. Upon transcription activation, FOXA3 appears to remain associated with Pol II: *FOXA3* mutants that fail to interact with Pol II also fail to significantly activate target-gene transcription. This crucial role of FOXA3 in the establishment of high-level, cell-specific gene expression is unique so far: it remains to be seen whether other pioneers may share similar activity. The association of FOXA3 with Pol II may contribute to establishment of transcriptional memory, but this possibility remains to be assessed.

Cell fate specification by pioneers

The primary role of pioneer factors is to implement cell fate decisions during development. By opening previously inaccessible regulatory elements, pioneers enable combinations of transcription factors to drive cell fate. In this developmental context of their function, pioneers exert activities classically associated with selector genes, which are responsible for specification of tissue identity, whereas their cooperating transcription factors often represent determination factors, which establish cell-type identity (FIG. 4). TABLE 1 lists pioneers and their cooperating transcription factors involved in cell fate decisions. In this section, we provide a discussion of cell fate specification by pioneers that is centred on their activity; for further information, we direct the reader to recent reviews on pioneers and cell fate specification[3] and on pioneers and cell-type reprogramming[16,17].

Activation of zygotic gene expression

The earliest developmental decision is activation of the zygote gene expression programme. This process relies on maternally expressed mRNAs for the production of factors that activate zygotic genes: these factors are pioneer factors. Two pioneers have been particularly studied in this context, Zelda in *D. melanogaster*[24] and double homeobox protein 4 (DUX4) in mammals[95]. Although Zelda is responsible for activation of the earliest zygotic genes, its activity is later complemented by another zinc-finger transcription factor pioneer, GAGA[96]. Later, at the cellularization step, another zinc-finger transcription factor, Odd paired (Opa) participates in driving the transition from pair-rule to segmental patterns of gene expression[97]; Opa appears to have pioneer-like properties[98]. There is no homologue of

Zelda in vertebrates, where other transcription factors appear to be involved in zygotic gene activation.

In mice and humans, the related pioneers DUX and DUX4 (REF.[95]), respectively, activate the cleavage-stage gene expression programme[99]. The function of DUX4 also involves directing the insertion of the histone variants H3.3X and H3.3Y in place of H3 at the body of targeted coding genes: this chromatin remodelling enhances transcription and may provide transcriptional memory[100]. Zygote gene activation appears to require the pioneers Oct4 (also known as Pou5f3), Sox and Nanog in *Xenopus tropicalis* and zebrafish[101–103]; these pioneers are the orthologues of the pluripotency reprograming factors described later.

Pioneers specify endoderm derivatives

The occupancy of liver-specific enhancers by FOXA in endoderm before activation of the liver gene-expression programme was the first evidence that this factor may act as a pioneer factor[22]. GATA4 is also expressed early in endoderm development, but its recruitment to many sites appears to depend on FOXA; hence, in these instances GATA4 may function as a cooperating transcription factor for FOXA function[5]. FOXA was also the first pioneer shown to bind nucleosomal DNA[104] and to initiate chromatin opening[105]. In *X. tropicalis*, the function of Foxa may be preceded by another pioneer, Foxh1, which already occupies regulatory sequences required for mesendoderm gene expression in the early blastula[106]. In mice, at least one of the related factors FOXA1 and FOXA2 is required for initiation of liver development[94]. Redundancy is also observed for two members of the GATA family, GATA4 and GATA6, which are required for liver development[107,108]. These factors also contribute to specification of other endoderm-derived tissues, for example, GATA4 for development of the ventral pancreas[109]. The model of cooperating factors acting together with FOXA2 for specification of different endoderm derivatives is supported by experiments relying on reprogramming of embryonic stem cells[68].

Cell differentiation into pancreatic tissues is also dependent on the *Foxa1* and *Foxa2* genes in mice[94], suggesting that either one or both of these factors are required for pancreatic differentiation in mice[94]. Further investigation using human pluripotent stem cells showed that FOXA2 is required for pancreatic differentiation in humans, whereas FOXA1 is not[56]. Pancreatic specifications also require GATA6, and both FOXA2 and GATA6 are essential for opening of pancreas-specific enhancers[56].

Haematopoietic cell lineages

The specification of haematopoietic cell lineages is a complex process that involves several transcription factors, with many of the factors being involved in differentiation of different lineages (TABLE 1). The sequential specification and restriction of cell identity leads to the establishment of differentiated cells of the myeloid and lymphoid lineages. A subset of the transcription factors responsible has pioneer activity with strong phenotypes, both in development and in cell-type reprogramming schemes (FIG. 4a). For example, PU.1 drives haematopoietic progenitors towards the myeloid and macrophage fates[10] and it acts in cooperation with the pioneer C/EBPα[11]. These two pioneers can also reprogramme both pre-B cells and fibroblasts into macrophage-like cells[11,110].

For differentiation to the lymphoid lineages, implementation of the B cell programme requires the pioneer factor EBF1 (REFS[9,43]), and its inactivation leads to failure of B cell differentiation[111] (FIG. 4a). EBF1 is required for expression of the downstream transcription factor PAX5 (REFS[112,113]), which cooperates with EBF1 in driving lineage commitment and maintenance of the B cell identity. The combination of transcription factors required for differentiation and maintenance of B cells also includes the bHLH factor E2A and FOXO1 (also known as FKHR)[112,114].

