

# **ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation**

Nataliya Nady<sup>1,6</sup>, Ankit Gupta<sup>2,6</sup>, Ziyang Ma<sup>1</sup>, Tomasz Swigut<sup>1</sup>, Akiko Koide<sup>2</sup>, Shohei Koide<sup>2,7</sup>, Joanna Wysocka<sup>1,3,4,5,7</sup>

<sup>1</sup> Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>2</sup> Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois, USA

<sup>3</sup> Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>4</sup> Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>5</sup> Institute of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California 94305, USA

<sup>6</sup> these authors contributed equally

<sup>7</sup> co-corresponding authors

1    **Competing Interests Statement**

2    JW. and NN. declare no competing interests. SK and AK are inventors on a patent application filed by  
3    the University of Chicago that covers monobody library design (US 13/813,409).  
4    The monobodies described in this work are available from SK under a material transfer agreement with  
5    the University of Chicago.

6



## 7   **Abstract**

8   Prdm14 is a sequence-specific transcriptional regulator of embryonic stem cell (ESC) pluripotency and  
9   primordial germ cell (PGC) formation. It exerts its function, at least in part, through repressing genes  
10   associated with epigenetic modification and cell differentiation. Here, we show that this repressive  
11   function is mediated through an ETO-family co-repressor Mtgr1, which tightly binds to the pre-  
12   SET/SET domains of Prdm14 and co-occupies its genomic targets in mouse ESCs. We generated two  
13   monobodies, synthetic binding proteins, targeting the Prdm14 SET domain and demonstrate their  
14   utility, respectively, in facilitating crystallization and structure determination of the Prdm14-Mtgr1  
15   complex, or as genetically encoded inhibitor of the Prdm14-Mtgr1 interaction. Structure-guided point  
16   mutants and the monobody abrogated the Prdm14-Mtgr1 association and disrupted Prdm14's function  
17   in mESC gene expression and PGC formation *in vitro*. Altogether, our work uncovers the molecular  
18   mechanism underlying Prdm14-mediated repression and provides renewable reagents for studying and  
19   controlling Prdm14 functions.

20

## Introduction

Prdm14 is a sequence-specific transcriptional regulator that plays key roles in promoting primordial germ cell (PGC) specification and safeguarding pluripotency of embryonic stem cells (ESCs) (Nakaki and Saitou, 2014). During mouse embryogenesis, Prdm14 is expressed in preimplantation embryos, where its asymmetric expression promotes allocation of cells toward the pluripotent inner cell mass (ICM) fate (Burton et al., 2013; Nakaki and Saitou, 2014). Prdm14 expression ceases in postimplantation epiblast cells and their differentiated progeny. However, during PGC specification from the epiblast, cells reacquire many transcriptional and epigenetic characteristics of the preimplantation state, and Prdm14 is re-expressed along with several other pluripotency-associated factors (reviewed in (Magnúsdóttir and Surani, 2014; Saitou et al., 2012). Loss of Prdm14 in mice results in sterility associated with early germ cell deficiency, as cells fated to become PGCs fail to reacquire expression of key pluripotency factors and undergo epigenetic reprogramming (Yamaji et al., 2008). Furthermore, overexpression of Prdm14 in epiblast-like cells (EpiLCs) is sufficient to induce PGCs *in vitro* (albeit with low frequency), suggesting a central role of Prdm14 in the mouse PGC regulatory network (Magnúsdóttir et al., 2013; Nakaki et al., 2013). Furthermore, Prdm14 is repressed in normal somatic tissues, but it is aberrantly reactivated in human malignancies of various tissue origin, including leukemias and lymphomas, breast, testicular, and lung cancers (Carofino et al., 2013; Dettman et al., 2011; Nishikawa et al., 2007; Ruark et al., 2013; Zhang et al., 2013).

Given the poor accessibility and transient nature of preimplantation embryo cells and PGCs *in vivo*, mechanistic studies of Prdm14 function in early development have been chiefly conducted in the context of mouse ESCs (mESCs). These cells represent a so-called 'naïve' pluripotent state, thought to resemble preimplantation embryo ICM and serve as a useful system for understanding early cell fate decisions (Nichols and Smith, 2009). Loss of Prdm14 destabilizes mESCs and sensitizes them to differentiation stimuli, leading to acquisition of alternative embryonic states, such as the postimplantation epiblast state or extraembryonic endoderm state, and to eventual depletion of the

naïve cell subpopulation (Ma et al., 2011; Yamaji et al., 2013). The differentiation in Prdm14<sup>-/-</sup> cells is thought to result from upregulation of signaling pathways such as FGFR pathway and by widespread DNA hypermethylation (Grabole et al., 2013; Hackett et al., 2013; Leitch et al., 2013). Indeed, genome-wide Prdm14 occupancy studies by ChIP-seq suggest that these are direct effects of the loss of Prdm14, as Prdm14 occupies and represses the regulatory elements of genes involved in FGFR signaling and *de novo* DNA methylation (Leitch et al., 2013; Ma et al., 2011; Magnúsdóttir et al., 2013; Yamaji et al., 2013). Nonetheless, Prdm14<sup>-/-</sup> mESCs can be maintained indefinitely under 2i conditions (Grabole et al., 2013; Payer et al., 2013; Yamaji et al., 2013), in which differentiation stimuli, including FGFR signaling, are chemically inhibited, providing an opportunity to study the early effects of Prdm14 deficiency upon release from such inhibition.

