© 2018. Published by The Company of Biologists Ltd | Development (2018) 145, dev170209. doi:10.1242/dev.170209

Targeted DamID reveals differential binding of mammalian pluripotency factors

Seth W. Cheetham*,¶, Wolfram H. Gruhn¶, Jelle van den Ameele¶, Robert Krautz¶, Tony D. Southall‡, Toshihiro Kobayashi§, M. Azim Surani and Andrea H. Brand**

ABSTRACT

The precise control of gene expression by transcription factor networks is crucial to organismal development. The predominant approach for mapping transcription factor-chromatin interactions has been chromatin immunoprecipitation (ChIP). However, ChIP requires a large number of homogeneous cells and antisera with high specificity. A second approach, DamID, has the drawback that high levels of Dam methylase are toxic. Here, we modify our targeted DamID approach (TaDa) to enable cell type-specific expression in mammalian systems, generating an inducible system (mammalian TaDa or MaTaDa) to identify genome-wide protein/DNA interactions in 100 to 1000 times fewer cells than ChIP-based approaches. We mapped the binding sites of two key pluripotency factors, OCT4 and PRDM14, in mouse embryonic stem cells, epiblast-like cells and primordial germ cell-like cells (PGCLCs). PGCLCs are an important system for elucidating primordial germ cell development in mice. We monitored PRDM14 binding during the specification of PGCLCs, identifying direct targets of PRDM14 that are key to understanding its crucial role in PGCLC development. We show that MaTaDa is a sensitive and accurate method for assessing cell type-specific transcription factor binding in limited numbers of cells.

KEY WORDS: ChIP-seq, Embryonic stem cells, Oct4, Prdm14, Primordial germ cells, Targeted DamID

INTRODUCTION

Chromatin immunoprecipitation (ChIP) has been widely used to characterise transcription factor-chromatin interactions (Furey, 2012), but this approach is limited by the requirement for large numbers of homogeneous cell populations and specific antibodies (Tsankov et al., 2015). As a result, mapping transcription-factor occupancy *in vivo* in rare cell types, such as stem cells, is technically challenging. DNA adenine methylation identification (DamID) has recently emerged as an alternative approach for genome-wide profiling (Marshall and Brand, 2015; Marshall et al., 2016; Otsuki et al., 2014; Southall et al., 2013; van Steensel and Henikoff, 2000). In DamID, a DNA- or chromatin-binding protein is fused to an *E. coli* Dam methylase. Wherever the Dam-fusion protein binds to DNA or chromatin, it methylates adenine within the sequence GATC. Endogenous adenine methylation is extremely rare in eukaryotes (Koziol et al., 2015; Wu et al., 2016; Zhang et al., 2015) so the tagged sequences can be detected easily by digestion with *Dpn*I, which only cuts at methylated GATC sites. In this way, binding sites can be identified genome-wide without cell isolation, fixation or immunoprecipitation.

Although DamID is particularly well suited for *in vivo* analysis, a major caveat is cytotoxicity resulting from high levels of expression of the Dam methylase (Southall et al., 2013; Catherine Davidson and A.H.B., unpublished). As a result, DamID in mammalian cells has generally relied on low-level, ubiquitous expression from an uninduced heat-shock promoter (van Steensel and Henikoff, 2000; Vogel et al., 2007). However, this precludes the identification of cell type-specific binding or detection of dynamic changes in DNA or chromatin interactions. To overcome these limitations, we modified targeted DamID (TaDa) (Southall et al., 2013) for use in mammalian cells, enabling the rapid, accurate and sensitive identification of transcription factor-binding sites. Mammalian TaDa (MaTaDa) enables genome-wide profiling of protein-DNA interactions in a temporally regulated, cell type-specific fashion.

First, we validated MaTaDa in murine embryonic stem cells by reanalysing OCT4 occupancy during the transition from the naïve to primed pluripotent cell state (Buecker et al., 2014). Next, we mapped the binding sites of the transcription factor PRDM14, which controls embryonic stem cell (ESC) pluripotency and is of pivotal importance for acquisition of primordial germ cell (PGC) fate in mice. Making use of *in vitro* specification of PGC-like cells (PGCLCs), we identified a set of novel cis-regulatory elements bound by PRDM14 specifically in PGCLCs. Analysis of these loci suggests that PRDM14 is involved in the suppression of EGFR/MAPK signalling and the regulation of genes associated with cell migration.

RESULTS

A mammalian system for targeted DamID

We engineered a construct for conditional expression of Dam-fusion proteins in mammalian cells, comprising a ubiquitous promoter (PGK) driving expression of a transcript encoding a primary open reading frame encoding mCherry (ORF1 246 amino acids). The primary ORF is followed by two TAA stop codons and a single nucleotide frameshift upstream of a secondary open reading frame encoding the Dam fusion protein (ORF2; Fig. 1A). We have shown previously that translation of this bicistronic mRNA results in expression of ORF1 followed by rare ribosomal re-entry and translational re-initiation, resulting in extremely low levels of expression of ORF2, the Dam fusion protein (Southall et al., 2013).

For spatial and temporal control of MaTaDa, we inserted a GFP or puromycin resistance-coding sequence and SV40 terminator, flanked by loxP sites, between the PGK promoter and the TaDa

The Gurdon Institute and Department of Physiology, Development and Neuroscience, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK.
*Present address: Mater Research Institute-University of Queensland, Woolloongabba QLD 4102, Australia. ‡Present address: Department of Life Sciences, Imperial College London, Sir Ernst Chain Building, London SW7 2AZ, UK. §Present address: National Institute for Physiological Sciences, Okazaki, Aichi 444-8585, Japan.
¶These authors contributed equally to this work

**Author for correspondence (a.brand@gurdon.cam.ac.uk)

S.W.C., 0000-0001-6428-3175; J.V.D.A., 0000-0002-2744-0810; M.A.S., 0000-0002-8640-4318; A.H.B., 0000-0002-2089-6954

Received 16 July 2018; Accepted 23 August 2018

Development (2018) 145, dev170209. doi:10.1242/dev.170209

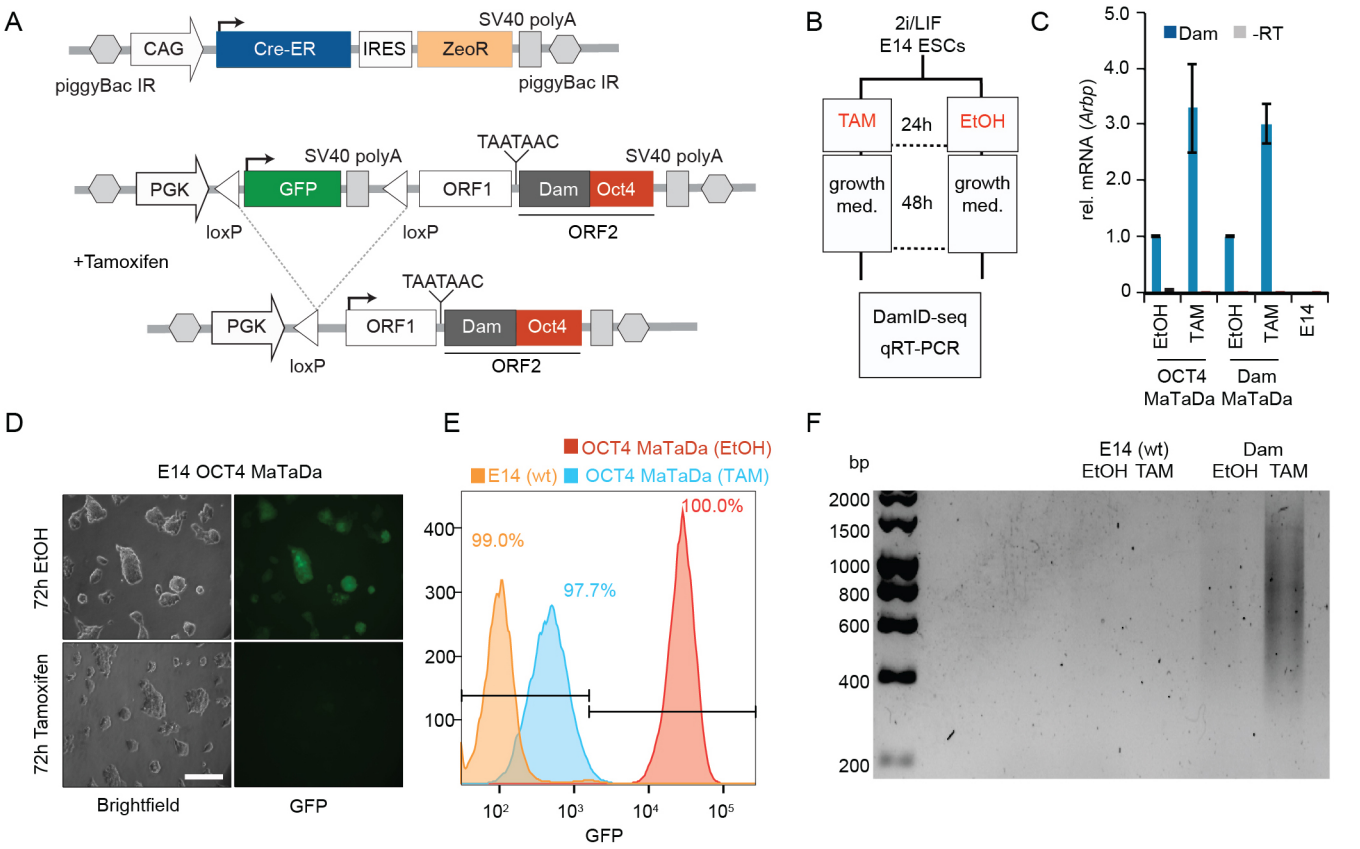

Fig. 1. Conditional expression of Dam-OCT4 in ESCs. (A) The PGK promoter drives expression of a floxed GFP cassette (green). Upon Cre induction, the floxed cassette is excised, allowing expression of ORF1 (white; 246 amino acids). Rare translational re-initiation results in low-level expression of the Dam-OCT4 fusion protein (ORF2; grey/red). Cre-ER is constitutively expressed and translocates to the nucleus upon tamoxifen treatment. (B) E14 mESCs were transformed with Dam-alone or Dam-OCT4 MaTaDa constructs were treated with tamoxifen or ethanol for 24 h and then allowed to grow for 48 h. The cells were then processed for DamID-seq and qRT-PCR. (C) Induction of Dam transcription after tamoxifen treatment compared with ethanol control measured by qRT-PCR. Data are mean ±s.e.m. (D) Following tamoxifen treatment, MaTaDa-containing ESCs rapidly lose GFP fluorescence. Scale bar: 100 µm. (E) FACS analysis demonstrates the efficient loss of GFP fluorescence in MaTaDa-expressing cells following Cre induction, with some perdurance of GFP protein at 72 h resulting in higher levels of fluorescence compared with the parental cell line (E14). (F) Amplification of mouse genomic DNA methylated by MaTaDa after tamoxifen treatment.