For differentiation to the T cell lineage, the pioneer TCF1 (also known as TCF7) opens chromatin at specific enhancers, which are required to establish the lineage[115] (FIG. 4a). Other transcription factors involved in establishment of the T cell lineage with TCF1 include GATA3 and RUNX1 (REFS[116,117]).

The intermediate-pituitary fate

The pituitary intermediate lobe sustains direct contact between tissues derived from the surface ectoderm and neural ectoderm; this contact is essential for pituitary development. The intermediate lobe is a homogeneous tissue containing one hormone-secretory cell type, the melanotrope cell, which produces α-melanotropin (also known as melanocyte-stimulating hormone-α). Early in development, the melanotrope cell acquires its identity through expression of PAX7 (REF.[118]) (FIG. 4b). Subsequent terminal differentiation of two pituitary lineages that express the same hormone precursor, pro-opiomelanocortin, is achieved by the same transcription factor, TPIT[119–121]. PAX7 modulates the function of TPIT in at least two ways: first, it opens chromatin at a few thousand enhancers that direct the melanotrope fate, and second, it represses a subset of genes that constitute the alternative corticotrope cell fate[7,31,118]. In this context, TPIT acts as a cooperating factor with PAX7 to implement the melanotrope cell fate[31]. *PAX7* is thus a selector gene that generally specifies the tissue identity of the intermediate lobe and acts as a binary switch to specify alternative cell fates, whereas TPIT determines and implements melanotrope cell differentiation (FIG. 4b).

The cardiac fate

Establishment of the cardiac fate depends on many transcription factors acting coordinately[122], including pioneers such as GATA4 (REFS[123,124]). A number of transcription factors that are crucial for cardiac-fate determination have been associated with gains in chromatin accessibility. However, it is not always clear whether these gains of accessibility reflect true pioneer ability or whether they may just represent the activity of cooperating factors. To demonstrate pioneer ability, recruitment to completely compact, inaccessible chromatin and the appearance of accessibility must be shown, which is not always easy. A good case in point is the purported pioneer activity of ISL1 for specification of the cardiac fate[49].

Corticotrope cell
Endocrine cell of the pituitary anterior lobe that secretes the pro-opiomelanocortin-derived hormone adrenocorticotropin.

Cell replacement therapies
Therapies designed to replace deficient cells by competent cells, such as pancreatic β-cells in diabetes or dopaminergic neurons in Parkinson disease. This replacement could be achieved by transplantation of reprogrammed cells or by reprograming cells in vivo.

Although the biological importance of ISL1 for cardiac development and function is clear, the reported gains of chromatin accessibility related with ISL1 occupancy suggest a further gain of accessibility from pre-existing accessibility rather than accessibility gain from an inaccessible state. The hypothesis that ISL1 possess pioneer activity is, however, partly supported by nucleosome binding experiments[49].

Cell-type reprogramming by pioneers

Cell reprogramming directed by pioneer factors is often determined by gain-of-function experiments that both confirm the properties of pioneers in cell fate specification and at the same time constitute tremendous tools for developing novel cell replacement therapies. TABLE 2 provides examples of reprogramming activities for a selection of transcription factor combinations.

Induced stem cell pluripotency

The reprogramming experiment performed with *Xenopus laevis* nuclei by John Gurdon in 1962 was so striking that its result remained suspicious to many for a long time[125]. Indeed, the successful replacement of an oocyte nucleus with a skin cell nucleus seemed unbelievable. But the discovery by Shinya Yamanaka and colleagues of a cocktail of transcription factors that can reprogramme fibroblasts into induced pluripotent stem cells[126,127] dramatically demonstrated the power of a few pioneers to enable this reprogramming through epigenome remodelling[34]. Reprogramming into induced pluripotent stem cells is a slow and inefficient process that involves multiple steps, beginning with the initial actions of the pioneers SOX2 and OCT4 at specific enhancers and followed by extensive chromatin remodelling extending far from the initial sites of chromatin opening[32]. As for other pioneers (TABLE 2), the pluripotency pioneer factors SOX2, OCT4 and KLF4 cooperate with each other in different combinations at specific subsets of enhancers[128]. This process also involves the cooperating factor MYC[32].

The neural and myogenic cell fates

The pioneer ability of transcription factors that control neurogenesis is supported mostly by gain-of-function experiments, namely by their reprogramming capacity (TABLE 2). Three neurogenic bHLH pioneer factors,

Table 2 | **Cell reprogramming by pioneer factors and cooperating transcription factors**