Although the cellular and molecular phenotypes associated with loss of Prdm14 in mESCs have been well characterized, much less is known about molecular mechanisms and partners through which Prdm14 acts. As a member of the PRDM family, Prdm14 contains both a zinc finger array, responsible for sequence-dependent DNA binding (Ma et al., 2011), and a PR domain that is related to the SET domain (Su(var)3-9, Enhancer-of-zeste and Trithorax) (Fog et al., 2012; Hohenauer and Moore, 2012). Many SET domains harbor methyltransferase activity for either histone or non-histone substrates (Del Rizzo and Trievel, 2011). However, to date no enzymatic activity has been reported for Prdm14, and interestingly multiple members of the PRDM family appear to be catalytically inactive. Instead, candidate-based co-immunoprecipitation studies implicated Polycomb complex PRC2 as a mediator of Prdm14-dependent repression (Chan et al., 2013; Yamaji et al., 2013). Nonetheless, it remains unclear whether PRC2 is a major or auxiliary partner of Prdm14, and what other molecular players are important for Prdm14's function.

To address these questions, we used an unbiased biochemical approach to uncover major Prdm14-associated proteins in mESCs. We identified an ETO-family corepressor, Mtgr1 (Myeloid translocation gene related 1, a.k.a. Mtg8r, Cbfa2t2, and Zmynd3), as a direct, stoichiometric partner of

Prdm14. We demonstrate that Mtgr1 co-occupies Prdm14 target loci, and its deletion in mESCs results in phenotypes and gene expression defects highly similar to those observed upon loss of Prdm14. Moreover, Mtgr1 knockout cells show impaired induction of PGC-like cells *in vitro*. To further facilitate studies of the Prdm14-Mtgr1 complex, we mapped interaction domains and developed multiple synthetic binding proteins, termed monobodies, that specifically recognize the SET domain of Prdm14 in a manner independent of, or alternatively, competitive with Mtgr1. Taking advantage of the stabilizing effect of one such monobody, we obtained a crystal structure of the Prdm14-Mtgr1 complex, revealing an extensive interface and electrostatic interactions mediating the association of the two proteins. Furthermore, structure-guided mutagenesis of the interface and the use of an inhibitory monobody demonstrated the function of the complex in safeguarding pluripotency and PGC-like cell induction. Altogether, we report a multi-disciplinary study that advances our understanding of Prdm14 function, identifying Mtgr1 as the major partner of Prdm14 in its roles in pluripotency and PGC induction, and providing the community with renewable, genetically encoded reagents that can be used both *in vitro* and *in vivo* to study and control Prdm14 function in development and malignancy.

## Results

### Identification of Mtgr1 as a novel Prdm14 partner

To identify Prdm14 partners in an unbiased manner, we employed a two-step immunopurification strategy [FLAG followed by HA (FH)] from a previously described clonal mESC transgenic line stably expressing tagged FH-Prdm14 (Ma et al., 2011). Examination of recovered proteins by SDS-PAGE and silver staining revealed two major polypeptides that were present in similar quantities in the FH-Prdm14 purifications, but not in the control immunoprecipitates (Figure 1A). These polypeptides were subsequently identified by mass spectrometry as Prdm14 and Mtgr1 (Figure 1A, Suppl. Table 1), the latter of which is one of the three members of the ETO family of co-repressors (Davis et al., 2003). Mass spectrometry analysis also identified additional polypeptides enriched

96 uniquely in the FH-Prdm14 purifications, including the other two ETO proteins Mtg8 and Mtg16, their  
97 known repressive complex partners Tbl1/Tblr1 and HDACs, as well as Brg1 complex components and  
98 Oct4, among others (Suppl. Table 1). Of note, we did not detect components of the Polycomb complex  
99 PRC2 (Chan et al., 2013; Yamaji et al., 2013).

100 Mtgr1 was the major polypeptide identified in our analysis and its stoichiometric recovery in  
101 our purifications indicated it might represent a strong and direct partner of Prdm14. To confirm this, we  
102 first verified Prdm14 and Mtgr1 association using reciprocal co-immunoprecipitations from mESC  
103 nuclear extracts (Figure 1B). Next, we mapped the minimal regions within Prdm14 and Mtgr1 that  
104 were required for the interaction by overexpressing differentially-tagged proteins (V5-Mtgr1 and FH-  
105 Prdm14) in HEK293 cells, followed by IP-Western analysis (Figure 1C, D). This strategy revealed that  
106 the NHR1 (nervy homology region 1) domain of Mtgr1 was necessary and sufficient for the interaction  
107 with Prdm14, whereas both Prdm14 SET domain and the region directly preceding it (pre-SET) were  
108 important for efficient binding to Mtgr1 (Figure 1C, D). To quantify the strength of the Prdm14-Mtgr1  
109 interaction we next expressed and purified recombinant proteins corresponding to the NHR1 domain of  
110 Mtgr1 (residues 98-206) and pre-SET+SET domains of Prdm14 (residues 184-373), and performed a  
111 bead-based binding assay in reciprocal orientations (Nishikori et al., 2012). The obtained binding  
112 measurements yielded a dissociation constant ( $K_D$ ) in the low nanomolar range (Figure 1E, F), which is  
113 consistent with a robust, direct interaction between the two proteins. On the other hand, the binding of  
114 Prdm14 SET domain alone (residues 232-373) to Mtgr1 NHR1 domain was barely detectable (Figure  
115 1G), further supporting that both pre-SET and SET domains are required for the high affinity  
116 interaction with Mtgr1.

117 Altogether, our approach identified an ETO protein Mtgr1 as a novel, direct partner of Prdm14  
118 in mESCs. While the ETO proteins, especially Mtg8 (a.k.a. ETO), have been studied in the context of  
119 acute myeloid leukemias (AML) (reviewed in Hatlen et al., 2012), their function in ESCs and during  
120 early embryogenesis has not been explored. Notably, all three ETO family members have the capacity

121 to interact with Prdm14 (not shown), but the high expression of Mtgr1 in mESCs as compared with  
122 Mtg8 and Mtg16 (Figure 1–figure supplement 1) likely accounts for the preferential recovery of Mtgr1  
123 in our experiments and suggests that this family member may be most relevant in the context of  
124 mESCs. We therefore proceeded to explore the functional significance of the Prdm14-Mtgr1 interaction  
125 in mESC biology.