construct (Fig. 1A, Fig. S2A). To excise the floxed cassette, we used a Cre-estrogen-receptor fusion (Cre-ER) (Fig. 1A, Fig. S2A). Cre-ER is constitutively expressed, but only translocates to the nucleus and induces Dam expression upon tamoxifen administration (Fig. 1B,C). In the absence of Cre, GFP or puromycin resistance are expressed and transcription is terminated upstream of ORF1. The expression of GFP or the puromycin resistance cassette can be used to assess transfection efficiency or to select transfected cells. Excision of the stop-cassette results in loss of GFP or puromycin resistance expression and induction of TaDa. In this way, MaTaDa enables both spatial and temporal control, directed by targeted expression of Cre, in either cell lines or transgenic animals. The low levels of Dam methylase expression are non-toxic and preclude dominant effects that might result from the overexpression of transcription factors. To lessen the potential for steric effects between the Dam methylase and fused proteins (Ramialison et al., 2017), we inserted a myc tag as a spacer.

Mapping binding sites of pluripotency factors with MaTaDa

During early mammalian development, embryonic cells have the potential to form all cell types of the embryo. Naïve and primed pluripotent states have been characterised in mouse based on the functional, transcriptional and epigenetic characteristics of the pre- and postimplantation epiblast (Nichols and Smith, 2009). Mouse embryonic stem cells (mESCs) are used extensively as an *in vitro* model to study the molecular mechanisms of pluripotency. In the presence of small molecules that promote Wnt/β-catenin and inhibit FGF/MAPK signalling, mESCs remain in a naïve pluripotent state similar to the pre-implantation epiblast (Ying et al., 2008) (Fig. S1A). Stimulation of the FGF signalling pathway promotes differentiation of ESCs into epiblast-like cells (EpiLCs), a primed pluripotency state. The unifying mechanistic features of pluripotency are key transcription factors such as OCT4 (POU5F1). OCT4 is a master regulator of both primed and naïve pluripotency in ESCs (Zeineddine et al., 2014). During the transition from naïve to primed pluripotency, the OCT4-binding pattern changes dynamically due to the availability of co-factors (Buecker et al., 2014). PRDM14, which can function as a transcriptional repressor (Nady et al., 2015; Yamaji et al., 2013), promotes naïve pluripotency in mESCs (Ma et al., 2011; Yamaji et al., 2013). PRDM14 is also crucial for the development of primordial germ cells (Yamaji et al., 2008).

To assess whether MaTaDa can detect differential binding of transcription factors, we generated stable mESC lines carrying MaTaDa constructs encoding either Dam alone or a Dam-transcription factor fusion protein, together with the CreER-expression construct with a zeocine resistance gene (Fig. 1A, Fig. S2A). Cells were selected for GFP expression or puromycin resistance and co-selected with zeocine and expanded.

Development (2018) 145, dev170209. doi:10.1242/dev.170209

Induction of CreER with tamoxifen for 24 h resulted in induction of Dam expression and the loss of GFP fluorescence from virtually all cells (97.7%) 48 h later (Fig. 1D,E). Robust methylation of genomic DNA was detected in treated cells when compared with untransfected or ethanol-treated cells (Fig. 1F). Some faint DNA amplification was observed in cells treated with ethanol (Fig. 1F), possibly owing to low-level expression before tamoxifen treatment and high sensitivity of the DamID technique. However, the considerably higher induction upon tamoxifen treatment allowed us to identify changes in transcription factor-binding patterns during differentiation. Importantly, tamoxifen treatment of Dam-OCT4 or Dam-PRDM14 expressing ESCs did not lead to a large increase in *Oct4* or *Prdm14* expression, and consequently there was no effect on pluripotency or differentiation (Figs S1B, S2C).

MaTaDa identifies genome-wide transcription factor occupancy with high accuracy and sensitivity

We sequenced DamID libraries from 150,000 mESCs expressing either Dam alone (control), Dam-OCT4 or Dam-PRDM14. OCT4 bound to 18,103 sites and PRDM14 to 8784 sites when a cut-off of $q<10^{-25}$ was used (Fig. S3). MaTaDa peaks were consistently detected between biological replicates (Fig. S3A-D), and genome-wide correlations between replicate experiments are shown in Fig. S3E,F. We detected binding to key OCT4 targets, such as *Nanog*, *Sox2*, *Klf4* and *Myc* (Fig. 2A). Similarly, PRDM14 was found to bind key genes involved in pluripotency, including *Oct4* and *Nanog*, and differentiation, such as *Fgfr1* and *Xist* (Fig. S4A).

Next, we compared our data for OCT4 and PRDM14 binding in 150,000 mESCs to published ChIP-seq data from 50 and 20 million mESCs, respectively (Buecker et al., 2014; Ma et al., 2011). Importantly, owing to the nature of DamID in acquiring counts only at GATC motifs in the genome, the resolution of signal across the genome is different from ChIP-seq. In addition, random methylation requires stringent normalization to the Dam-only control, and a ratio is calculated between Dam-fusion and DAM-only. This prevents direct comparisons between MaTaDa and ChIP-seq in a quantitative manner. We therefore used a peak-centred analysis for genome-wide comparisons between both techniques. Both the OCT4 and PRDM14 MaTaDa signals were highly enriched over ChIP peaks (Fig. 2B, Fig. S4B). Conversely, ChIP-seq signal is highly enriched over MaTaDa peaks (Fig. 2C, Fig. S4C). Overlap between peaks was dependent on the stringency of peak calling (Fig. 2D-G, Fig. S4D-F), but at $q<10^{-25}$, 1901 of 3880 (49%) OCT4 ChIP peaks overlapped with MaTaDa peaks (18,096 peaks at $q<10^{-25}$) (Fig. 2B-G, Fig. S5A-C), as did 1824 of 5681 (32%) of PRDM14 ChIP-seq peaks (8784 MaTaDa peaks at $q<10^{-25}$) (Fig. S4B-J) (Ma et al., 2011). Nevertheless, at any given q-value, a subset of peaks was always specific to either technique (Fig. 2F,G, Figs S4E,F, S5A-E). Interestingly, although for both ChIP-seq and for DamID it is generally thought that peak intensity grossly correlates with binding strength, the correlation (R^2) for peak intensity of peaks common to both techniques (at $q<10^{-25}$) was only 0.07 for OCT4 and 0.12 for PRDM14 (Fig. 2H, Fig. S4H).

We next conducted motif and genomic feature enrichment analysis (Imrichová et al., 2015) on the OCT4 MaTaDa peaks ($q<10^{-25}$). The three most highly ranked transcription factor motifs for which a position weight matrix was available all corresponded to OCT4-related motifs (Fig. 2I). Interestingly, presence of OCT4 motifs at any given q-value was higher under common and ChIP-specific than under MaTaDa-specific peaks (Fig. S5F). This could suggest that MaTaDa captures more indirect OCT4-binding events than ChIP-seq in these conditions. Peaks were also enriched for enhancers and active chromatin, as illustrated by the enrichment for ESC DNAse-accessible sites and H3K27ac and H3K4me1 histone marks (Fig. 2J). The enrichment for genomic features did not change considerably when either MaTaDa- or ChIP-seq-specific peaks were analysed (Fig. S5G,H).

We conclude that MaTaDa was able to profile genome-wide transcription factor occupancy accurately and with high sensitivity, and can therefore function as an alternative or complementary approach to ChIP-seq.

MaTaDa is sufficiently sensitive to profile rare populations of cells

A major advantage of TaDa over ChIP is the ability to profile rare populations of cells *in vivo* (Otsuki et al., 2014; Southall et al., 2013; Marshall and Brand, 2017; Aughey et al., 2017). To test the sensitivity of MaTaDa, we tested whether MaTaDa could profile binding in only 10,000 cells. Remarkably, the binding profiles for PRDM14 were strikingly similar to those obtained from 150,000 cells (Fig. 3A-C). However, peak calling at any given q-value always resulted in more peaks being called from 10,000 than 150,000 cells (Fig. 3B-E,G), owing to a lower signal-to-noise ratio, as expected from the low cell number. Nevertheless, the sensitivity of this low-cell number MaTaDa was very high, with nearly all peaks from 150,000 cells ($q<10^{-100}$) eventually being recovered by PRDM14 MaTaDa on 10,000 cells (Fig. 3F). The ability to profile transcription factor binding in small numbers of cells suggests that MaTaDa has sufficient sensitivity to uncover transcription factor-genome interactions in rare cell types *in vivo*.