Reprogamming into	Cell of origin	Pioneer factor	Cooperating transcription factor	Refs
Pluripotent cells				
iPS cell	Fibroblast (m)	KLF4, SOX2, OCT4	MYC	127
Endoderm derivatives				
Hepatocyte	Fibroblast (m)	FOXA3, GATA4	HNF1α	54
Hepatocyte	Fibroblast (m)	FOXA	HNF4α	55
Bipotent hepatic progenitor	Fibroblast (m)	FOXA3	HNF1β	159
Neuronal lineages				
Glutamatergic neuron	Fibroblast (m)	ASCL1	BRN2, MYT1L	131
Dopaminergic neuron	Fibroblast (h)	ASCL1, FOXA2	BRN2, MYT1L, LMX1A	160
Motor neuron	Fibroblast (h, m)	ASCL1	BRN2, MYT1L, LHX3, HB9, ISL1, NGN2	161
Neuron	Hepatocyte (m)	ASCL1	BRN2, MYT1L	162
Tripotent neural progenitor	Fibroblast (m)	SOX2	FOXG1, BRN2	163
Haematopoietic lineages				
Haematopoietic progenitor	Fibroblast (h)	OCT4	–	164
Haematopoietic progenitor	Fibroblast (m)	GATA2	GFI1B, FOS, ETV6	165
Macrophage	Fibroblast (m)	PU.1	C/EBP	110
Other lineages				
Muscle	Fibroblast (m)	MYOD	–	133,134
Cardiomyocyte	Fibroblast (m)	GATA4	MEF2C, TBX5	166
Cardiomyocyte	Cardiac progenitor (m)	ISL1	GATA4	49
Osteoblast	Fibroblast (h)	OCT4	LMYC, RUNX2, OSX	167
Melanotrope	Corticotrope (m)	PAX7	TPIT	7,118

Selected examples of pioneer and cooperating transcription factor combinations used to reprogramme cells towards different fates with emphasis on diversity of outcomes. This list is far from exhaustive; recent reviews provide comprehensive data[16,17]. ASCL1, Achaete–Scute homologue 1; BRN2, brain-specific homeobox/POU domain protein 2; C/EBP, CCAAT/enhancer-binding protein; FOXA, forkhead box protein A; FOXG1, forkhead box protein G1; GFI1B, zinc-finger protein GFI1B; h, human; HNF, hepatocyte nuclear factor; iPS cell, induced pluripotent stem cell; KLF4, Krüppel-like factor 4; LHX3, LIM homeobox protein 3; LMX1A, LIM homeobox transcription factor 1α; m, mouse; MEF2C, myocyte-specific enhancer factor 2C; MYT1L, myelin transcription factor 1-like protein; NGN2, neurogenin 2; OCT4, octamer-binding protein 4; OSX, zinc-finger protein osterix; RUNX2, Runt-related transcription factor 2; TBX5, T-box transcription factor 5; TPIT, T-box transcription factor, pituitary.

ASCL1 (also known as MASH1), NEUROD1 and NEUROD2, have the ability to initiate neurogenic differentiation. ASCL1 can initiate neurogenic differentiation in embryonic stem cells or in fibroblasts[129–131], NEUROD1 can initiate neurogenic differentiation in embryonic stem cells only[132] and NEUROD2 can initiate neurogenic differentiation in embryocarcinoma cells[78]. The pioneer ASCL1 is essential for reprogramming fibroblasts into neuronal cells[131], and the cooperating transcription factor BRN2 (also known as POU3F2) contributes to this process[129]. The pioneer ability of NEUROD2 was correlated with changes in DNA methylation at neuronal enhancers during normal brain development; the DNA demethylase TET2 acts at NEUROD2-targeted enhancers, suggesting that enhancer demethylation is actively initiated through recruitment of TET2 (REF.[78]).

The first bHLH pioneer transcription factor shown to reprogramme fibroblasts was MYOD, which directs myogenic differentiation[133,134]. This reprogramming activity was directly compared with that of ASCL1, and although both factors are initially recruited to overlapping subsets of chromatin sites in fibroblasts, their unique reprogramming function relies on the recruitment of specific cooperating factors. These cooperating factors include transcriptional repressors such as myelin transcription factor 1-like protein (MYT1L), which contributes to reprogramming into neurons through repression of myogenic genes[135]. It is presently unknown whether ASCL1 requires a cooperating factor for implementation of the neurogenic programme. Thus, pioneer-dependent reprogramming depends on strong recruitment to closed chromatin and usually requires cooperating factors, which explains differences in reprogramming ability.

The bHLH pioneers ASCL1 and NEUROD1 are also expressed in a variety of non-neuronal tissues (for example, in the developing pituitary and pancreas), but their potent reprogramming activity is essentially neuronal. Why are ASCL1 and NEUROD1 unable to exert their neuronal reprogramming ability in the pituitary or pancreas: is it because of the activity of a repressor similar to MYT1L, or could the pituitary and pancreas heterochromatin environment be incompatible with their function? The answer to this question will likely tell us much about the mechanisms of pioneer function.

Reprograming to endoderm derivatives

Reprogramming towards the hepatic fate is dependent on the FOXA pioneers and it is achieved with various combinations of the pioneer GATA4 with the transcription factor HNF4α or HNF1α, for example, FOXA3 with GATA4 and HNF1α[54] or FOXA1, FOXA2 or FOXA3 with HNF4α[55] (TABLE 2). The FOXA family of pioneers illustrates the redundancy between family members for reprogramming, as any of FOXA1, FOXA2 or FOXA3 is suitable for reprogramming even though FOXA3 has properties different from those of the other two and is not sufficient in vivo for liver specification[91,94]. Thus, partial functional redundancy between family members may fulfil the requirements for reprogramming. In addition, positive cross-regulation between pioneers and transcription factors involved in reprogramming ensures that once a fate is engaged, it is reinforced and maintained.