126

## 127 **Prdm14 and Mtgr1 co-occupy genomic targets**

128 Prdm14 is a sequence-dependent DNA-binding protein that binds many genomic loci in  
129 mESCs, corresponding primarily to distal regulatory elements, whereas ETO proteins do not contain  
130 domains implicated in direct DNA sequence recognition (Rossetti et al., 2004, 2008). To examine if  
131 Mtgr1 is brought to genomic targets occupied by Prdm14, we performed Mtgr1 chromatin  
132 immunoprecipitations coupled with high throughput DNA sequencing (ChIP-seq) from wt mESCs, *FH-*  
133 *Prdm14* overexpressing mESCs, and as a control for antibody specificity, *Mtgr1*<sup>-/-</sup> mESCs (generation  
134 of which is described in more detail later), cultured for 5 days under serum+LIF conditions. In parallel,  
135 we profiled Prdm14 occupancy by performing ChIP-seq analysis from *FH-Prdm14* cells, using an anti-  
136 HA antibody due to the unavailability of ChIP-grade Prdm14 antibodies. Overall, we identified ~ 8000  
137 Mtgr1 peaks present in both *FH-Prdm14* and wt mESCs, but absent in *Mtgr1*<sup>-/-</sup> mESCs. These bound  
138 sites include loci known to be occupied and repressed by Prdm14 (e.g. near *Prdm14*, *Dnmt3b*, *Wnt8a*,  
139 *Peg10*, and targets of the FGFR pathway *Fgfr2* and *Shc1*; Figure 2A).

140 Generally, the genomic occupancies of Mtgr1 and Prdm14 were well correlated (correlation coefficient  
141 ~0.9, Figure 2–figure supplement 1A) and we have not been able to detect a substantial class of  
142 Prdm14-bound sites devoid of Mtgr1 occupancy (Figure 2B). Not surprisingly, the Prdm14 and Mtgr1  
143 sites shared common functional ontologies, with enrichment for processes involved in embryonic  
144 development and cell differentiation (Figure 2–figure supplement 1B). Furthermore, the most highly  
145 enriched DNA sequence motif at Mtgr1-bound sites corresponded to the previously identified Prdm14

motif (Figure 2C). Interestingly, we noted that at many targets Mtgr1 binding was enhanced by Prdm14 overexpression (see tracks in Figure 2A, compare wt and *FH-Prdm14* ESC). This observation prompted us to quantitatively compare Mtgr1 ChIP-seq enrichments in wt ESCs and *FH-Prdm14* cells that are characterized by ~5 fold overexpression of Prdm14. We observed that Mtgr1 enrichments were higher in *FH-Prdm14* than in wt ESCs at most target sites, consistent with Prdm14-mediated recruitment of Mtgr1 to chromatin (Figure 2D). However, we also noticed that a subset of Mtgr1 sites was bound more weakly in *FH-Prdm14* cells than in wt ESCs (Figure 2D, examples shown in Figure 2-figure supplement 2A). The major distinction between these two populations was presence of the Prdm14 sequence motif and Prdm14 occupancy at the sites where Mtgr1 binding was enhanced by Prdm14 overexpression, and lack of the Prdm14 sequence motif with low/no Prdm14 occupancy at the sites where Mtgr1 binding was diminished by Prdm14 overexpression (Figure 2D). Of note, at the Prdm14 motif-lacking sites, the most enriched sequence motifs corresponded to helix-loop-helix transcription factor recognition sites, suggesting that a TF from this family may be involved in mediating Mtgr1 binding at these sites (Figure 2 - figure supplement 2C). Regardless, our results indicate that Prdm14 is sufficient to augment interaction of Mtgr1 with chromatin at its cognate binding sites and, at high levels, redirect it away from the motif-lacking sites. Thus, Prdm14 might be a limiting factor for Mtgr1 recruitment to chromatin. To test this notion further, we performed Mtgr1 ChIP-seq analysis from Prdm14<sup>-/-</sup> ESCs and generated average signal profiles at Prdm14 motif-containing and Prdm14 motif-lacking sites across all our Mtgr1 ChIP-seq datasets. We observed that at Prdm14 motif-containing sites, Mtgr1 binding is increased in FH-Prdm14 overexpressing cells and diminished (but not completely abrogated) in Prdm14<sup>-/-</sup> cells (Figure 2 - figure supplement 2B, left panel). On the other hand, at Prdm14 motif-lacking sites Mtgr1 binding is depleted by FH-Prdm14 overexpression, but it is also moderately affected in Prdm14<sup>-/-</sup> cells despite low/no Prdm14 binding at these sites, suggesting an indirect effect (Figure 2 - figure supplement 2B, right panel). Altogether, these results are consistent with the Mtgr1 genomic occupancy being sensitive to the Prdm14 dosage (either loss or gain) at the

171 Prdm14-motif containing sites. However, these results also demonstrate that even in the absence of  
172 Prdm14, some Mtgr1 binding remains at the motif-containing sites, suggesting partial redundancies in  
173 the recruitment mechanisms.