MaTaDa captures cell type-specific transcription factor binding

We tested whether MaTaDa could profile differential transcription factor occupancy between different but related cell types by generating OCT4-binding profiles during ESC differentiation. Removal of 2i and LIF, and addition of FGF2 and activin drives the transition of naïve ESCs to epiblast-like stem cells (EpiLCs) (Hayashi et al., 2011). EpiLCs are analogous to the cells of the post-implantation epiblast and are in a primed pluripotent state. During the transition from ESCs to EpiLCs, OCT4 interacts with OTX2, and together they bind a distinct set of enhancers to promote the activation of EpiLC-associated genes (Buecker et al., 2014).

First, we generated genome-wide MaTaDa profiles of OCT4 occupancy in naïve ESCs and EpiLCs (Fig. 4). Next, we compared our data with a previously defined set of binding sites that are bound by OCT4 predominantly in either naïve ESCs or EpiLCs (Buecker et al., 2014). We found that ground-state pluripotency genes, such as *Klf4* (Fig. 4A), were bound primarily in ESCs, whereas primed pluripotency-associated genes, such as *Fgf5* (Fig. 4B), were bound exclusively in EpiLCs, but not in the ESCs from which they were derived. Crucially, ESC-specific enhancers were not strongly bound following differentiation (Fig. 4C), demonstrating that MaTaDa is able to capture differential transcription factor occupancy, enabling the detection of spatially and temporally restricted protein-chromatin interactions.

PRDM14 binding in ESCs and PGCLCs

In addition to controlling ESC pluripotency, PRDM14 is essential for specifying PGCs from post-implantation epiblast cells (Yamaji et al., 2008). Paralleling mouse embryonic development, the establishment of a primed pluripotent state in EpiLCs is required for specification of PGCLCs *in vitro* (Hayashi et al., 2011; Ohinata et al., 2009). Notably, PGCLCs are functionally equivalent to the

Development (2018) 145, dev170209. doi:10.1242/dev.170209

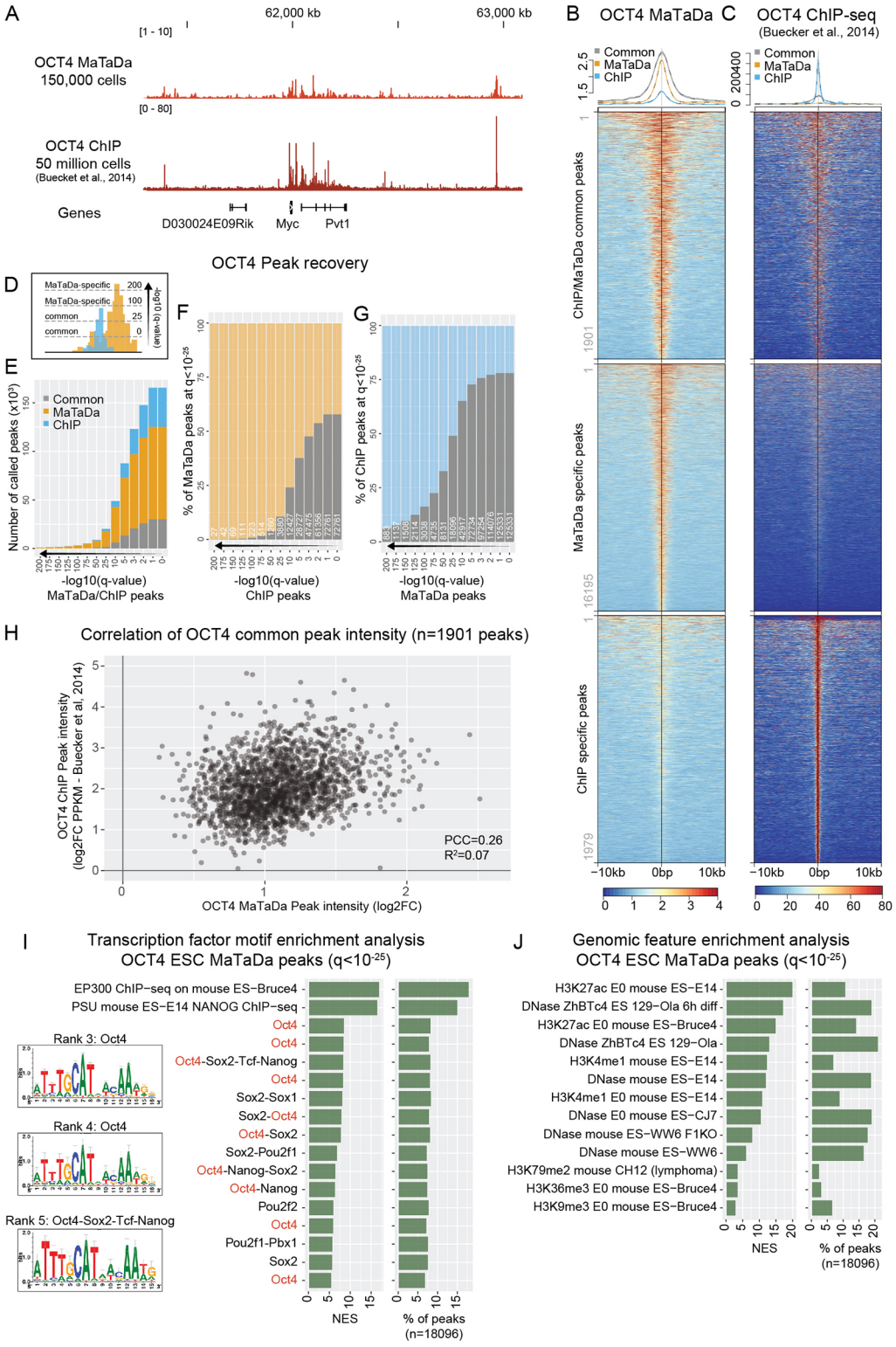

Fig. 2. MaTaDa accurately profiles genome-wide transcription-factor occupancy. (A) Genome browser view of OCT4 binding at the Myc locus (MaTaDa; average of three replicates) compared with ChIP. MaTaDa data are represented as fold enrichment of Dam-fusion over Dam-only; ChIP-seq data are represented as aligned reads. (B,C) OCT4 MaTaDa (B) and ChIP-seq (C) ESC signal is plotted over a 10 kb window either side of the peak midpoint, for peaks common to MaTaDa and ChIP-seq (top), specific to MaTaDa only (middle) and specific to ChIP-seq only (bottom) at $q<10^{-25}$. Above are metaplots of the MaTaDa (B) and ChIP-seq (C) signal. (D) Schematic to illustrate how peak recovery between two different datasets can vary depending on the q-value. (E-G) Number (E) or percentage (F,G) of peaks called upon changing the q-value for peak detection, either for MaTaDa and ChIP-seq in parallel (E), or compared with a fixed q-value $<10^{-25}$ for MaTaDa (F) or ChIP-seq (G). Common peaks are grey, MaTaDa-specific peaks are orange, ChIP-seq-specific peaks are blue. (H) Scatterplot of peak intensity for peaks ($q<10^{-25}$) common to OCT4 MaTaDa and ChIP-seq. PCC, Pearson correlation coefficient. (I,J) Transcription factor motif (I) and genomic feature (J) enrichment analysis of OCT4 MaTaDa peaks ($q<10^{-25}$, 18,096 peaks). Position weight matrices (PWM) are shown for the top three enriched motifs for which a PWM was available. Normalised enrichment score (NES) and percentage of peaks containing the feature are indicated.

PGCs found in the early mouse embryo and hence represent an important system for studying mammalian germ line development (Hayashi et al., 2011). PGC identity is controlled by a transcription factor network consisting of BLIMP1, PRDM14 and TFAp2c, which suppresses expression of somatic genes and promotes transcription of germ cell genes (Magnúsdóttir et al., 2013; Nakaki et al., 2013). Expression of PRDM14 was shown to be sufficient to drive differentiation of EpiLCs into PGCLCs (Nakaki et al., 2013). However, deriving sufficient quantities of PGCLCs is difficult and has limited the mechanistic understanding of this pivotal developmental process. In particular, it has not been possible to determine whether PRDM14 binding changes during PGCLC development and the key PRDM14 targets that drive the EpiLC-PGCLC transformation have not been identified previously.

