Distinct combinations of transcription factors are necessary to elicit cell fate changes in embryonic development. Yet within each group of fate-changing transcription factors, a subset called ‘pioneer factors’ are dominant in their ability to engage silent, unmarked chromatin and initiate the recruitment of other factors, thereby imparting new function to regulatory DNA sequences. Recent studies have shown that pioneer factors are also crucial for cellular reprogramming and that they are implicated in the marked changes in gene regulatory networks that occur in various cancers. Here, we provide an overview of the contexts in which pioneer factors function, how they can target silent genes, and their limitations at regions of heterochromatin. Understanding how pioneer factors regulate gene expression greatly enhances our understanding of how specific developmental lineages are established as well as how cell fates can be manipulated.

KEY WORDS: Cell fate control, Pioneer factors, Transcription

Introduction

A differentiated, multicellular organism is made up of diverse cell types that are induced and maintained by cell type-specific chromatin states and patterns of gene expression. When new cell fates are induced during embryonic development, tissue regeneration or cell reprogramming, regulatory transcription factors must suppress genes specific to the original cell fate and activate genes specific to the new cell fate. Rather than one regulatory protein being singularly responsible for determining each cell type, it is well established that particular combinations of transcription factors elicit cell fate changes and maintain cell identity (Yamamoto, 1985). But although groups of transcription factors elicit cell fate changes, each member of a group does not function identically in the mechanism of changing gene expression patterns. Data derived from chromatin-binding studies in embryos, biochemical reconstitution of chromatin in vitro, genomics, and genetics have revealed that certain transcription factors, termed ‘pioneer factors’, have the distinct ability to target nucleosomal DNA sites at silent genes that have not been marked for activity, and thus initiate the process whereby a collective of regulatory proteins...
can assemble at particular sequences and activate genes specific to a new fate. In this way, pioneer factors are fundamental to development and reprogramming, and have even been shown to play a role in cancer. Here and in the accompanying poster, we provide an overview of pioneer factors by summarizing their role in cell fate specification and conversion, as well as the molecular basis for their activity. Understanding how pioneer factors interact with chromatin and the limits of their activity will enhance our ability to manipulate cell fate control for diverse research and therapeutic purposes.

The molecular basis of pioneer factor activity
Gene regulation occurs in the context of chromatin, the complex of DNA and histone proteins that makes up the chromosomes. Within the chromatin, cellular DNA is wrapped nearly twice around an octamer of the four core histones to form arrays of nucleosomes. Linker histones bind to nucleosomes and stabilize condensed, repressive states. Chromatin can become ‘opened’ at gene regulatory regions, which allows RNA polymerase to function. This type of open chromatin is known as active chromatin, or chromatin type A. After a domain becomes open, it can typically accommodate the binding of any transcription factor and is accompanied by ‘active’ covalent modifications of the core histones, including histone H3 methylation on lysine 4 (H3K4me) and H3K9 and H3K27 acetylation (The ENCODE Project Consortium, 2011; Kharchenko et al., 2011). At promoters, these modifications can flank a nucleosome-free region immediately upstream of the transcription start site.

Large-scale chromatin-mapping studies have revealed that about half of the genome in a given cell possesses vast stretches of ‘low signal state’ chromatin that does not accumulate either active or repressive histone modifications (Kharchenko et al., 2011; Ho et al., 2014; Roadmap Epigenomics Consortium et al., 2015), and thus can be considered as unprogrammed. This type of chromatin is called low signal chromatin, or chromatin type L. Such unprogrammed chromatin is likely to be bound non-specifically by linker histones, which are nearly as abundant as core histones and are inherently repressive to transcription. Although many transcription factors cannot access target sites in the type L chromatin, pioneer factors are indeed capable of targeting this type of chromatin (Soufi et al., 2012; van Oevelen et al., 2015). In contrast to active and low signal chromatin, actively repressed chromatin domains – also known as repressed chromatin or chromatin type R – are inhibitory to the binding of most transcription factors, including pioneer factors, and thus strongly repress transcription. Such closed domains are typically accompanied by repressive covalent modifications of the core histones, including H3K9 or H3K27 methylation, although usually type R domains do not possess both marks at the same time, thus reflecting the different mechanisms of chromatin repression (Ho et al., 2014; Becker et al., 2016).