174

### 175 **Loss of Mtgr1 phenocopies requirement for Prdm14 in safeguarding pluripotency**

176 Prdm14 has well-characterized roles in pluripotency and PGC formation, and if Mtgr1 is a key  
177 mediator of Prdm14's functions then the loss of Mtgr1 should impact these processes in a similar  
178 manner. To test this hypothesis, we used CRISPR-Cas9 with a guide RNA targeting the third exon of  
179 the *Mtgr1* gene to generate *Mtgr1*<sup>-/-</sup> mESCs, and verified the presence of the homozygous deletions and  
180 loss of the Mtgr1 protein in the three clonal lines selected for further analysis (Figure 3–figure  
181 supplement 1). As a reference for comparison, we also isolated and characterized two *Prdm14*<sup>-/-</sup> mESC  
182 lines by targeting the second exon of the *Prdm14* gene (Figure 3–figure supplement 2). Moreover, we  
183 reconstituted each of the *Mtgr1*<sup>-/-</sup> and *Prdm14*<sup>-/-</sup> cell lines with *FH-Mtgr1* or *FH-Prdm14* cDNA,  
184 respectively, to generate 'rescue' cell lines and ensure specificity of the observed phenotypes. All  
185 aforementioned cell lines were isolated and maintained under the serum-free 2i+LIF conditions, in  
186 which the major differentiation cues are inhibited, and which support self-renewal even in the absence  
187 of Prdm14 (Grabole et al., 2013; Yamaji et al., 2013). After being transferred into standard serum+LIF  
188 growth conditions, the *Mtgr1*<sup>-/-</sup> lines exhibited changes in morphological appearance with less compact  
189 colonies, diminished cell-cell interactions and cell flattening, as previously reported for loss of Prdm14  
190 in mESCs and reproduced here with our *Prdm14*<sup>-/-</sup> lines (Ma et al., 2011; Yamaji et al., 2013) (Figure  
191 3–figure supplement 3). These features were not observed in wt mESCs or after rescue with the  
192 respective protein constructs (Figure 3–figure supplement 3).

193 Loss of Prdm14 has been shown to sensitize mESC to differentiation stimuli, resulting in  
194 upregulation of genes associated with epiblast and extraembryonic endoderm fates (Ma et al., 2011;  
195 Yamaji et al., 2013). To examine if these molecular phenotypes are also observed upon loss of Mtgr1,



196 we conducted RNA-seq transcriptome analyses from wt mESCs, *Prdm14*<sup>-/-</sup> and *Mtgr1*<sup>-/-</sup> cell lines, and  
197 their respective rescue lines after transfer from 2i+LIF to serum+LIF conditions. As seen in *Prdm14*<sup>-/-</sup>  
198 mESCs, *Mtgr1*<sup>-/-</sup> mESCs showed upregulation of epiblast (e.g. *Fgf5*, *Dnmt3b*, *Oct6*, *Wnt8a*) and  
199 extraembryonic endoderm (e.g. *Krt19*, *Sparc*, *H19*, *Fgfr2*) markers, and downregulation of naïve  
200 pluripotency genes (e.g. *Esrrb*, *Zfp42*, *Tbx3*, *Tet2*), as compared with either wild type or *FH-Mtgr1*  
201 rescue mESCs (Figure 3A, Figure 3—figure supplement 4A).

202 Next, we identified genes showing the most variable expression across our datasets and we  
203 visualized their expression changes in each of our RNA-seq datasets as a heatmap (Figure 3B). Most of  
204 the differentially expressed genes were concordantly upregulated in *Prdm14*<sup>-/-</sup> and *Mtgr1*<sup>-/-</sup> cells, as  
205 compared with wt mESCs, in agreement with the proposed function of the Prdm14-Mtgr1 complex in  
206 gene repression. A more systematic comparison of all genes upregulated at least twofold upon loss of  
207 either Prdm14 or Mtgr1 revealed that while indeed, the majority of genes that are upregulated in either  
208 knockout are upregulated in both (Figure 3—figure supplement 4B, purple dots), a subset of transcripts  
209 is preferentially affected only in one of the knockouts (red or blue dots).

210 Importantly, in the *FH-Prdm14* or *FH-Mtgr1* reconstituted knockout cells the derepression  
211 defects were rescued (Figure 3B). Interestingly, while *FH-Mtgr1* cells showed expression patterns  
212 highly similar to that of wt mESCs cultured in serum, *FH-Prdm14* cells were more similar to the  
213 mESCs grown under 2i +LIF, despite being cultured in serum at the time of analysis (Figure 3B).  
214 Indeed, many of the expression differences observed between wt mESC grown in 2i versus serum were  
215 recapitulated in *FH-Prdm14* mESCs (Figure 3B). Of note, a subset of transcripts was upregulated only  
216 in *FH-Prdm14* cells; many of those genes represent markers of the so-called 2-cell (2C) state (Amano  
217 et al., 2013; Dan et al., 2013; Macfarlan et al., 2012).

218 In our analysis of variably expressed genes, *Prdm14*<sup>-/-</sup> and *Mtgr1*<sup>-/-</sup> cell lines clustered together,  
219 but separately from the respective rescue lines and wt mESCs (Figure 3B). These observations were  
220 further confirmed by the global comparisons of transcriptomes with principal component analysis

(PCA), in which *Prdm14*<sup>-/-</sup> and *Mtgr1*<sup>-/-</sup> cells were closest to each other and clustered away from the remaining cell lines (Figure 3C). Additionally, the PCA analysis corroborated higher similarity of *FH-Mtgr1* cells to wt mESCs grown in serum+LIF, and *FH-Prdm14* cells to wt mESCs grown under 2i+LIF conditions (Figure 3C). Given that: (i) *FH-Prdm14* cell lines in our study express Prdm14 at levels ~5-6 higher than wt ESCs, (ii) Prdm14 has an autonomous DNA-binding activity, and (iii) Prdm14 overexpression can augment Mtgr1 recruitment to the target genes (as shown in Figure 2), we propose that expression changes observed in *FH-Prdm14* cell lines are associated with more robust repression of differentiation programs by Prdm14-Mtgr1 complex as compared with wt cells and consequently, with the stabilization of the naïve pluripotency program. Similar gain-of-function effects are not observed in *FH-Mtgr1* cells, likely because Mtgr1 lacks the autonomous ability to access its genomic targets and, at least in mESCs, Prdm14 is limiting for its chromatin association. Altogether, our data uncover the function of Mtgr1 in safeguarding mESC pluripotency and demonstrate that loss of Mtgr1 phenocopies gene expression defects associated with Prdm14 deletion.