We induced differentiation of EpiLCs into PGCLCs using the growth factors BMP4, BMP8, SCF, EGF and LIF. PGCLCs

Development (2018) 145, dev170209. doi:10.1242/dev.170209

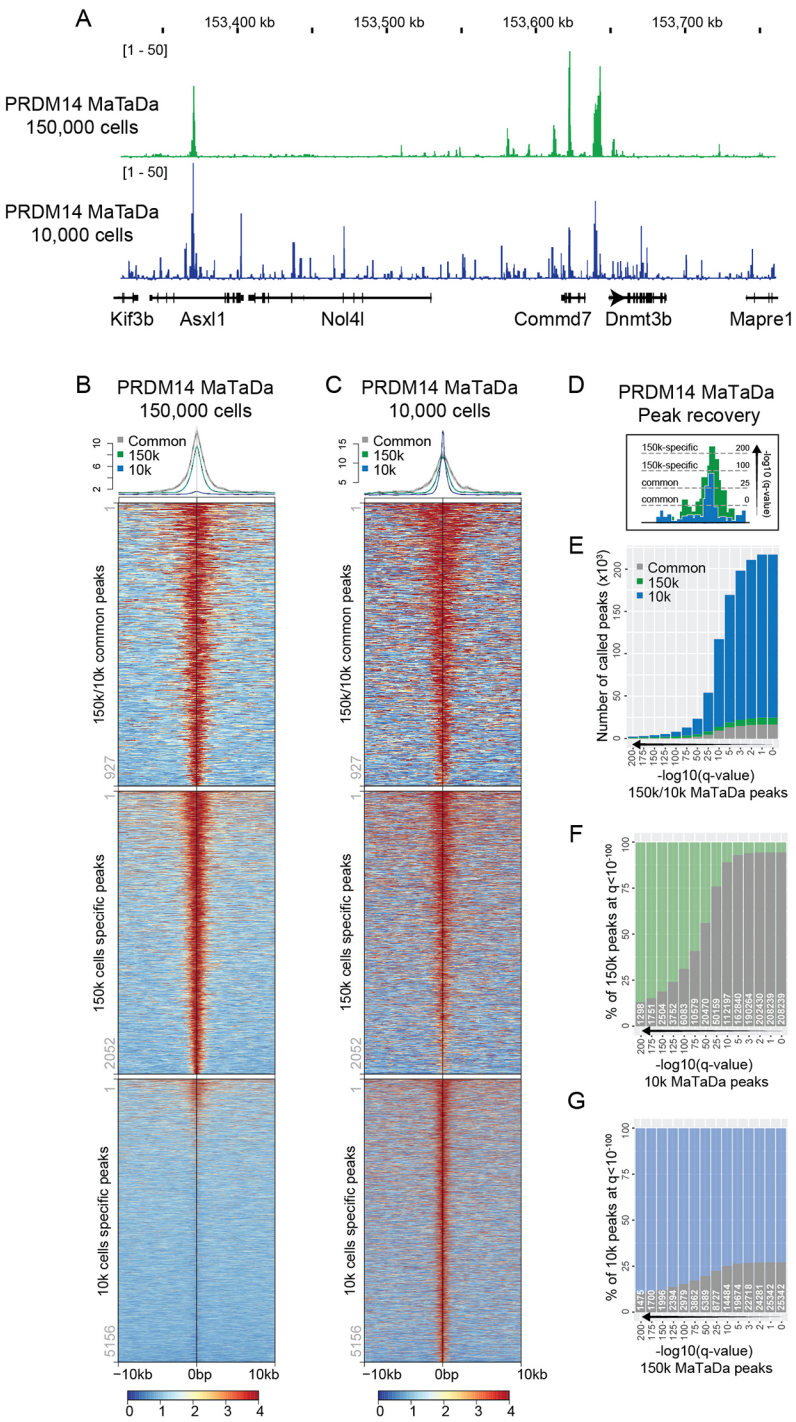

Fig. 3. MaTaDa has sufficient sensitivity to profile rare cell populations. (A) Genome browser view of PRDM14 occupancy at the Dnmt3b locus from 10,000 (10k) cells (average of five replicates) and 150,000 (150k) cells (average of three replicates). Data are represented as fold enrichment of Dam fusion over Dam only. (B,C) PRDM14 MaTaDa signal from 150,000 ESCs (B) and 10,000 ESCs (C) is plotted over a 10 kb window either side of the peak midpoint for peaks common to 150k and 10k (top), specific to 150k only (middle) and specific to 10k only (bottom) at $q<10^{-100}$. Above are metaplots of the 150k (B) and 10k (C) signal. (D) Schematic to illustrate how peak recovery can vary depending on the q-value. (E-G) Number (E) or percentage (F,G) of peaks called upon changing the q-value for peak detection, either for 150k and 10k ESCs in parallel (E), or compared with a fixed q-value $<10^{-100}$ for 150k (F) or 10k (G). Common peaks are grey; 150k-specific peaks are green; 10k-specific peaks are blue.

(~10-15% PGCLC specification efficiency, up to 150,000 PGCLCs per replicate) were FACS purified using endogenous DPPA3-GFP and ESG1-tdTomato reporters (see Materials and Methods, Fig. S6). Using MaTaDa, we monitored PRDM14 binding in the first 3 days of PGCLC specification (Fig. S6A-D) and found genome-wide changes in PRDM14 binding compared with ESCs (Fig. 5A-C). Whereas 77% of binding sites ($q<10^{-100}$) were shared between ESCs and PGCLCs, the level of PRDM14 occupancy was strikingly different: 2450 sites were preferentially bound by PRDM14 in ESCs (>2 fold higher in ESCs) and 698 in PGCLCs (>2 fold higher in PGCLCs), suggesting that PRDM14 regulates a functionally distinct set of genes in ESCs and PGCLCs (Fig. 5C). Most PRDM14 peaks occur more than 2 kb from transcriptional start sites (Fig. S7A-C) at distinct sites in ESCs and PGCLCs (Fig. S7C). Chen et al. (2012) defined 1277 putative enhancers in ESCs, based upon chromatin marks and transcription factor occupancy. Interestingly, we found that ESC-enriched PRDM14 sites coincided with these presumptive regulatory regions, but PGCLC-enriched sites did not (Fig. 5D). The PGCLC-enriched sites were enriched for the PRDM14 core motif (GGTCTCTAA; *P*=4.39e-6). Interestingly, by *de novo* analysis we discovered another motif (Fig. S7E; E=2.1e-33) that is similar to motifs recognised by RXRG (E=4.80e-5), the pluripotency factors NR5A2 (E=1.52e-4) and NR6A1 (E=4.16e-2), indicating that PGCLC-enriched sites may be regulatory regions bound by

Development (2018) 145, dev170209. doi:10.1242/dev.170209

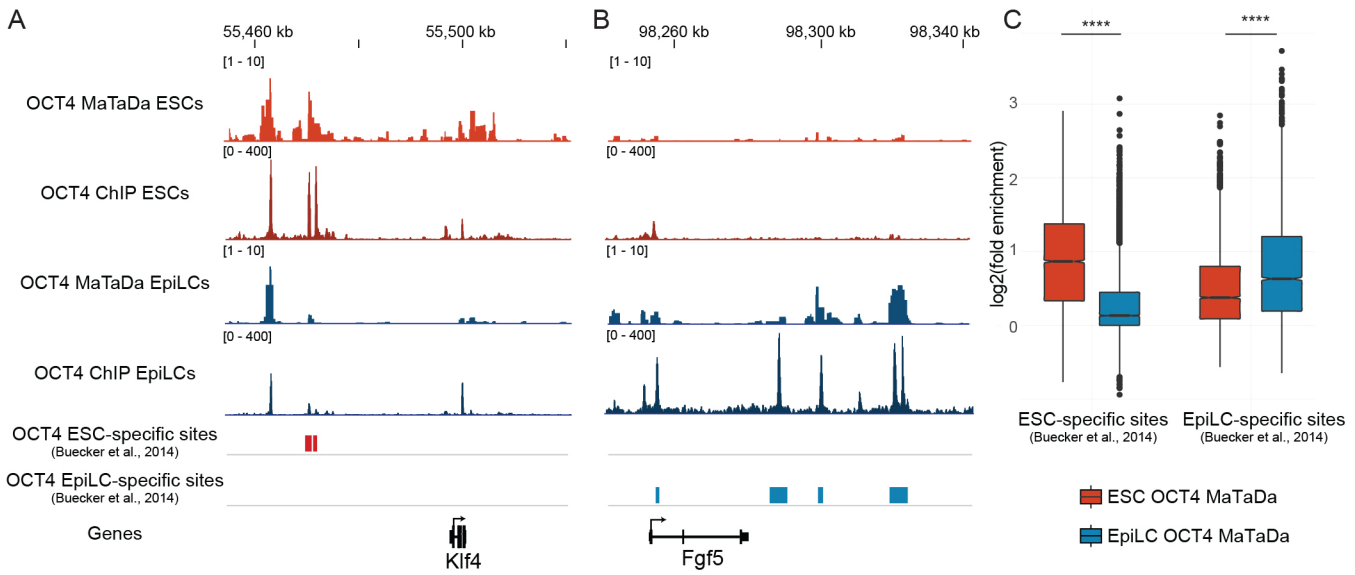

Fig. 4. Differential binding of OCT4 in ESCs and EpiLCs, as identified by MaTaDa. (A) Genome browser view of the ground state pluripotency gene *Klf4*, showing OCT4 binding to nearby enhancers in ESCs (average of three replicates) but not EpiLCs (average of two replicates). (B) Genome browser view of the pro-differentiation gene *Fgf5*, showing OCT4 binding to nearby enhancers in EpiLCs (average of two replicates) but not ESCs (average of three replicates). (C) Box plot demonstrating that MaTaDa recapitulates the dynamic binding of OCT4 to EpiLC and ESC-specific sites. *P*<1e-5 unpaired, unequal variance *t*-test. Boxplots represent median, first and third quartiles (hinges) and 1.5× interquartile range extending from the hinges (whiskers).

several pluripotency-associated factors. A second motif (Fig. S7E; E=2.0e-2) similar to the SOX motif is present at all 698 binding sites. Although SOX proteins have been shown to recognise similar motifs, they become restricted in their binding patterns through interactions with specific co-factors (Hou et al., 2017; Kondoh and Kamachi, 2010). Our data suggest that PRDM14 binds at novel, previously unidentified, PGCLC-specific enhancers.

Next, we analysed genes associated with PRDM14 binding in ESCs and PGCLCs. We found that ESC-enriched PRDM14 target genes are implicated in the regulation of embryonic development and negative regulation of cell differentiation (Fig. S7D), which includes genes differentially expressed in PRDM14 mutants, such as *Fgfr1*, *Fgfr2* and *Dnmt3b* (Costello et al., 2011; Grabole et al., 2013).