A central question regarding changes in cell fate is how genes in silent, low signal chromatin (type L) and repressed chromatin (type R) can be accessed and activated, despite the fact that the DNA is nucleosomal and thus inherently closed to most binding factors. The ability to target low signal chromatin is based on the inherent capacity of pioneer factors to recognize their target DNA sequences on the nucleosome (Cirillo et al., 1998), either by full or partial motif recognition. The FoxA DNA-binding domain (DBD) possesses a ‘winged helix’ structure that resembles the globular, nucleosomal binding domain of linker histone, binding to the full motif on one side of a DNA helix and leaving the other side of DNA free to bind core histones (Clark et al., 1993; Ramakrishnan et al., 1993; Cirillo et al., 1998). In vivo genetic and in vitro biochemical studies indicate that FoxA binding displaces linker histones from the local chromatin and keeps nucleosomes accessible (Cirillo et al., 1998; Iwafuchi-Doi et al., 2016). Furthermore, the C-terminal domain of FoxA can bind directly to core histone proteins and is required for opening chromatin, without ATP or ATP-dependent chromatin remodelers (Cirillo et al., 2002). In contrast to the full motif recognition binding of FoxA, some pioneer factors recognize only a partial motif: Oct3/4 (Pou5f1), Klf4 and basic helix-loop-helix type pioneer factors, such as Ascl1, target a partial DNA sequence of their canonical binding motifs on DNA, which is compatible with nucleosome binding (Soufi et al., 2015). It is now possible to predict a transcription factor’s pioneer activity by extensive computational analysis of genomic DNA motifs and the factor’s 3D structure (Soufi et al., 2015). In addition, pioneer factors can be identified empirically by their ability to target nucleosomal DNA in vivo (Soufi et al., 2015) and/or elicit an open region of chromatin (Sherwood et al., 2014).

Despite the importance of pioneer factor binding for initiating changes in cell fate, the process itself is insufficient to induce changes in gene expression directly. For this to take place, pioneer factor binding must occur cooperatively with other factors that, by themselves, cannot initiate such events in silent chromatin. Indeed, FoxA can recruit other activators or repressors to their target sites (Carroll et al., 2005; Zhang et al., 2005; Sekiya and Zaret, 2007; Li et al., 2012a; Iwafuchi-Doi et al., 2016). In Caenorhabditis elegans development, PHA-4, a homolog of FoxA, frequently binds promoters and recruits RNA polymerase II (Hsu et al., 2015). Thus, the initial chromatin-binding activity of pioneer factors is followed by the assembly of activating or repressing complexes, as dictated by the local DNA sequence and the presence of factors that are able to bind and further modify the chromatin.

Pioneer factors and developmental competence
The emergence of distinct cell types during embryonic development results from an intricate and dynamic cascade of regulatory changes in gene expression. Pioneer factors are likely to play a role in establishing a competence for many different cell fates, and indeed have already been shown to be fundamental in many developmental contexts, such as Pax7 for putitary melanotrope development and PU.1 (Spi1) for myeloid and lymphoid development (summarized by Iwafuchi-Doi and Zaret, 2014). In Drosophila embryos, the maternal transcription factor Zelda (Vielfaltig – FlyBase) plays a primary role in zygotic genome activation following fertilization (Liang et al., 2008). Zelda is present in nuclei earlier than other master regulators, such as Bicoid and Dorsal, and locally opens inactive regulatory regions to allow other transcription factors to bind (Li et al., 2014; Schulz et al., 2015; Sun et al., 2015). Similarly, maternal Class V POU factors (e. g. Oct3/4 and Pou5f3) and SoxB1 factors (e.g. Sox2 and Sox19) are primary regulators of zygotic genome activation in mouse (Foygel et al., 2008; Pan and Schultz, 2011) and zebrafish (Lee et al., 2013; Leichsenring et al., 2013).