### **Mtgr1 is required for PGC specification *in vitro***

Prdm14 is critical for the specification of PGCs from the post-implantation epiblast cells (Magnúsdóttir and Surani, 2014; Yamaji et al., 2008). To address if Mtgr1 also plays a role in PGC formation, we used a previously established *in vitro* model, in which naïve mESCs are first differentiated to a primed, post-implantation epiblast-like state (mEpiLCs), from which PGC-like cells (mPGC-LCs) are then induced via addition of various cytokines (Hayashi and Saitou, 2013; Hayashi et al., 2011). The mPGC-LCs formation is monitored with the fluorescent reporter, *Stella*:GFP and quantified by FACS analysis (Figure 3D). *In vitro* derived mPGC-LCs have been shown to be competent to differentiate to sperm/oocytes upon transplantation and produce viable, fertile offspring and this *in vitro* differentiation is therefore considered a useful tool to study mechanisms underlying PGC specification (Hayashi et al., 2011; Nakaki et al., 2013). To examine the role of Prdm14 and

246 Mtgr1 in the context of this model, we derived *Prdm14*<sup>-/-</sup> and *Mtgr1*<sup>-/-</sup> mESCs in the *Stella*:GFP reporter  
247 background; this reporter recapitulates endogenous *Stella* induction which occurs during the PGC  
248 formation (Payer et al., 2006).

249 In agreement with previous reports, differentiation of *Stella*:GFP mESCs consistently produced  
250 mPGC-LCs with an efficiency of ~7-8% on day 6 of differentiation, while low levels (~1%) of GFP-  
251 positive cells were detected in mESCs and further diminished upon mEpiLC formation (Figure 3E). In  
252 contrast, both *Prdm14*<sup>-/-</sup> and *Mtgr1*<sup>-/-</sup> mESCs showed significantly decreased efficiency of mPGC-LCs  
253 formation (Figure 3F and G, Figure 3—figure supplement 5). These defects were rescued by the re-  
254 introduction of *FH-Prdm14* or *FH-Mtgr1*, respectively (Figure 3F and G, Figure 3—figure supplement  
255 5). These data suggest that Mtgr1, like Prdm14, is important for mouse PGC establishment.  
256 Significantly, loss of Mtgr1 does not result in general block in differentiation, as germ layer markers  
257 are expressed at comparable levels in embryoid bodies induced from *Mtgr1*<sup>-/-</sup> versus wt ESCs (Figure 3  
258 - figure supplement 5).

259

## 260 **Generation of renewable monobody reagents to study and inhibit Prdm14-Mtgr1 complex**

261 To develop new tools for understanding the function of Prdm14 and to aid structure  
262 determination, we generated designer binding proteins termed ‘monobodies’ recognizing the SET  
263 domain of human and mouse Prdm14, as a part of a larger project aimed at developing new reagents for  
264 controlling epigenetic regulatory proteins. Monobodies are small binding proteins (~10 kDa) generated  
265 from combinatorial phage-display libraries built on the antibody-like scaffold of the tenth human  
266 fibronectin type III domain (FN3; Figure 4A) (Koide et al., 1998, 2012b). Monobodies can recognize  
267 their targets with high affinity and specificity and have a strong tendency to recognize functional  
268 binding sites in their target molecules including clefts and planar surfaces, and thus they often are  
269 potent inhibitors (Koide et al., 2012a; Sha et al., 2013; Wojcik et al., 2010). Furthermore, unlike  
270 antibodies whose folding depends on disulfide bond formation, monobodies are cysteine-free and thus

functional when expressed under reducing environments such as the nucleus and cytoplasm. These attributes make monobodies particularly attractive as genetically encoded intracellular inhibitors.

We isolated Prdm14-binding monobodies from two combinatorial phage display libraries termed the ‘loop’ library and the ‘side’ library (Koide et al., 2012a). Following phage display selection, we performed gene shuffling for affinity maturation and further selection in the yeast display format to isolate clones with high affinity. We identified a total of 12 clones that bound to human PRDM14 (hPRDM14) with  $K_D < 100$  nM as measured by yeast surface display (Figure 4A; Figure 4–figure supplement 1). Among them we identified two clones, Mb(hPRDM14\_S4) and Mb(hPRDM14\_S14), that bound hPRDM14 (residues 238-487) at least five fold stronger than the closest homologues, hPRDM6 (residues 194-405) and hPRDM12 (residues 60-229) (Figure 4C). We will use shorthand names, Mb(S4) and Mb(S14), for referring to them hereafter for brevity. Additionally, these two monobodies show comparable binding to the mouse homologue of hPRDM14, Prdm14. Of note, hPRDM14 was able to substitute for the mouse Prdm14 in rescue of the Prdm14<sup>-/-</sup> ESC defects, suggesting both biochemical and biological conservation of function between mouse and humans (Figure 4–figure supplement 2).

We produced these two monobodies as purified proteins for further characterization. Consistent with analysis using yeast surface display, these purified monobodies showed high affinity with  $K_D < 50$  nM to both hPRDM14 and Prdm14 in bead-based assays (Figure 4B). We then examined whether these monobodies inhibited the interaction of Prdm14 with Mtgr1. Mb(S14) potently competed against the binding of Mtgr1 to Prdm14 but Mb(S4) did not (Figure 4D). This result suggests that the two monobodies bind to distinct surfaces of Prdm14 and the epitope for Mb(S14) overlaps with and therefore occludes the Mtgr1-binding surface (Figure 4D).