FOXA2 also appears to be the major regulator of pancreatic cell identity[136] with regard to pancreatic β-cells[136]. The emphasis of research efforts has been mostly to reprogramme exocrine cells towards the β-cell fate for therapeutic purposes. For this purpose, the combination of neurogenin 3, pancreas/duodenum homeobox protein 1 and MAFA is effective, but it is not yet clear whether these factors exert pioneer activity[137].

Pioneer function in cancer development

By altering the fate of cancer cells, pioneers may initiate, promote and/or alter tumorigenesis. Both aberrant (over)expression of pioneers and the production of chimeric proteins resulting from chromosomal translocations involving pioneer-encoding genes could give rise to novel reprogramming activities that may contribute to tumorigenesis. As a primer for entire gene regulatory networks, pioneer expression may be paramount for cell phenotypes such as hormone responsiveness. However, in other instances, the ability of pioneers to enhance cell differentiation phenotypes provides therapeutic opportunities, as discussed below for ASCL1.

The FOXA pioneers were first implicated in hormone-dependent cancers as it became clear that they pioneer recruitment of cooperating nuclear receptors[57–60]. This cooperation is essential to establish the hormone-dependent gene regulatory network that characterizes such cancers[138]. FOXA1 is required to recruit the oestrogen receptor and for oestrogen responsiveness of oestrogen receptor-positive breast cancer cells[139] and also to establish hormone responsiveness to androgens in prostate cancer[140]. Consistent with this role, *FOXA1* point mutations found in prostate cancers were associated with decreased androgen signalling and increased tumour growth[18,141].

Overexpression of FOXA1 is associated with poor prognosis in prostate cancer[142], and in primary prostate cancer it is often associated with non-coding mutations in *cis*-regulatory sequences[143]. FOXA1 overexpression is associated with a better prognosis in breast cancer[144]. Recent reviews have exhaustively discussed the roles of pioneers in prostate cancer[19]. In addition to FOXA, which is overexpressed in hormone-dependent cancers as discussed above, overexpression of FOXM1 is observed in subsets of breast, non-Hodgkin lymphoma and peripheral nerve sheath tumours[145–147]. However, it is not clear whether this overexpression involves pioneer-type activity. A clearer case of pioneer overexpression is the association of HOXA9 with leukaemia through activation of leukaemia-associated enhancers[20]. This activity was associated with recruitment of the MLL3–MLL4 histone methyltransferase complex.

Interestingly, some rhabdomyosarcomas harbour chromosomal translocations that lead to the production of chimeric proteins containing the amino terminus of PAX3 or PAX7 fused to the carboxy-terminal transactivation domain of FOXO1. Both PAX7 and FOXO1 have pioneer activity[118,148] and the PAX3–FOXO1 chimaera activates regulatory networks involved in

tumour development[21]. The related *FOXO3* and *FOXO4* genes undergo translocation with the MLL gene in acute lymphoblastic leukaemia[149]. It is also noteworthy that ectopic expression of the homeodomain pioneer DUX4 in muscle causes facioscapulohumeral muscular dystrophy[150,151] and that its activity may compete with PAX7 for binding target sites[152].

An interesting therapeutic application of pioneer expression in cancer was reported in glioblastomas expressing high levels of the pioneer ASCL1, in which treatment with Notch inhibitors resulted in even higher ASCL1 expression levels, enhanced neuronal differentiation and reduction of stem properties and of tumorigenecity[130]. In this case, activation of the pioneer function provides a therapeutic opportunity by enhancing differentiation of poorly differentiated tumour cells.

Future perspective

Pioneer factors provide great therapeutic opportunities considering their powerful role as master regulators of the epigenome. Their ability to initiate cell reprogramming provides a realistic approach for cell replacement therapies or for reversing the phenotypes of cancer cells, as illustrated by ASCL1 in glioblastomas[130]. But to use these tools soundly, we must understand the molecular basis of pioneer activity. The early steps of pioneer function remain the least understood. We have only a vague understanding of the molecular basis of pioneer access to closed chromatin: beyond knowing that constitutive heterochromatin is a barrier to many pioneers and that pioneers interact with nucleosomal DNA, are there other heterochromatin barriers or permissive features that constitute a code for pioneer access? Could different pioneers have differential access to different types of chromatin? Differential chromatin access could thus be an important component of the sequential action of pioneers in differentiation cascades. In other words, is there more to chromatin structure than simply closed and open?

Once a pioneer is stably recruited to closed chromatin, what triggers the cascade leading to chromatin opening? This initial step remains unknown. If there is an initiating mechanism, are the following steps merely its consequence, or is the process subject to checkpoints and is it reversible? These questions are relevant to the nature of the epigenetic memory that is established by pioneers. Do primed enhancers represent this epigenetic memory? What are the requirements for long-term maintenance of the primed-enhancer memory status? These basic questions are important since they may be relevant to understanding the molecular basis of memory in many biological systems, such as in immunity. Clearly, as we better understand the mechanisms of pioneer function and recognize their contribution to development, we can exploit the power of pioneers for diagnosis and treatment of diseases from cancer to degenerative disorders.