The TALE family of homeodomain transcription factors, including pre-B cell leukemia homeobox (Pbx) and Meis homeobox, are important co-factors of Hox proteins. They engage the hoxb1a promoter during early zebrafish embryogenesis and recruit chromatin-modifying enzymes and RNA polymerase II to establish a poised state. Hoxb1b is then recruited and activates the hoxb1a promoter at a later stage (Choe et al., 2014). A recent study showed that Hoxa2 is recruited to a subset of sites that are pre-bound by Meis, which specifies the second branchial arches (Amin et al., 2012).
2015). Pbx also acts as a pioneer factor to enable the action of MyoD in specifying muscle fate (Berkes et al., 2004).

**Pioneer factors FoxA and GATA**

As a paradigm, the pioneer factors FoxA and GATA have been studied in great detail. Direct assessment of transcription factor occupancy on silent, liver-specific genes in early mouse embryos revealed that binding sites for FoxA and GATA transcription factors, but not sites for various other factors expressed in the liver lineage, were occupied in the undifferentiated foregut endoderm (Gualdi et al., 1996). Upon induction of the liver bud, other factors became engaged with the chromatin to initiate the liver-specific gene expression program. FoxA and GATA proteins are also expressed in areas outside the prospective liver region, for example in the medial-posterior endoderm. Here, their chromatin-binding activity still endows competence for the induction of liver genes, but this does not eventuate owing to the restrictive mesodermon interactions that normally inhibit the liver fate in this tissue (Bossard and Zaret, 2000; McLin et al., 2007).

Early evidence for the pioneer factor activity of FoxA and GATA came from attempts to model their binding with purified components in vitro (Cirillo et al., 1998; Cirillo and Zaret, 1999). Various liver-specific transcription factors were tested for their ability to bind to their sites on nucleosomal DNA and to engage nucleosome arrays harboring their binding sites, where the arrays were compacted by binding of linker histone. Remarkably, only purified FoxA protein could engage its target sequence on nucleosomes and enhance GATA factor binding, whereas the other factors tested could not. Furthermore, FoxA protein, and to a lesser extent Gata4, but not other liver-specific factors, could engage their target sites on the compacted nucleosome arrays and create a local open domain of chromatin, independent of nucleosome remodelers. Indeed, this is how the name ‘pioneer factors’ was coined (Cirillo et al., 2002).

The occupation by FoxA and GATA of a liver-specific enhancer in endoderm chromatin prior to the specification of liver cells, and the fact that at least FoxA could target nucleosomal DNA, suggested that these proteins functioned as competence factors (Zaret, 1999). Consistent with this, genetic inactivation of FoxA genes in the endoderm resulted in a failure of the endoderm to initiate hepatic differentiation (Lee et al., 2005). In C. elegans, when the FoxA homolog PHA-4 is first expressed it occupies target sequences of highest affinity, but as its expression is elevated during endoderm development, PHA-4 begins to occupy lower affinity targets, indicating that its action in binding to chromatin is concentration dependent (Gaudet and Mango, 2002). Recent genomics studies of pancreatic differentiation of human embryonic stem cells showed that FoxA binding occurs at genes in endoderm chromatin is concentration dependent (Gaudet and Mango, 2002). Upon induction of the liver bud, other factors became engaged with the chromatin to initiate the liver-specific gene expression program. FoxA and GATA proteins are also expressed in areas outside the prospective liver region, for example in the medial-posterior endoderm. Here, their chromatin-binding activity still endows competence for the induction of liver genes, but this does not eventuate owing to the restrictive mesodermon interactions that normally inhibit the liver fate in this tissue (Bossard and Zaret, 2000; McLin et al., 2007).