We next tested if these monobodies can immunoprecipitate Prdm14 from mESC lysates. Mb(S4) captured Prdm14 and co-immunoprecipitated vast majority of Mtgr1 from the *FH-Prdm14* cell extracts (Figure 4E). On the other hand, Mb(S14) captured lower levels of Prdm14 and its associated

296 Mtgr1, in agreement with its competition for the same binding surface as Mtgr1 (Figure 4E). Since  
297 PRC2 complex has been previously reported to associate with Prdm14 (Chan et al., 2013; Payer et al.,  
298 2013; Yamaji et al., 2013), we also looked for the presence of Suz12, a PRC2 component. We did not  
299 detect immunoprecipitated Suz12 in the elution fraction for either of the two monobodies. Next, we  
300 used monobodies to precipitate endogenous Prdm14 from wt ESCs. Immunoblot analysis with  $\alpha$ -Mtgr1  
301 antibody showed that Mb(S4) monobody, which does not disrupt Prdm14-Mtgr1 interaction, recovered  
302 endogenous Mtgr1 (and was able to deplete most of it from the extract, Figure 4—figure supplement 3A  
303 and B). In addition, we performed Prdm14 Mb(S4)-precipitation/mass spec analysis from wt ESCs,  
304 which readily detected Prdm14- and Mtgr1-originating peptides, but did not recover any other PRDM  
305 proteins confirming the high specificity of this monobody (Figure 4—figure supplement 3B and 3C).  
306 Common to monobodies generated to recognize native, folded proteins, neither Mb(S4) nor Mb(S14)  
307 detected denatured Prdm14 in immunoblotting (not shown). Overall, we report here the generation of  
308 the first recombinant affinity reagents targeting two distinct sites of Prdm14 SET domain.

309

### 310 **Monobody- and fusion-assisted crystal structure determination of the Prdm14-Mtgr1 complex**

311 To understand how Prdm14 and Mtgr1 interact at the atomic level, we attempted crystallization  
312 of the Prdm14-Mtgr1 complex. However, aggregation of both proteins resulted in low yields of the  
313 complex suitable for crystallization. To overcome this problem, we designed a fusion construct of  
314 Prdm14 and Mtgr1 in which the two proteins were linked with a ten-residue linker (GSSGSSGS), a  
315 common strategy for stabilizing heterodimers (Ernst et al., 2014; Kobe et al., 2015; Reddy Chichili et  
316 al., 2013; Zhou et al., 2015). To confirm that the linker did not distort the Prdm14-Mtgr1 complex, we  
317 performed a series of experiments. The fusion protein had the same retention time on size-exclusion  
318 chromatography as the unlinked complex (Figure 5—figure supplement 1A), indicating that the linker  
319 did not alter the stoichiometry of the complex. We then compared the fusion and unlinked complex  
320 using solution NMR spectroscopy. Most of the cross peaks in the HSQC spectrum of  $^{15}\text{N}$ -Prdm14 in

321 complex with unlabeled Mtgr1 (where we observe signals only from  $^{15}\text{N}$ -Prdm14) overlapped with  
322 those in the HSQC spectrum of  $^{15}\text{N}$ -labeled Prdm14-linker-Mtgr1 (where we observe signals from the  
323 entire fusion protein including both Prdm14 and Mtgr1) (Figure 5- figure supplement 1B). The large  
324 number of overlapping peaks in the two spectra strongly suggests that the Prdm14 protein takes on  
325 nearly identical average conformation in the unlinked complex and in the fusion protein (Figure 5-  
326 figure supplement 1B). Furthermore, the fusion protein and the unlinked complex had the same affinity  
327 to Mb(S4), indicating that the linker did not distort the Prdm14 epitope for the monobody (Figure 5-  
328 figure supplement 1C). This fusion construct allowed us to overcome the aggregation problem, but it  
329 still yielded no crystals in crystallization trials using over 500 conditions.

330 We then used Mb(S4), the monobody that did not inhibit the Prdm14-Mtgr1 interaction, as a  
331 crystallization chaperone. Monobodies, like antibody fragments, often facilitate the crystallization of  
332 otherwise recalcitrant systems (Koide, 2009; Stockbridge et al., 2015). The addition of Mb(S4) readily  
333 led to crystallization of Prdm14-linker-Mtgr1, and we determined its structure to a resolution of 3.06 Å  
334 through single wavelength anomalous diffraction (SAD) phasing using selenomethionine-labeled  
335 crystals (Figure 5A; Table 1; Figure 5-figure supplement 2A). The crystallized complex had two  
336 Prdm14-linker-Mtgr1/Mb(S4) complexes in the asymmetric unit. As expected, the monobody bound  
337 exclusively to Prdm14, burying 604 Å<sup>2</sup> surface areas, a similar interface size to other monobody/target  
338 complexes (Figure 5 - figure supplement 2B, 2E) (Gilbreth et al., 2008; Wojcik et al., 2010). In the  
339 crystal, monobody-monobody interactions facilitated crystal contacts via face-to-face interactions of  
340 the β-sheet surfaces (not involved in Prdm14 interaction), illustrating the importance of Mb(S4) as a  
341 crystallization chaperone for this complex (Figure 5-figure supplement 2D).

342 The Prdm14 SET domain (residues 240-356) in the crystal structure is flanked by pre-SET and  
343 post-SET regions. The crystallization construct includes residues 184-239 that precede the SET domain  
344 (pre-SET) and residues 357-373 following the SET domain (post-SET). Unlike many SET domain-  
345 containing proteins, Prdm14 does not have cysteine-rich Zn finger domains adjacent to the SET

346 domain, commonly termed pre-SET and post-SET domains. Thus, in the absence of well-defined  
347 domains, we refer to these adjacent segments as pre-SET and post-SET regions. The crystal structure of  
348 the mouse Prdm14 SET domain is very similar to that of the hPRDM12 SET domain, the closest  
349 human homologue of Prdm14 (PDB ID 3EP0; C $\alpha$  RMSD=0.99; Figure 5B). The Prdm14 SET domain  
350 in our structure has a total of nine beta-strands ( $\beta$ 1- $\beta$ 9) arranged in three antiparallel beta-sheets with a  
351 short  $3_{10}$  helix ( $\eta$ 1) inserted between  $\beta$ 6 and  $\beta$ 7. The pre-SET region in our construct has a short helix at  
352 the N-terminus followed by a long structurally disordered region a part of which (residues 217-239) has  
353 no detectable electron density even at 0.5  $\sigma$  contour levels (2Fo-Fc). The residues that constitute the  
354 post-SET region at the C-terminus to the SET domain are arranged in an antiparallel beta sheet ( $\beta$ 10  
355 and  $\beta$ 11). Overall, the structural features of the PR/SET domain in Prdm14 show no major differences  
356 with other PR/SET domains in Prdm proteins.