By comparing the transcriptomes of EpiLCs and PGCLCs (Sasaki et al., 2015), we found that 445/2889 differentially expressed genes (15%) are direct targets of PRDM14 (*P*<1e-16, Fig. 5E,F). Key PGC specification genes, including *Tfap2c*, *Dppa3*, *Nr5a2* and *Esrrb* were both bound by PRDM14 and upregulated in PGCLCs, whereas EpiLC-associated, PRMD14-bound genes, including *Wnt8a*, *Otx2*, *Pou3f1* and *Dnmt3a* were downregulated. PRDM14 may thus play a key role in PGC specification by upregulating key reprogramming genes and repressing EpiLC genes.

PGCLC-enriched PRDM14 targets also included genes involved in EGFR and MAPK signalling (Fig. S7D). Notably, inhibition of the MAPK pathway is sufficient to upregulate key PGC markers context dependently (Kimura et al., 2014).

Most intriguingly, in both ESCs and PGCLCs, PRDM14 binding is enriched in the vicinity of genes that function in neuronal development, cell migration and cell morphology (Fig. S7D). Consistent with this finding, genes involved in neurogenesis become induced upon depletion of PRDM14 in ESCs (Yamaji et al., 2013). In the developing mouse embryo, PGCs migrate from the area of specification into the endoderm epithelium of hindgut and colonise the developing genital ridge at E10.5 (Anderson et al., 2000; Clark and Eddy, 1975). To investigate a potential function of PRDM14 in PGC migration, we focused on *Wnt5a* and *Tnc*, which are significantly bound by PRDM14 in PGCLCs (Fig. 5A, Fig. S8A) and have been implicated in cell migration (Chawengsaksophak et al., 2012; Nishio et al., 2005). Expression of these genes is extremely low in ESCs and EpiLCs (Fig. S8B,C). Upon PGCLC induction, *Wnt5a* and *Tnc* expression become induced only in those cells that fail to acquire the PGC identity, whereas PRDM14-expressing PGCLCs (Fig. S7D) repress both migration-associated genes. In contrast, the Wnt5a receptors *Ror2* and *Fzd5* are not repressed in specified PGCLCs (Fig. S8E,F) (Ishikawa et al., 2001; Niehrs, 2012). Together, our results suggest that PRDM14 plays a direct role in controlling PGCLC migration towards the genital ridge.

DISCUSSION

The study of genome-wide interactions of transcription factors and chromatin has been largely dominated by a single biochemical technique: ChIP. ChIP-seq experiments typically require millions of cells, precluding the identification of transcription factor targets in small populations of cells, including stem cells *in vivo*. In addition, the accuracy in detection of binding sites, as determined by ChIP, is difficult to assess due to the paucity of alternative techniques. Here, we have developed mammalian targeted DamID (MaTaDa), which enables transcription factor occupancy to be profiled with high sensitivity in a temporally and spatially controlled manner. MaTaDa overcomes the potential toxicity associated with expression of high levels of Dam methylase and avoids potential artefacts caused by overexpression of a transcription factor. Notably, expression of Dam-PRDM14 in PGCLCs did not result in adverse effects on cell growth or PGCLC specification efficiency compared with parental cells or cell lines expressing Dam alone (Fig. S2B-D). A key finding is that MaTaDa can reveal the binding sites of master regulators of pluripotency in as few as 10,000 cells and potentially even fewer. Although we used pure cell populations in this study, our previous work in *Drosophila* shows that TaDa profiles can be generated from a tiny proportion of cells in complex heterogeneous

Development (2018) 145, dev170209. doi:10.1242/dev.170209

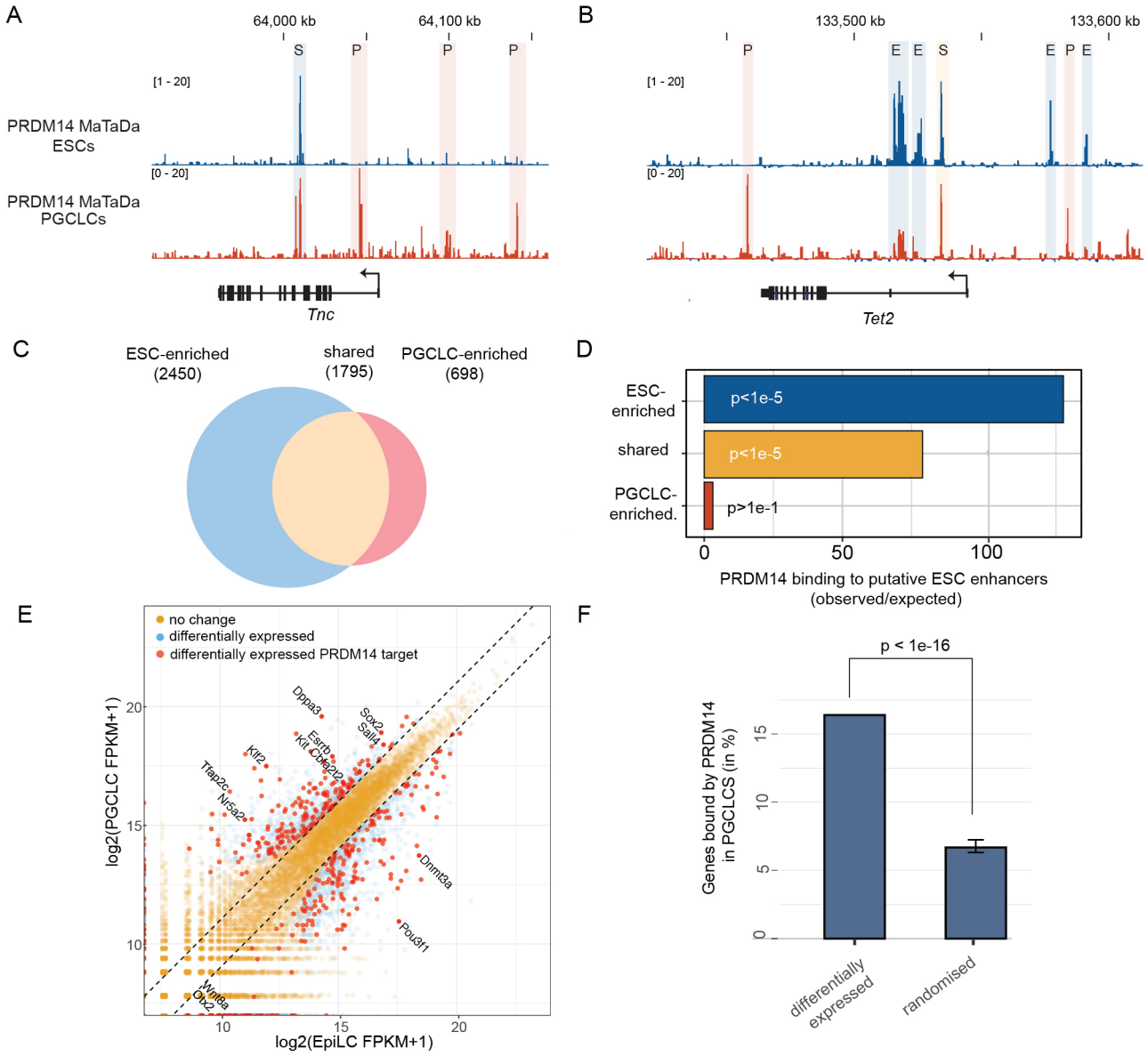

Fig. 5. Distinct chromatin association of PRDM14 in ESCs and PGCLCs. (A,B) Genome browser view of PGCLC-enriched PRDM14 occupancy in ESCs and PGCLCs (both average of three replicates) at the *Tnc* (A) and the *Tet2* (B) loci. Shared, PGCLC-enriched and ESC-enriched sites (S, P and E, respectively) are shown. (C) PRDM14-bound regions are subdivided into ESC-enriched (>2 fold), shared in ESCs/PGCLCs and PGCLC-enriched (>2 fold). (D) Overlap between ESC-defined enhancers and genomic loci occupied by PRDM14 in ESCs (ESC-enriched), PGCLCs (PGCLC-enriched) or both (shared). *P*-values are calculated using the Genomic Association Test (Heger et al., 2013). (E) Comparison between genes differentially expressed during PGCLC specification and genes bound by PRDM14 in PGCLCs. Gene expression is plotted in log2 FPKM (fragments per kilobase mapped reads). (F) A large proportion of PGCLC-PRDM14 targets are differentially expressed between EpiLCs and PGCLCs. *P*-values are calculated by an empirical test based on a normal distribution.

tissues (Southall et al., 2013). This will allow MaTaDa to be applied *in vivo* to assess chromatin occupancy in rare and previously inaccessible cell populations.

Despite the overall similarity of binding profiles, MaTaDa and ChIP-seq results differed in notable ways. Many binding sites were detected using one method but not the other, despite the use of identical cell lines and culture conditions (Fig. 2, Figs S4, S5). In this respect, it was striking that the degree of overlap between ChIP-seq peaks and peaks obtained by MaTaDa was similar for OCT4 and PRDM14. This indicates that the incongruities between MaTaDa and ChIP might derive from fundamental differences between the two techniques. Distinguishing which approach is a more accurate reflection of the binding of a transcription factor is not straightforward. Indeed, no studies have thus far systematically analysed the similarities between alternative methods for determining the genome-wide occupancy of transcription factors. Although ChIP signals were strongly enriched over MaTaDa peaks (Fig. 2, Fig. S4), the intensity of peaks did not always correlate (Fig. 2H, Fig. S4H). An underlying assumption of ChIP is that regions that are most strongly detected are most commonly or strongly bound, and are thus ‘key targets’. However, additional factors, such as binding kinetics, steric effects of fusion proteins (Ramialison et al., 2017), accessibility of antibody-targeted epitopes, and bias in PCR amplification, cross-linking or sonication could all affect the observed MaTaDa and ChIP signal intensities (Meyer and Liu, 2014). MaTaDa and ChIP may therefore

Development (2018) 145, dev170209. doi:10.1242/dev.170209

represent complementary approaches for understanding transcription factor-genome interactions.