Published online 9 March 2022

1. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development* **143**, 1833–1837 (2016).
2. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).
3. Larson, E. D., Marsh, A. J. & Harrison, M. M. Pioneering the developmental frontier. *Mol. Cell* **81**, 1640–1650 (2021).
4. Stewart-Morgan, K. R., Petryk, N. & Groth, A. Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* **22**, 361–371 (2020).
5. 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).
6. Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).
7. Mayran, A. et al. Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat. Genet.* **50**, 259–269 (2018).
8. Barozzi, I. et al. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol. Cell* **54**, 844–857 (2014).
9. Boller, S. et al. Pioneering activity of the C-terminal domain of EBF1 shapes the chromatin landscape for B cell programming. *Immunity* **44**, 527–541 (2016).
10. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
11. van Oevelen, C. et al. C/EBPalpha activates pre-existing and de novo macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. *Stem Cell Rep.* **5**, 232–247 (2015).
12. Iwafuchi-Doi, M. et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).
13. Cirillo, L. A. & Zaret, K. S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol. Cell* **4**, 961–969 (1999).
14. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
15. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
16. Aydin, B. & Mazzoni, E. O. Cell reprogramming: the many roads to success. *Annu. Rev. Cell Dev. Biol.* **35**, 433–452 (2019).
17. Horisawa, K. & Suzuki, A. Direct cell-fate conversion of somatic cells: toward regenerative medicine and industries. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **96**, 131–158 (2020).
18. Adams, E. J. et al. FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. *Nature* **571**, 408–412 (2019).
19. Hankey, W., Chen, Z. & Wang, Q. Shaping chromatin states in prostate cancer by pioneer transcription factors. *Cancer Res.* **80**, 2427–2436 (2020).
20. Sun, Y. et al. HOXA9 reprograms the enhancer landscape to promote leukemogenesis. *Cancer Cell* **34**, 643–658.e5 (2018).
21. Cao, L. et al. Genome-wide identification of PAX3-FKHR binding sites in rhabdomyosarcoma reveals candidate target genes important for development and cancer. *Cancer Res.* **70**, 6497–6508 (2010).
22. Gualdi, R. et al. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev.* **10**, 1670–1682 (1996).
23. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).
24. McDaniel, S. L. et al. Continued activity of the pioneer factor Zelda is required to drive zygotic genome activation. *Mol. Cell* **74**, 185–195.e4 (2019).
25. Zhu, F. et al. The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
26. 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).
27. Michael, A. K. et al. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* **368**, 1460–1465 (2020).
28. Tanaka, H. et al. Interaction of the pioneer transcription factor GATA3 with nucleosomes. *Nat. Commun.* **11**, 4136 (2020).
29. Fernandez Garcia, M. et al. Structural features of transcription factors associating with nucleosome binding. *Mol. Cell* **75**, 921–932.e6 (2019).
30. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
31. Mayran, A. et al. Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. *Nat. Commun.* **10**, 3807 (2019).
32. 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).
33. Nicetto, D. et al. H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science* **363**, 294–297 (2019).
34. Koche, R. P. et al. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8**, 96–105 (2011).
35. Becker, J. S. et al. Genomic and proteomic resolution of heterochromatin and its restriction of alternate fate genes. *Mol. Cell* **68**, 1023–1037.e15 (2017).
36. Hildebrand, E. M. & Dekker, J. Mechanisms and functions of chromosome compartmentalization. *Trends Biochem. Sci.* **45**, 385–396 (2020).
37. Poleshko, A. et al. Genome-nuclear lamina interactions regulate cardiac stem cell lineage restriction. *Cell* **171**, 573–587.e14 (2017).
38. Stadhouders, R. et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* **50**, 238–249 (2018).
39. Bolt, C. C. et al. Context-dependent enhancer function revealed by targeted inter-TAD relocation. Preprint at *bioRxiv* https://doi.org/10.1101/2022.01.19.476888 (2022).
40. Larson, A. G. et al. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
41. Giresi, P. G. & Lieb, J. D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (formaldehyde assisted isolation of regulatory elements). *Methods* **48**, 233–239 (2009).
42. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open

chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
43. Li, R. et al. Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. *Genes Dev.* **32**, 96–111 (2018).
44. Wapinski, O. L. et al. Rapid chromatin switch in the direct reprogramming of fibroblasts to neurons. *Cell Rep.* **20**, 3236–3247 (2017).
45. Jozwik, K. M., Chernukhin, I., Serandour, A. A., Nagarajan, S. & Carroll, J. S. FOXA1 directs H3K4 monomethylation at enhancers via recruitment of the methyltransferase MLL3. *Cell Rep.* **17**, 2715–2723 (2016).
46. King, H. A., Trotter, K. W. & Archer, T. K. Chromatin remodeling during glucocorticoid receptor regulated transactivation. *Biochim. Biophys. Acta* **1819**, 716–726 (2012).
47. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife* **6**, e22631 (2017).
48. Takaku, M. et al. GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol.* **17**, 36 (2016).
49. Gao, R. et al. Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate. *Cell Res.* **29**, 486–501 (2019).
50. Hoffman, J. A., Trotter, K. W., Ward, J. M. & Archer, T. K. BRG1 governs glucocorticoid receptor interactions with chromatin and pioneer factors across the genome. *eLife* **7**, e35073 (2018).
51. Dyson, H. J. & Wright, P. E. Role of intrinsic protein disorder in the function and interactions of the transcriptional coactivators CREB-binding protein (CBP) and p300. *J. Biol. Chem.* **291**, 6714–6722 (2016).
52. Alver, B. H. et al. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* **8**, 14648 (2017).
53. Jindal, G. A. & Farley, E. K. Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev. Cell* **56**, 575–587 (2021).
54. Huang, P. et al. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386–389 (2011).
55. Sekiya, S. & Suzuki, A. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* **475**, 390–393 (2011).
56. Lee, K. et al. FOXA2 is required for enhancer priming during pancreatic differentiation. *Cell Rep.* **28**, 382–393.e7 (2019).
57. Laganiere, J. et al. Location analysis of estrogen receptor α target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc. Natl Acad. Sci. USA* **102**, 11651–11656 (2005).
58. Carroll, J. S. et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
59. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
60. Wang, Q. et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245–256 (2009).
61. Voss, T. C. et al. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544–554 (2011).
62. Sherwood, R. I. et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
63. Swinstead, E. E. et al. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165**, 593–605 (2016).
64. Glont, S. E., Chernukhin, I. & Carroll, J. S. Comprehensive genomic analysis reveals that the pioneering function of FOXA1 is independent of hormonal signaling. *Cell Rep.* **26**, 2558–2565.e3 (2019).
65. Paakinaho, V., Swinstead, E. E., Presman, D. M., Grøntved, L. & Hager, G. L. Meta-analysis of chromatin programming by steroid receptors. *Cell Rep.* **28**, 3523–3534.e2 (2019).
66. Nevil, M., Gibson, T. J., Bartolutti, C., Iyengar, A. & Harrison, M. M. Establishment of chromatin accessibility by the conserved transcription factor Grainy head is developmentally regulated. *Development* **147**, dev185009 (2020).
67. Jacobs, J. et al. The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat. Genet.* **50**, 1011–1020 (2018).
68. Cernilogar, F. M. et al. Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. *Nucleic Acids Res.* **47**, 9069–9086 (2019).
69. Ungerbäck, J. et al. Pioneering, chromatin remodeling, and epigenetic constraint in early T-cell gene regulation by SPI1 (PU.1). *Genome Res.* **28**, 1508–1519 (2018).
70. Escobar, T. M. et al. Active and repressed chromatin domains exhibit distinct nucleosome segregation during DNA replication. *Cell* **179**, 953–963.e11 (2019).
71. Busslinger, M., Hurst, J. & Flavell, R. A. DNA methylation and the regulation of globin gene expression. *Cell* **34**, 197–206 (1983).
72. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213 (1986).
73. Neiman, D. et al. Islet cells share promoter hypomethylation independently of expression, but exhibit cell-type-specific methylation in enhancers. *Proc. Natl Acad. Sci. USA* **114**, 13525–13530 (2017).
74. Orlanski, S. et al. Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. *Proc. Natl Acad. Sci. USA* **113**, 5018–5023 (2016).
75. Jadhav, U. et al. Extensive recovery of embryonic enhancer and gene memory stored in hypomethylated enhancer DNA. *Mol. Cell* **74**, 542–554.e5 (2019).
76. Reizel, Y. et al. FoxA-dependent demethylation of DNA initiates epigenetic memory of cellular identity. *Dev. Cell* **56**, 602–612.e4 (2021).
77. Sardina, J. L. et al. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell* **23**, 727–741.e9 (2018).
78. Hahn, M. A. et al. Reprogramming of DNA methylation at NEUROD2-bound sequences during cortical neuron differentiation. *Sci. Adv.* **5**, eaax0080 (2019).
79. Lio, C. J. et al. TET methylcytosine oxidases: new insights from a decade of research. *J. Biosci.* **45**, 21 (2020).
80. Yang, Y. A. et al. FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. *Nucleic Acids Res.* **44**, 8153–8164 (2016).