357 The Mtgr1 NHR1 domain (also called the TAFH domain) is highly conserved in the ETO  
358 family. The Mtgr1 NHR1 domain in the crystal structure contains four well-defined  $\alpha$ -helices arranged  
359 in a bundle ( $\alpha$ A- $\alpha$ D; Figure 5C). Currently, three NMR structures for the NHR1 domain of human  
360 MTG8, another member of the ETO family, are available (Park et al., 2009; Plevin et al., 2006; Wei et  
361 al., 2007). Mtgr1 in the crystal structure has almost identical topology as the average solution NMR  
362 structure of the NHR1 domain from MTG8 in complex with a stabilizing peptide (PDB ID 2KNH; C $\alpha$   
363 RMSD=1.77 Å; Figure 5C). Similar structural features and high sequence identity for the NHR1  
364 domain in ETO proteins explain the pull-down of Mtg8 and Mtg16 in the FH-Prdm14 purifications  
365 (Figure 5F (left); Figure 1A; Suppl. Table 1).

### 366 367 **Mtgr1 contacts with the SET domain and pre-SET region of Prdm14**

368 The Mtgr1-Prdm14 interaction interface buries 2180 Å<sup>2</sup>, a large interface but still within the  
369 observed range for high-affinity protein-protein interaction interfaces with a low nanomolar  $K_D$  value  
370 (Lo Conte et al., 1999). We observe that Mtgr1 interacts with both SET domain and pre-SET region of

371 Prdm14 (Figure 5A), as was predicted from our earlier interaction mapping co-immunoprecipitation  
372 experiments (Figure 1C). The interactions between the Prdm14 SET domain and Mtgr1, mediated  
373 primarily by  $\alpha$ A and  $\alpha$ D, contribute 65% of the total buried surface area. The substantial interaction  
374 between the pre-SET region and Mtgr1 (35% of the interface) rationalizes the importance of the pre-  
375 SET region in the Prdm14-Mtgr1 interaction (Figure 1C).

376         Several features in the interface are notable. Arg105 located at the N-terminal end of Mtgr1  $\alpha$ A  
377 sits in a pocket formed by Ser290, Met292, Cys319 and Tyr339 of Prdm14 and its guanidinium moiety  
378 makes hydrogen bonds with the side chains of all of these residues (Figure 5D, Figure 5 - figure  
379 supplement 2C). Mutating Mtgr1 Arg105 to Asp resulted in a loss of detectable binding (Figure 5E),  
380 and replacing Prdm14 Tyr339 of the pocket with Arg substantially reduced binding (Figure 5E). In  
381 addition, Lys109 in Mtgr1  $\alpha$ A interacts with Glu294, Tyr355 and Pro366 of Prdm14 SET domain.  
382 Lys109 in Mtgr1 forms a salt bridge with Glu294 from  $\beta$ 5 of the Prdm14 SET domain (Figure 5D).  
383 Mutating either of these residues led to a loss of detectable binding (Figure 5E). Remarkably, charge  
384 reversal of this ionic interaction, i.e. the combination of Mtgr1 K109E and Prdm14 E294K restored  
385 binding, in agreement with the salt bridge formation across the binding interface in solution (Figure 5E  
386 middle). These mutation experiments together support the authenticity of the binding interface  
387 observed in the crystal structure and identified critical electrostatic interactions.

388         The Mtgr1 interaction with Prdm14 pre-SET region involves residues from Mtgr1 helix  $\alpha$ A,  $\alpha$ B  
389 and  $\alpha$ D (Figure 5A). Several Leu and Val residues from the helix in the pre-SET region contribute to  
390 hydrophobic interactions with residues in Mtgr1 helices (Figure 5F). In addition Mtgr1 Asn141 forms  
391 hydrogen bonds with residues Phe185 and Phe187 from the helix in pre-SET region. These results  
392 further reinforce the observation that Prdm14 utilizes both pre-SET region and the SET domain to  
393 interact with Mtgr1.

394



## 395 **Association of Prdm14 with Mtgr1 is required for mESC maintenance and PGC-LC formation**

396        Given strong association between Prdm14 and Mtgr1, as well as phenotypic similarities upon  
397 loss of either protein, we next examined whether the lack of Prdm14-Mtgr1 interaction would have a  
398 biological effect on mESC gene expression and on the efficiency of PGC-LC induction. To this end, we  
399 first confirmed that the point mutations designed based on the crystal structure (Figure 5D and 5E) also  
400 disrupted the interaction between the full-length proteins in cells, and that the combined charge reversal  
401 mutations restored the interaction (Figure 6A). Then we reconstituted *Stella*:GFP Prdm14<sup>-/-</sup> mESCs  
402 with cDNAs encoding either wt or mutant (E294K or Y339R) FH-Prdm14 and confirmed that all three  
403 proteins were expressed at similar levels (Figure 6B). Next, we used RNA-seq to compare gene  
404 expression patterns of these mESCs after transition from 2i+LIF to serum+LIF culture (Figure 6C). In  
405 the Prdm14<sup>-/-</sup> cells reconstituted with mutant Prdm14, we noted elevated expression of genes associated  
406 with differentiation to epiblast and extraembryonic endoderm, and diminished expression of naïve  
407 pluripotency genes (with an exception of *Tet2*), similar to attenuated expression patterns (albeit not to  
408 the same degree) as we observed upon loss of Prdm14 (Figure 6C, compare to Figure 3A). Principal  
409 component and clustering based on differential gene expression further support the notion that Prdm14  
410 E294K lines show hypomorphic expression profile, which falls in between the Prdm14<sup>-/-</sup> ESCs and  
411 those reconstituted with wt Prdm14 (Figure 6—figure supplement 1). Finally, the efficiency of mPGC-  
412 LCs formation from mESC reconstituted with Prdm14 mutants (E294K, Y339R) was significantly  
413 diminished compared to cells rescued with the wt Prdm14 to almost the same degree as in *Prdm14*<sup>-/-</sup>  
414 cells (Figure 6D).