We designed MaTaDa to take advantage of the large collection of Cre-expressing constructs, cell lines and model organisms for targeted expression *in vivo*. As has been demonstrated in *Drosophila* (Cheetham and Brand, 2018; Marshall and Brand, 2017; Southall et al., 2013; Aughey et al., 2018), the generation of MaTaDa transgenic animals for key transcription factors, lncRNAs, chromatin complexes and RNA polymerase II will permit the characterisation of the molecular landscape of gene regulation in almost any cell type. Analysing the interactions between transcription factors and enhancers in small and pure populations of stem cells *in vivo* will be of vital importance for an increased understanding of the transcriptional control of development. Although we observe some leakiness *in vitro* (Fig. 1F), our system can clearly identify rearrangements in transcription factor occupancy.

Taking advantage of the high sensitivity of MaTaDa we were able for the first time to monitor PRDM14-chromatin association during the course of PGCLC specification. PRDM14 DNA binding in PGCLCs and ESCs differs significantly in location and intensity, which may be a consequence of co-factor availability or the distinct epigenetic state of each cell type. Although 2i ESCs reside in an epigenetic state characterised by low abundance of repressive chromatin marks, such as DNA methylation or H3K9me2, PGCLCs are specified from EpiLCs, which are associated with elevated levels of repressive chromatin modifications (Leitch et al., 2013; Zylicz et al., 2015). Hence, the epigenetic environment established in EpiLCs could potentially restrict PRDM14 binding during early PGCLC specification. We found that 15% of genes differentially expressed in the transition between EpiLCs and PGCLCs are direct PRDM14 targets. Interestingly, PRDM14 may function not only as a repressor but also as an activator, as it was bound at both up- and downregulated genes. Interestingly, PGCLC-enriched PRDM14 binding sites did not correspond to predicted ESC enhancers (Chen et al., 2012) and may identify novel PGCLC-specific enhancers. Several of these presumptive enhancers regulate components of the MAPK pathway. In mice, PGCs are specified in the proximity of cells undergoing mesodermal differentiation. Consequently, inductive signals like Wnt and BMP that initiate PGC specification in the postimplantation embryo are also involved in defining the mesoderm linage (Behringer et al., 1999; Winnier et al., 1995). Furthermore, inhibition of the MAPK pathway during mesodermal differentiation results in the upregulation of PGC marker genes (Kimura et al., 2014). This suggests that PRDM14 functions during early murine PGC specification by inhibiting the MAPK signalling pathway and thereby prevents establishment of the mesodermal cell fate.

Interestingly, we find that PRDM14 binds in the vicinity of genes associated with cell migration (Fig. S7D), such as *Wnt5a* and *Tnc*, which are most significantly bound by PRDM14 in PGCLCs (Fig. 5A, Fig. S8A). Wnt5a and its receptor Ror2 function in PGC migration. Loss of Wnt5a signalling strongly impairs PGC migration to the genital ridge (Chawengsaksophak et al., 2012; Laird et al., 2011). Here, we find that *Wnt5a* and *Tnc* but not Wnt5a receptors are repressed in PRDM14-expressing PGCLCs, which suggests that Wnt5a is secreted from somatic cells to promote directed PGC migration, while Wnt5a repression in PGCs may prevent autocrine stimulation

Ectopic expression of PRDM14 has been linked to several types of cancer, such as lymphatic leukaemia, lung carcinoma and most prominently breast cancer (Dettman et al., 2011; Nishikawa et al., 2007; Taniguchi et al., 2017; Zhang et al., 2013). A comprehensive understanding of Wnt5a function in breast cancer remains elusive; however, there is evidence that decreased Wnt5a expression in these tumours is associated with a poorer prognosis (Zeng et al., 2016). Hence, a link between PRMD14, Wnt5a and cell migration might be of clinical relevance. We conclude that MaTaDa holds great promise for the *in vivo* analysis of transcription factor and chromatin protein interactions during development and disease.

MATERIALS AND METHODS

Embryonic stem cell culture

E14tg2a embryonic stem cells (ESCs) were cultured in serum/LIF medium for maintenance [GMEM (Invitrogen; 21710-025), 10% FBS (Invitrogen 10270-106), 1% non essential amino acid (Invitrogen; 11140), 1 mM sodium pyruvate (Invitrogen; 1130-070), 2 mM L-glutamine (Invitrogen; 25030-024), 1% penicillin-streptomycin (Invitrogen; 15140-22), 0.2% 2-mercaptoethanol (Invitrogen; 21985-023) and 0.1% LIF (obtained from CSCR Cambridge)]. Cells were grown on gelatine-coated cell culture dishes (ThermoFisher) and passaged by dissociating to ESC colonies with TrypLE (Invitrogen 12604-021). For experiments, ESCs were grown in 2i/LIF medium [N2B27 medium, 1 µM PD0325901, 3 µM CHIR99021 (Stemgent) and 0.1% LIF (obtained from CSCR Cambridge)] on fibronectin-coated dishes (17 µg ml^{-1}; Millipore) for at least four passages.

Induction of epiblast-like cells (EpiLCs) and primordial germ cell-like cells (PGCLCs)

EpiLCs were induced as described previously (Hayashi et al., 2011). In brief, 2i ESCs were differentiated into EpilCs by treatment with FGF2 and activing A for 40 h. Subsequently, PGCLCs specification was induced by a cytokine cocktail consisting of BMP4 (0.5 µg/ml), BMP8 (0.5 µg/ml), SCF (0.1 µg/ml), EGF (0.05 µg/ml) and LIF (1×; made by CSCR Cambridge) (Fig. S5A). The induction of PGCLC specification by cytokines is inefficient and, hence, FACS purification of successfully specified cells is required. Here, we made use of a dual reporter ESC line harbouring stable integrations of GFP and tdTomato in the endogenous *Dppa3* and *Esg1* loci, respectively (Hackett et al. 2018 preprint). Whereas expression of Dppa3/GFP is used to identify specifying PGCLCs, high Esg1/tdTomato expression marks undesired cell types such as ESCs or EpiLCs.

piggyBac transposition

PBase (2.5 µg), 0.5 µg CreER plasmid and 2.5 µg of MaTaDa plasmid were diluted in 50 µl of Gibco Opti-MEM Media. Lipofectamine 2000 (5 µl, Invitrogen) was diluted in 45 µl of OptiMem and mixed with the plasmid solution and vortexed. The solution was incubated for 10 min at room temperature and then added to E14 ESCs in culture. The cells were incubated for 4-6 h at 37°C in 5% CO_2. The media was then changed. Transfection efficiency usually ranges between 3-30%. Cells were selected using 25 µg/ml zeocine for 7 days after transfection.

Flow cytometry

Embryoids were washed with PBS (Gibco), dissociated by incubation in 10 mM tissue culture grade EDTA (Invitrogen) for 3-5 min at 37°C and subjected to FACS using the Sony SH800S Cell Sorter. FACS data were analysed using the Flowjo software.

qRT-PCR

Total RNA was isolated from 20,000-200,000 cells using the RNeasy Mini Kit (Qiagen) including an on-column DNase digest. cDNA was generated using the SuperScript III Reverse Transcriptase kit (ThermoFisher) and 250 ng random primer (Invitrogen) per reaction. The cDNA was quantified using the SYBR Select Master Mix (Applied Biosystems) and the QuantStudio 6 Flex Real-Time PCR System (ThermoFisher). PCR reaction mix and qPCR program were prepared according to manufacturer's instructions.

MaTaDa constructs

To clone the maTaDa construct PGK-LGL-Dam, LT3-Dam-Myc was amplified and inserted into a vector with the PGK promoter driving

Development (2018) 145, dev170209. doi:10.1242/dev.170209

expression of a floxed GFP cassette using Gibson assembly (Fig. S9A). The OCT4 CDS was amplified from mESC cDNA and inserted into PGK-LGL-Dam using the restriction enzymes *Bgl*II and *Not*I (Fig. S9B). PGK-LPL-Dam was constructed by replacing the floxed GFP cassette with a puromycin resistance cassette (Fig. S9C). PRDM14 was amplified from mESC cDNA and inserted into PGK-LPL-Dam with *Bgl*II and *Not*I (Fig. S9D).

DamID-seq

DamID-seq was performed as described previously (Marshall et al., 2016). Briefly, cells were dissociated with TrypLE, washed and counted. gDNA was extracted using the Qiagen QIAamp DNA Micro Kit. The DNA was then digested overnight at 37°C with *Dpn*I to cut methylated GATC sites (New England Biolabs) and purified with a QIAquick PCR Purification Kit. Adaptors were blunt-end ligated for 2 h at 16°C using T4 DNA ligase (New England Biolabs) and heat inactivated at 65°C for 20 min. The ligated DNA was then digested with *Dpn*II to cleave any unmethylated GATC sites (New England Biolabs) and purified with a 1:1 ratio of Seramag beads. Adaptor-ligated fragments were then amplified with MyTaq (Bioline) and PCR purified. The amplified DNA was then sonicated and digested with *Alw*I (New England Biolabs) to remove the adaptors, generating diverse DNA ends. The fragments were then prepared for Illumina sequencing according to the modified TruSeq protocol described by Marshall et al. (2017). All sequencing was performed as single end 50 bp reads generated by the Gurdon Institute NGS Core using an Illumina HiSeq 1500.

Published data acquisition

Sra files were acquired from the Gene Expression Omnibus (Clough and Barrett, 2016) via wget (v1.17.1) and converted with fastq-dump (v2.3.5) to fastq files or with abi-dump (v2.3.5) to csfasta and qual files for colorspace data.