81. Zhang, Y. et al. Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. *Nat. Genet.* **48**, 1003–1013 (2016).
82. Vanzan, L. et al. High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its binding sites. *Nat. Commun.* **12**, 3337 (2021).
83. von Meyenn, F. et al. Comparative principles of DNA methylation reprogramming during human and mouse in vitro primordial germ cell specification. *Dev. Cell* **39**, 104–115 (2016).
84. Bronner, C., Alhosin, M., Hamiche, A. & Mousli, M. Coordinated dialogue between UHRF1 and DNMT1 to ensure faithful inheritance of methylated DNA patterns. *Genes* **10**, 65 (2019).
85. Dufourt, J. et al. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat. Commun.* **9**, 5194 (2018).
86. Tu, W. J. et al. Priming of transcriptional memory responses via the chromatin accessibility landscape in T cells. *Sci. Rep.* **7**, 44825 (2017).
87. Netea, M. G. et al. Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* **20**, 375–388 (2020).
88. Pacis, A. et al. Gene activation precedes DNA demethylation in response to infection in human dendritic cells. *Proc. Natl Acad. Sci. USA* **116**, 6938–6943 (2019).
89. Luo, C. et al. Global DNA methylation remodeling during direct reprogramming of fibroblasts to neurons. *eLife* **8**, e40197 (2019).
90. D'Urso, A. et al. Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. *eLife* **5**, e16691 (2016).
91. Horisawa, K. et al. The dynamics of transcriptional activation by hepatic reprogramming factors. *Mol. Cell* **79**, 660–676.e8 (2020).
92. Pelletier, A., Mayran, A., Omichinski, J. & Drouin, J. Pax7 pioneer action requires both paired and homeo DNA binding domains. *Nucleic Acids Res.* **49**, 7424–7436 (2021).
93. Hsu, H. T. et al. Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4. *Science* **348**, 1372–1376 (2015).
94. Lee, C. S., Friedman, J. R., Fulmer, J. T. & Kaestner, K. H. The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**, 944–947 (2005).
95. Choi, S. H. et al. DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
96. Gaskill, M. M., Gibson, T. J., Larson, E. D. & Harrison, M. M. GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. *eLife* **10**, e66668 (2021).
97. Koromila, T. et al. Odd-paired is a pioneer-like factor that coordinates with Zelda to control gene expression in embryos. *eLife* **9**, e59610 (2020).
98. Soluri, I. V., Zumerling, L. M., Payan Parra, O. A., Clark, E. G. & Blythe, S. A. Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the Drosophila segmentation network. *eLife* **9**, e53916 (2020).
99. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
100. Resnick, R. et al. DUX4-induced histone variants H3.X and H3.Y Mark DUX4 target genes for expression. *Cell Rep.* **29**, 1812–1820.e5 (2019).
101. Gentsch, G. E., Spruce, T., Owens, N. D. L. & Smith, J. C. Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nat. Commun.* **10**, 4269 (2019).
102. Pälfy, M., Schulze, G., Valen, E. & Vastenhouw, N. L. Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS Genet.* **16**, e1008546 (2020).
103. Veil, M., Yampolsky, L. Y., Grüning, B. & Onichtchouk, D. Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. *Genome Res.* **29**, 383–395 (2019).
104. Cirillo, L. A. et al. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* **17**, 244–254 (1998).
105. Serandour, A. A. et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* **21**, 555–565 (2011).
106. Charney, R. M. et al. Foxh1 occupies cis-regulatory modules prior to dynamic transcription factor interactions controlling the mesendoderm gene program. *Dev. Cell* **40**, 595–607.e4 (2017).
107. Holtzinger, A. & Evans, T. Gata4 regulates the formation of multiple organs. *Development* **132**, 4005–4014 (2005).
108. Zhao, R. et al. GATA6 is essential for embryonic development of the liver but dispensable for early heart formation. *Mol. Cell Biol.* **25**, 2622–2631 (2005).
109. Watt, A. J., Zhao, R., Li, J. & Duncan, S. A. Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Dev. Biol.* **7**, 37 (2007).
110. Feng, R. et al. PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105**, 6057–6062 (2008).
111. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263–267 (1995).
112. Lin, Y. C. et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* **11**, 635–643 (2010).
113. Johanson, T. M. et al. Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. *Nat. Immunol.* **19**, 1257–1264 (2018).
114. Mansson, R. et al. Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *Proc. Natl Acad. Sci. USA* **109**, 21028–21033 (2012).
115. Johnson, J. L. et al. Lineage-determining transcription factor TCF-1 initiates the epigenetic identity of T cells. *Immunity* **48**, 243–257.e10 (2018).
116. Germar, K. et al. T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. *Proc. Natl Acad. Sci. USA* **108**, 20060–20065 (2011).
117. Weber, B. N. et al. A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* **476**, 63–68 (2011).
118. Budry, L. et al. The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes Dev.* **26**, 2299–2310 (2012).