The occupation by FoxA and GATA of a liver-specific enhancer in endoderm chromatin prior to the specification of liver cells, and the fact that at least FoxA could target nucleosomal DNA, suggested that these proteins functioned as competence factors (Zaret, 1999). Consistent with this, genetic inactivation of FoxA genes in the endoderm resulted in a failure of the endoderm to initiate hepatic differentiation (Lee et al., 2005). In C. elegans, when the FoxA homolog PHA-4 is first expressed it occupies target sequences of highest affinity, but as its expression is elevated during endoderm development, PHA-4 begins to occupy lower affinity targets, indicating that its action in binding to chromatin is concentration dependent (Gaudet and Mango, 2002). Recent genomics studies of pancreatic differentiation of human embryonic stem cells showed that FoxA binding occurs at genes in endoderm chromatin is concentration dependent (Gaudet and Mango, 2002). Upon induction of the liver bud, other factors became engaged with the chromatin to initiate the liver-specific gene expression program. FoxA and GATA proteins are also expressed in areas outside the prospective liver region, for example in the medial-posterior endoderm. Here, their chromatin-binding activity still endows competence for the induction of liver genes, but this does not eventuate owing to the restrictive mesodermon interactions that normally inhibit the liver fate in this tissue (Bossard and Zaret, 2000; McLin et al., 2007).

**Pioneer factors in cell fate conversions**

To date, diverse cell fate conversions have been reported by various combinations of transcription factors, typically including pioneer factors. The most dramatic example is the reprogramming of fibroblasts into induced pluripotent stem cells (iPSCs) by only four transcription factors: Oct3/4, Sox2, Klf4 and c-Myc (Takahashi and Yamanaka, 2006). Of these factors, Oct3/4, Sox2 and Klf4 act as pioneer factors in that they can access silent, low signal state chromatin, regardless of whether they bind together or alone (Soufi et al., 2012, 2015). By contrast, c-Myc alone prefers to bind to active chromatin, but can also bind low signal chromatin sites when cooperating with the other factors (Soufi et al., 2012, 2015). There are more examples of a combination of cell type-specific pioneer factors and co-factors directing cell conversion: PU.1 and C/EBPα can reprogram macrophage-like cells from fibroblasts (Feng et al., 2008; van Oevelen et al., 2015); Gata4, Met2c, Tbx5, Hand2 and Nkx2-5 can reprogram cardiomyocyte-like cells from fibroblasts (Ieda et al., 2010; Addis et al., 2013); Ascl1, Brm2 (Pou3f2) and Myt11 can reprogram functional glattameric neurons from fibroblasts (Vierbuchen et al., 2010); and FoxA, Gata4 and Hnf4α/1α can reprogram hepatocyte-like cells from fibroblasts (Huang et al., 2011; Sekiya and Suzuki, 2011). In all of these cases, at least one component of the factor combination plays a central role as a pioneer factor: PU.1 (macrophage-like cells), Gata4 (cardiomyocyte-like cells), Ascl1 (glutaminergic neurons) and FoxA (hepatocyte-like cells) (summarized by Iwafuchi-Doi and Zaret, 2014). For instance, Ascl1 binds silent, low signal chromatin and recruits Brm2 to target sites, and Brm2 is primarily required for the later stage of cell conversion by contributing to cell maturation (Wapinski et al., 2013). These findings demonstrate that even when the factors are expressed simultaneously, they function in a hierarchical manner, and that pioneer factors act first to establish a competence for a specific cell fate, which is followed by co-factor binding to instruct further differentiation.