Quality check

Quality check was performed for all files individually with FastQC (v0.11.4). Residual adaptor sequences and low quality bases were removed with cutadapt (v1.9.1) or TrimGalore (v0.4.5) when needed. Total and unique reads were summed to assess library size. The lengths of the reads was determined as additional quality check with awk (v4.1.3).

damidseq_pipeline

A file of GATC sites for GRCm38 genome was generated as gff file with gatc.track.maker.pl (see github.com/AHBrand-Lab). Analysis of fastq files from DamID experiments was performed with the damidseq pipeline script (Marshall and Brand, 2015) that maps reads to an indexed bowtie2 genome (i.e. GRCm38), bins into GATC fragments according to GATC sites and normalises reads against a Dam-only control. Binding intensities were quantile normalised across all replicates (i.e. across all bedgraph files) for the same experiment and subsequently averaged. Pearson correlation coefficients and R^2 values for comparisons of individual normalised replicates were calculated between pairs of bedgraph files in the RStudio environment with base functions (base, v3.4.3; RStudio, v1.1.423).

ChIP-seq mapping

Reads were mapped to the indexed mouse genome (mm10) with bowtie2 (v2.2.9) (Langmead and Salzberg, 2012) or optionally to the corresponding masked version, including only major chromosomes to improve data quality. Resultant sam files were converted to bam files, sorted and indexed with samtools (v1.3.1) (Li et al., 2009). Duplicates were removed with the MarkDuplicates picard tool (v.1.95) when needed. Total and unique mapped reads were counted with awk (v4.1.3) and bedtools (v2.25.0) (Quinlan and Hall, 2010).

Browser tracks and data visualisation

Reads were extended as well as binned, and resulting tracks converted with the bamCoverage deeptools command (v3.0.2). Files were converted into bw files with awk (v4.1.3) and bedGraphToBigWig (v4) or to tdf files with the Integrative Genomic Viewer (IGV) (Robinson et al., 2011). Data were visualised using IGV, with the midline for MaTaDa ratio tracks set at 1 and, for ChIP-seq, set at 0. Heatmaps were generated using Seqplots in the RStudio IDE (v1.12.0; Stempor and Ahringer, 2016).

ChIP-peak calling and quantification

Sorted bam-files for input (Buecker et al., 2014) or HA-/EGFP-flag samples (Ma et al., 2011; Yamaji et al., 2013) were merged with samtools (v1.3.1) to serve as combined control sample. Broad peaks were called with MACS2 (v2.1.0) for the individual bam files in comparison with the combined control sample with the following specifications: –keep-dup all –bw 300 –qvalue 0.05 –mfold 5 50 –broad –broad-cutoff 0.1. Peaks were called on the individual bam files for the experimental samples in comparison with the combined control sample. The number of significant peaks was read out at sequentially decreasing q-values [i.e. represented as '-log10(qvalue)' in line with MACS2]; peaks in accessory contigs and mitochondrial genome were filtered out. Residual peaks were sorted according to coordinates and data were converted into bed format. Reads were accumulated over peaks by intersecting bam file-derived read coordinates with peaks using bedtools (v2.25.0), summing the reads with awk (v4.1.3) and normalising them for library size and peak length. Data were prepared and plotted with tidy tools in the RStudio IDE as violin and scatterplots (i.e. ggplot2, v2.2.1; tibble, v1.4.2; tidyr, v0.8.0; readr 1.1.1; purr, v0.2.4; dplyr, v0.7.4; stringr, v1.3.0).

DamID-peak calling

bam files with extended reads for Dam only generated by the pipeline for every sample were merged and used as combined control for each individual Dam fusion sample during MACS2 (v2.1.0) peak calling. Additionally, peaks were called for a merged bam file consisting of all Dam fusion samples in comparison with the merged bam file for Dam only in line with peak calling for ChIP samples (i.e. merge versus merge).

Consensus peaks

Peaks were defined as reproducible across all replicates at a given q-value, when overlapping peaks from these biological replicates were consistently identified in more than 50% (Yang et al., 2014) of all cases (including the merge versus merge, e.g. 2 out of 3, 3 out of 4). Coordinates of consensus peaks were defined as the maximum area covered by all overlapping peaks, which prevents peak duplication.

DamID-peak quantification

DamID-binding intensities for identified peaks were aggregated by identifying all GATC fragments overlapping with the area of the peak, trimming the first and the last fragment to peak coordinates and summing the weighted scores associated with the fragments. Data were analysed and visualised in accordance with the corresponding ChIP datasets in RStudio (see 'ChIP-peak calling and quantification').

Common peaks

Peaks shared among ChIP and DamID or between experiments carried out using the same technique were identified by intersecting the corresponding peak collections with bedtools (v2.25.0) intersectBed. Coordinates of common peaks were defined, deduplicated and sorted similar to the methods used to generate consensus peaks. The extent of overlap between common peaks was evaluated depending on the q-value, which was either gradually changed for the sets of peaks from both techniques or changed for one dataset while keeping the q-value of the other dataset unchanged. The latter allows the evaluation of the recovery of peaks in the compared set, despite differing significance. Similarly, the distributions of common and individual peaks were determined by identifying the closest peak from the compared dataset to the summits of the investigated peak set dependent on the q-value with the closest tool of the bedtools suite (v2.25.0). These distributions were plotted as densities of peak numbers dependent on the distance to the reference peak summits with ggridges (v0.5.0) in RStudio.

Annotation

Peaks were annotated to overlapping genomic features or nearest gene, respectively (e.g. for intergenic/distal peaks), with the ChIPseeker-package

(v1.10.3) in the R-environment using annotations from TxDb.Mmusculus.UCSC.mm10.knownGene (v3.4.0) and gene IDs from org.Mm.eg.db (v3.4.0).

RNA-Seq

cutadapt (v1.9.1)-trimmed reads for RNA-seq from EpiLCs and PGCLCs (Sasaki et al., 2015) were reconverted from fastq to csfasta and qual file formats and Phred+33 scores translated into numeric Q scores. Reads were aligned and assembled with tophat (v2.0.14) (Kim et al., 2013), which used bowtie1 (v1.1.2) to map colorspace data using –no-coverage-search. Differentially expressed genes were identified with cufflinks (v2.2.0; i.e. –compatible-hits-norm –no-length-correction –library-type fr-secondstrand –max-mle-iterations 50,000), cuffmerge (v1.0.0) and cuffdiff (v2.2.0; i.e. –compatible-hits-norm –no-length-correction –library-type fr-secondstrand –max-mle-iterations 50,000 –frag-bias-correction –multi-read-correct) (Trapnell et al., 2012).

Significant genes were filtered by $q \leq 0.05$ and a fold-change of ≥ 2 (i.e. EpiLC-vs-d4BVSC), and overlapped with the list of annotated peaks associated with intergenic regions to identify differentially expressed genes associated with putative enhancer peaks. Similar numbers of genes were randomly sampled without replacement 10,000 times for empirically testing the enrichment of putative enhancer peaks with the associated genes by approximating a normal distribution.

Enrichment

Coordinates for promoters, exons, introns, and 5′ and 3′UTRs, as well as intergenic regions were retrieved for transcript and exon annotations from biomaRt (v2.30.0). A list of high probability enhancers was derived from (Chen et al., 2012). Enrichment analysis for ESC- and PGCLC-specific, as well as common peaks, sets [i.e. -log10(q-value)≥100, FC≥2] with gene features and enhancers was performed with gat (v1.2.2) (Heger et al., 2013) using 100,000 sampling iterations and mm10 chromosomes as workspace. Fold enrichment of genomic features (i.e. enhancers) or log2-fold enrichment (i.e. gene features) for the associations were displayed together with their respective q-values.

Motif detection

Motifs were detected *de novo* using the MEME suite program MEME and compared with known motifs using TOMTOM (Bailey et al., 2015). Enrichment of the PRDM14 motif was detected using AME (Bailey et al., 2015). Transcription factor motifs and overlap with chromatin marks, as well as DNaseI hypersensitivity sites in Oct4 peak sets, were screened with i-cisTarget (Imrichova et al., 2015).

Acknowledgements

We thank Kay Harnish from the Gurdon Institute NGS Core facility, Cambridge, UK for help with sequencing and Catherine Davidson for technical assistance.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: S.W.C., W.H.G., J.v.d.A., R.K., T.D.S., T.K., M.A.S., A.H.B.; Methodology: S.W.C., W.H.G., J.v.d.A., R.K.; Software: S.W.C., R.K.; Validation: S.W.C., W.H.G., J.v.d.A., R.K., A.H.B.; Formal analysis: S.W.C., W.H.G., J.v.d.A., R.K., A.H.B.; Resources: A.H.B.; Data curation: S.W.C.; Writing - original draft: S.W.C., W.H.G., J.v.d.A., R.K., A.H.B.; Writing - review & editing: S.W.C., W.H.G., J.v.d.A., R.K., A.H.B.; Supervision: M.A.S., A.H.B.; Project administration: A.H.B.; Funding acquisition: M.A.S., A.H.B.

Funding

This work was funded by a Wellcome Trust Senior Investigator Award (103792), a Wellcome Trust Program Grant (092545) and a Royal Society Darwin Trust Research Professorship to A.H.B. S.W.C. was funded by a Herchel Smith Fund Research Studentship. W.H.G. was supported by a European Molecular Biology Organization Long-term Fellowship (ALTF 263_2014). J.v.d.A. was supported by a European Molecular Biology Organization Long-term Fellowship (ALTF 1600_2014) and by a Wellcome Trust Postdoctoral Training Fellowship for Clinicians (105839). T.K. was supported by Postdoctoral Fellowships for Research Abroad, The Uehara Memorial Foundation Research Fellowship and Kanae Foundation for the Promotion of Medical Science. M.A.S. was supported by the Human Frontier Science Program (RGP0020/2012) and a Wellcome Trust Senior Investigator Award (096738). A.H.B. acknowledges core funding to the Gurdon Institute from the Wellcome Trust (092096) and Cancer Research UK (C6946/A14492). Deposited in PMC for release after 6 months.