119. Lamolet, B. et al. A pituitary cell-restricted T-box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell* **104**, 849–859 (2001).
120. Pulichino, A. M. et al. Tpit determines alternate fates during pituitary cell differentiation. *Genes Dev.* **17**, 738–747 (2003).
121. Pulichino, A. M. et al. Human and mouse Tpit gene mutations cause early onset pituitary ACTH deficiency. *Genes Dev.* **17**, 711–716 (2003).
122. Luna-Zurita, L. et al. Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. *Cell* **164**, 999–1014 (2016).
123. He, A. et al. Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat. Commun.* **5**, 4907 (2014).
124. Jia, G. et al. Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nat. Commun.* **9**, 4877 (2018).
125. Gurdon, J. B. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* **10**, 622–640 (1962).
126. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
127. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
128. Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459.e20 (2017).
129. Wapinski, O. L. et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621–635 (2013).
130. Park, N. I. et al. ASCL1 reorganizes chromatin to direct neuronal fate and suppress tumorigenicity of glioblastoma stem cells. *Cell Stem Cell* **21**, 209–224.e7 (2017).
131. Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
132. Pataskar, A. et al. NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.* **35**, 24–45 (2016).
133. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
134. Tapscott, S. J. et al. MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science* **242**, 405–411 (1988).
135. Lee, Q. Y. et al. Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes. *Nat. Cell Biol.* **22**, 401–411 (2020).
136. Milan, M. et al. FOXA2 controls the cis-regulatory networks of pancreatic cancer cells in a differentiation grade-specific manner. *EMBO J.* **38**, e102161 (2019).
137. Zhou, Q. & Melton, D. A. Pancreas regeneration. *Nature* **557**, 351–358 (2018).
138. Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).
139. Carroll, J. S. et al. Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**, 1289–1297 (2006).
140. Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
141. Grasso, C. S. et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
142. Sahu, B. et al. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J.* **30**, 3962–3976 (2011).
143. Zhou, S. et al. Noncoding mutations target cis-regulatory elements of the FOXA1 plexus in prostate cancer. *Nat. Commun.* **11**, 441 (2020).
144. Thorat, M. A. et al. Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. *J. Clin. Pathol.* **61**, 327–332 (2008).
145. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
146. Green, M. R. et al. Integrative genomic profiling reveals conserved genetic mechanisms for tumorigenesis in common entities of non-Hodgkin's lymphoma. *Genes Chromosom. Cancer* **50**, 313–326 (2011).
147. Yu, J. et al. Array-based comparative genomic hybridization identifies CDK4 and FOXM1 alterations as independent predictors of survival in malignant peripheral nerve sheath tumor. *Clin. Cancer Res.* **17**, 1924–1934 (2011).
148. Hatta, M. & Cirillo, L. A. Chromatin opening and stable perturbation of core histone:DNA contacts by FoxO1. *J. Biol. Chem.* **282**, 35583–35593 (2007).
149. So, C. W. & Cleary, M. L. Common mechanism for oncogenic activation of MLL by forkhead family proteins. *Blood* **101**, 633–639 (2003).
150. Young, J. M. et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet.* **9**, e1003947 (2013).
151. Lim, K. R. Q., Nguyen, Q. & Yokota, T. DUX4 signalling in the pathogenesis of facioscapulohumeral muscular dystrophy. *Int. J. Mol. Sci.* **21**, 729 (2020).
152. Bosnakovski, D. et al. The DUX4 homeodomains mediate inhibition of myogenesis and are functionally exchangeable with the Pax7 homeodomain. *J. Cell Sci.* **130**, 3685–3697 (2017).
153. Minderjahn, J. et al. Mechanisms governing the pioneering and redistribution capabilities of the non-classical pioneer PU.1. *Nat. Commun.* **11**, 402 (2020).
154. Raposo, A. et al. Ascl1 coordinately regulates gene expression and the chromatin landscape during neurogenesis. *Cell Rep.* **10**, 1544–1556 (2015).
155. Gerber, A. N., Klesert, T. R., Bergstrom, D. A. & Tapscott, S. J. Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes Dev.* **11**, 436–450 (1997).
156. Biddie, S. C. et al. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol. Cell* **43**, 145–155 (2011).
157. Fu, Y. et al. STAT5 promotes accessibility and is required for BATF-mediated plasticity at the Il9 locus. *Nat. Commun.* **11**, 4882 (2020).
158. Desanlis, I. et al. HOX13-dependent chromatin accessibility underlies the transition towards the digit development program. *Nat. Commun.* **11**, 2491 (2020).
159. Yu, B. et al. Reprogramming fibroblasts into bipotential hepatic stem cells by defined factors. *Cell Stem Cell* **13**, 328–340 (2013).
160. Pfisterer, U. et al. Direct conversion of human fibroblasts to dopaminergic neurons. *Proc. Natl Acad. Sci. USA* **108**, 10343–10348 (2011).
161. Son, E. Y. et al. Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* **9**, 205–218 (2011).
162. Marro, S. et al. Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* **9**, 374–382 (2011).
163. Lujan, E., Chanda, S., Ahlenius, H., Südhof, T. C. & Wernig, M. Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. *Proc. Natl Acad. Sci. USA* **109**, 2527–2532 (2012).
164. Szabo, E. et al. Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
165. Pereira, C. F. et al. Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell* **13**, 205–218 (2013).
166. Ieda, M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
167. Yamamoto, K. et al. Direct conversion of human fibroblasts into functional osteoblasts by defined factors. *Proc. Natl Acad. Sci. USA* **112**, 6152–6157 (2015).
168. Swinstead, E. E., Paakinaho, V., Presman, D. M. & Hager, G. L. Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: a new perspective: multiple transcription factors can effect chromatin pioneer functions through dynamic interactions with ATP-dependent chromatin remodeling factors. *Bioessays* **38**, 1150–1157 (2016).
169. Festuccia, N. et al. Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol.* **18**, 1139–1148 (2016).
170. Kadauke, S. et al. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* **150**, 725–737 (2012).
171. Caravaca, J. M. et al. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev.* **27**, 251–260 (2013).
172. Deluz, C. et al. A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes Dev.* **30**, 2538–2550 (2016).
173. Festuccia, N. et al. Transcription factor activity and nucleosome organization in mitosis. *Genome Res.* **29**, 250–260 (2019).
174. Soares, M. A. F. et al. Hierarchical reactivation of transcription during mitosis-to-G1 transition by Brn2 and Ascl1 in neural stem cells. *Genes Dev.* **35**, 1020–1034 (2021).

Acknowledgements
Work in the authors' laboratory was funded by a Foundation grant (FRN-154297) from the Canadian Institutes of Health.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Molecular Cell Biology thanks the anonymous reviewers for their contribution to the peer review of this work.

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