**Pioneer factors in cancers**

As pioneer factors play a primary role in gene regulation, it is no surprise that their mis-regulation can compromise human health. Indeed, in many forms of cancer, pioneer factors are up- or downregulated, mutated or amplified in their genomic region, or alternatively the DNA sequence of the pioneer factors’ binding sites is mutated. In esophageal and lung squamous cell carcinomas, chromosome segments containing SOX2 are often amplified (Bass et al., 2009). In the case of skin squamous-cell carcinoma, Sox2 is the most upregulated transcription factor, and conditional deletion of Sox2 markedly decreases skin tumor formation (Boumahdi et al., 2014). FoxA and GATA factors are involved in a variety of hormone-dependent cancers, such as estrogen receptor (ER)-positive breast cancer, androgen receptor (AR)-positive prostate cancer, and ER-dependent resistance and AR-mediated facilitation of liver cancer, and FOXA levels correlate well with clinical outcomes (reviewed by Jozwik and Carroll, 2012). Furthermore, single nucleotide polymorphisms at FOXA binding sites reduce binding of FOXA and ER in liver and correlate with hepatocellular carcinoma development in female patients (Li et al., 2012b).

Thus far, non-nuclear receptor transcription factors, such as pioneer factors, have been considered ‘undruggable’ targets for cancer treatment, but they could be attractive targets that avoid the issues of drug resistance that usually occur when targeting intracellular signaling pathways (Johnston and Carroll, 2015). The FoxM1-DNA interaction, which is upregulated in a wide range of cancers, is inhibited by direct interaction with a natural product called thiostepton (Hegde et al., 2011). Although findings such as these may hold promise for future cancer treatments, a better understanding of how pioneer factors function will be required in order to target them reliably for therapeutic breakthroughs.
Conclusions

Pioneer factors are among the master regulators of cell fate. They function by initiating chromatin targeting events on nucleosomal DNA, typically in low signal chromatin regions where the presence of linker histones represses transcription. The local exposure of chromatin brought about by pioneer factor binding allows other, non-pioneer transcription factors to access nucleosomal DNA, which in turn drives lineage-specific gene expression and selection of cell fate. The ability of pioneer factors to target silent genes and allow other factors to bind provides a mechanistic explanation for the long-standing phenomenon of developmental competence, in which a tissue gains the potential to execute a cell fate decision. However, pioneer factors do not occupy all in silico target sites in the genome; they are actively excluded from heterochromatic domains spanned by H3K9me2/3 (type R chromatin) (Lupien et al., 2008; Soufi et al., 2012), among others. By presenting a barrier to factor binding, heterochromatic, repressive domains provide a means for cells to stably retain their fate (Becker et al., 2016). We suggest that a possible reason for cell conversions being typically of a low efficiency and failing to shut off their initial genetic program (Cahan et al., 2014) may relate to the inefficiency of reprogramming factors in engaging with heterochromatic domains that span genes for which expression is required for the desired cell type. It is possible to alter chromatin state broadly by applying small molecules to target chromatin-modifying enzymes, but such changes will occur globally throughout the genome. We speculate that understanding how cell type-specific heterochromatic domains are established and how pioneer factors can overcome such barriers during development will provide more targeted ways to manipulate cell fate in health and disease. More broadly, further work in the field should be aimed at understanding how different pioneer factors target silent, low signal-state chromatin and how heterochromatic features at highly repressed chromatin might block pioneer factor binding. These detailed mechanistic insights will pave the way for the future ability to program and reprogram cell fates at will.

Acknowledgements

We thank Eileen Hulme for help in preparing the manuscript.

Competing interests

The authors declare no competing or financial interests.

Funding

Work on pioneer factors was supported by the National Institutes of Health [GM06477 to K.S.Z.] M.I.-D. was supported by postdoctoral fellowships from the Japan Society for the Promotion of Science and the Naito, Astellas and Uehara foundations.

Development at a Glance

A high-resolution version of the poster is available for downloading in the online version of this article at http://dev.biologists.org/content/143/11/1833/F1.poster.jpg

References