Data availability

Generated datasets have been deposited in GEO under accession number GSE101971.

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.170209.supplemental

References

Anderson, R., Copeland, T. K., Schöler, H., Heasman, J. and Wylie, C. (2000). The onset of germ cell migration in the mouse embryo. *Mech. Dev.* **91**, 61-68.

Aughey, G. N. and Southall, T. D. (2016). Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 25-37.

Aughey, G. N., Estacio Gomez, A., Thomson, J., Yin, H. and Southall, T. D. (2018). CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. *eLife* **7**, e32341.

Bailey, T. L., Johnson, J., Grant, C. E. and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* **43**, W39-W49.

Behringer, R. R., Bradley, A., Liu, P., Wakamiya, M., Shea, M. J. and Albrecht, U. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat. Genet.* **22**, 361-365.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T. and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838-853.

Chawengsaksophak, K., Svingen, T., Ng, E. T., Epp, T., Spiller, C. M., Clark, C., Cooper, H. and Koopman, P. (2012). Loss of Wnt5a disrupts primordial germ cell migration and male sexual development in mice. *Biol. Reprod.* **86**, 1-12.

Cheetham, S. W. and Brand, A. H. (2018). RNA-DamID reveals cell-type-specific binding of roX RNAs at chromatin-entry sites. *Nat. Struct. Mol. Biol.* **25**, 109-114.

Chen, C., Morris, Q. and Mitchell, J. A. (2012). Enhancer identification in mouse embryonic stem cells using integrative modeling of chromatin and genomic features. *BMC Genomics* **13**, 152.

Clark, J. M. and Eddy, E. M. (1975). Fine structural observations on the origin and associations of primordial germ cells of the mouse. *Dev. Biol.* **47**, 136-155.

Clough, E. and Barrett, T. (2016). The gene expression omnibus database. *Method. Mol. Biol.* **1418**, 93-110.

Costello, I., Pimeisl, I.-M., Dräger, S., Bikoff, E. K., Robertson, E. J. and Arnold, S. J. (2011). The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nat. Cell Biol.* **13**, 1084-1091.

Dettman, E. J., Simko, S. J., Ayanga, B., Carofino, B. L., Margolin, J. F., Morse, H. C. and Justice, M. J. (2011). Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling common lymphoid progenitors. *Oncogene* **30**, 2859-2873.

Furey, T. S. (2012). ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nat. Rev. Genet.* **13**, 840-852.

Grabole, N., Tischler, J., Hackett, J. A., Kim, S., Tang, F., Leitch, H. G., Magnúsdóttir, E. and Surani, M. A. (2013). Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep.* **14**, 629-637.

Hackett, J. A., Huang, Y., Gunesdogan, U., Holm-Gretarsson, K., Kobayashi, T. and Surani, A. M. (2018). Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *bioRxiv*, doi:10.1101/269811.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532.

Heger, A., Webber, C., Goodson, M., Ponting, C. P. and Lunter, G. (2013). GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**, 2046-2048.

Hou, L., Srivastava, Y. and Jauch, R. (2017). Molecular basis for the genome engagement by Sox proteins. *Semin. Cell Dev. Biol.* **63**, 2-12.

Imrichová, H., Hulselmans, G., Atak, Z. K., Potier, D. and Aerts, S. (2015). I-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res.* **43**, W57-W64.

Ishikawa, T., Tamai, Y., Zorn, A. M., Yoshida, H., Seldin, M. F., Nishikawa, S. and Taketo, M. M. (2001). Mouse Wnt receptor gene Fzd5 is essential for yolk sac and placental angiogenesis. *Development* **128**, 25-33.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36.

Kimura, T., Kaga, Y., Ohta, H., Odamoto, M., Sekita, Y., Li, K., Yamano, N., Fujikawa, K., Isotani, A., Sasaki, N. et al. (2014). Induction of primordial germ cell-like cells from mouse embryonic stem cells by ERK signal inhibition. *Stem Cells* **32**, 2668-2678.

Kondoh, H. and Kamachi, Y. (2010). SOX-partner code for cell specification: regulatory target selection and underlying molecular mechanisms. *Int. J. Biochem. Cell Biol.* **42**, 391-399.
Koziol, M. J., Bradshaw, C. R., Allen, G. E., Costa, A. S. H., Frezza, C. and Gurdon, J. B. (2015). Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24-30.
Laird, D. J., Altshuler-Keylin, S., Kissner, M. D., Zhou, X. and Anderson, K. V. (2011). Ror2 enhances polarity and directional migration of primordial germ cells. *PLoS Genet.* **7**, e1002428.
Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359.
Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J. G., Smith, A. et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311-316.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. and Wysocka, J. (2011). Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat. Struct. Mol. Biol.* **18**, 120-127.
Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B. and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* **15**, 905-915.
Marshall, O. J. and Brand, A. H. (2015). Damidseq-pipeline: An automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **31**, 3371-3373.
Marshall, O. J. and Brand, A. H. (2017). Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. *Nat. Commun.* **8**, 2271.
Marshall, O. J., Southall, T. D., Cheetham, S. W. and Brand, A. H. (2016). Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat. Protoc.* **11**, 1586-1598.
Meyer, C. A. and Liu, X. S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.* **15**, 709-721.
Nady, N., Gupta, A., Ma, Z., Swigut, T., Koide, A., Koide, S. and Wysocka, J. (2015). ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife* **4**, e10150.
Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y. and Saitou, M. (2013). Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222-226.
Nichols, J. and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487-492.
Niehrs, C. (2012). The complex world of WNT receptor signalling. *Nat. Rev. Mol. Cell Biol.* **13**, 767-779.
Nishikawa, N., Toyota, M., Suzuki, H., Honma, T., Fujikane, T., Ohmura, T., Nishidate, T., Ohe-Toyota, M., Maruyama, R., Sonoda, T. et al. (2007). Gene amplification and overexpression of PRDM14 in breast cancers. *Cancer Res.* **67**, 9649-9657.
Nishio, T., Kawaguchi, S., Yamamoto, M., Iseda, T., Kawasaki, T. and Hase, T. (2005). Tenascin-C regulates proliferation and migration of cultured astrocytes in a scratch wound assay. *Neuroscience* **132**, 87-102.
Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T. and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571-584.
Otsuki, L., Cheetham, S. W. and Brand, A. H. (2014). Freedom of expression: cell-type-specific gene profiling. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 429-443.
Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.
Ramialison, M., Waardenberg, A. J., Schonrock, N., Doan, T., de Jong, D., Bouveret, R. and Harvey, R. P. (2017). Analysis of steric effects in DamID profiling of transcription factor target genes. *Genomics* **109**, 75-82.
Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24-26.
Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H. et al. (2015). Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* **17**, 178-194.
Southall, T. D., Gold, K. S., Egger, B., Davidson, C. M., Caygill, E. E., Marshall, O. J. and Brand, A. H. (2013). Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. *Dev. Cell* **26**, 101-112.
Stempor, P. and Ahringer, J. (2016). SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. *Wellcome Open Res.* **1**, 14.
Taniguchi, H., Hoshino, D., Moriya, C., Zembutsu, H., Nishiyama, N., Yamamoto, H., Kataoka, K. and Imai, K. (2017). Silencing PRDM14 expression by an innovative RNAi therapy inhibits stemness, tumorigenicity, and metastasis of breast cancer. *Oncotarget* **8**, 46856-46874.
Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562-578.
Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., Gnirke, A. and Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344-349.
van Steensel, B. and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424-428.
Vogel, M. J., Peric-Hupkes, D. and van Steensel, B. (2007). Detection of in vivo protein–DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467-1478.
Winnier, G., Blessing, M., Labosky, P. A. and Hogan, B. L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9**, 2105-2116.
Wu, T. P., Wang, T., Seetin, M. G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S. D., Mackintosh, S. G., Zhong, M. et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329-333.
Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y. and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat. Genet.* **40**, 1016-1022.
Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., Saitou, M. et al. (2013). PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* **12**, 368-382.
Yang, Y., Fear, J., Hu, J., Haecker, I., Zhou, L., Renne, R., Bloom, D. and McIntyre, L. M. (2014). Leveraging biological replicates to improve analysis in ChIP-seq experiments. *Comput. Struct. Biotechnol. J.* **9**, e201401002.
Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523.
Zeineddine, D., Hammoud, A. A., Mortada, M. and Boeuf, H. (2014). The Oct4 protein: more than a magic stemness marker. *Am. J. Stem Cells* **3**, 74-82.
Zeng, R., Huang, J., Zhong, M.-Z., Li, L., Yang, G., Liu, L., Wu, Y., Yao, X., Shi, J. and Wu, Z. (2016). Multiple roles of WNT5A in breast cancer. *Med. Sci. Monit.* **22**, 5058-5067.
Zhang, T., Meng, L., Dong, W., Shen, H., Zhang, S., Liu, Q. and Du, J. (2013). High expression of PRDM14 correlates with cell differentiation and is a novel prognostic marker in resected non-small cell lung cancer. *Med. Oncol.* **30**, 605.
Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J. et al. (2015). N6-methyladenine DNA modification in drosophila. *Cell* **161**, 893-906.
Zylicz, J. J., Dietmann, S., Günesdogan, U., Hackett, J. A., Cougot, D., Lee, C. and Surani, M. A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *Elife* **4**, e09571.