415        Our results thus demonstrate that association of Prdm14 with Mtgr1 is required for mediating its  
416 functions in pluripotency and germ cell formation. Given that one of the monobodies we developed,  
417 Mb(S14), binds to Prdm14 competitively with Mtgr1, we hypothesized that this reagent can be utilized  
418 to inhibit Prdm14 function in living cells or organisms in a highly controlled manner. To provide a  
419 proof-of-principle for such strategy, we engineered piggyBac doxycycline-inducible constructs

420 encoding mCherry-Mb(S14) fusion or, as a control, mCherry alone, and introduced them into  
421 *Stella*:GFP reporter mESCs (Figure 6E). Next, we induced mPGC-LC formation from mESCs in the  
422 absence or presence of doxycycline (added during the induction of mPGC-LCs from mEpiLCs) to  
423 activate monobody expression. We observed consistent reduction in PGC-LC formation efficiency in  
424 cells expressing the mCherry-Mb(S14) fusion protein, as compared with the same cell population  
425 without addition of doxycycline or to cells expressing mCherry alone (Figure 6F, Figure 6-figure  
426 supplement 1). A more moderate effect of the monobody (as compared with the Prdm14 E294K  
427 mutation, Figure 6D) is probably due to the fact that this monobody needs to compete against the high-  
428 affinity interaction between Prdm14 and Mtgr1. In addition, we noticed a short half-life of the  
429 mCherry-Mb(S14) fusion protein (data not shown), which may contribute to the moderate effect but  
430 could also facilitate inhibition with high temporal resolution in the future. Thus, Mb(S14) represents a  
431 novel tool that can be utilized to perturb Prdm14 function in living cells during a dynamic biological  
432 process such as the PGC induction and can be further modified in the future for addressing particular  
433 questions using standard protein-engineering technologies.

434

## 435 **Discussion**

436 Our study identified the ETO family co-repressor Mtgr1 as a new regulator of mESC identity,  
437 which facilitates molecular functions of Prdm14 through direct binding to its pre-SET/SET region.  
438 Therefore, although no evidence has been previously found for the catalytic activity of the Prdm14 SET  
439 domain, our data demonstrate that this domain is nonetheless essential for Prdm14 function by  
440 mediating the interaction with its key partner, Mtgr1. Prdm14 and Mtgr1 co-occupy distal regulatory  
441 regions of many target genes linked to differentiation, DNA methylation and chromatin modification,  
442 consistent with their tight interaction. Loss of either protein results in upregulation of a subset of these  
443 target genes, and a gradual loss of mESC self-renewal. However, similar to what has been observed  
444 previously for Prdm14, Mtgr1 is not required for maintenance of mESC under 2i+LIF. This can be

445 explained by the fact that the Fgf/Erk pathway, a major signaling cascade driving both epiblast and  
446 extraembryonic endoderm differentiation (to which loss of Prdm14/Mtgr1 sensitizes cells), is inhibited  
447 under these conditions (Nichols and Smith, 2009).

448 Concordant upregulation of gene transcripts in *Prdm14*<sup>-/-</sup> or *Mtgr1*<sup>-/-</sup> cells as compared with wt  
449 mESCs supports function of the Prdm14-Mtgr1 complex in gene silencing and agrees with the reported  
450 role of Mtgr1 as an HDAC-recruiting co-repressor (Rossetti et al., 2004). Notably, we detected HDACs  
451 1-3 in our Prdm14 immunoprecipitates, and HDAC3 and other NCoR1 repressive complex components  
452 in our purifications with the Mb(S4) monobody, suggesting that the Prdm14-Mtgr1-dependent  
453 repression may indeed be facilitated by histone deacetylation and that Mtgr1 is required for recruiting  
454 HDACs to Prdm14-binding loci. Furthermore, although a subset of genomic sites occupied by Mtgr1  
455 occurs at Prdm14 motif-lacking sites, the high similarity of transcriptional changes observed upon loss  
456 of either protein suggests that the major impact of Mtgr1 on gene expression and cell identity of  
457 mESCs is in the context of its association with Prdm14. Moreover, our data suggest that Prdm14 can  
458 not only guide, but through change in its levels, quantitatively tune the degree of interaction of  
459 ubiquitously expressed Mtgr1 with chromatin.

460 Our results from the *in vitro* PGC-LC formation model strongly suggest that in addition to its  
461 role in mESCs, Mtgr1 is also a critical mediator of the Prdm14 function in germline development.  
462 Three lines of evidence support this notion: (i) Mtgr1 deletion, (ii) single point mutations in Prdm14  
463 disrupting association with Mtgr1, and (iii) expression of a monobody targeting the Prdm14-Mtgr1  
464 interaction surface, all of which hinder PGC-LC induction *in vitro*. It is therefore surprising that *Mtgr1*  
465 mouse knockout strain has been reported as viable and fertile (Amann et al., 2005). However, it  
466 remains unclear whether this strain represents a true loss-of-function, because in the mouse targeting  
467 strategy the first six exons are preserved (encoding amino acids 1-316 which span the Prdm14  
468 interaction region and NHR2 domain involved in dimerization and association with mSin3/HDAC  
469 complex) (Rossetti et al., 2004). In light of our findings and aforementioned caveats, we suggest that

470 the *in vivo* function of Mtgr1 in the germline should be revisited. It is also possible, however, that  
471 requirement for Mtgr1 in germ cell development *in vivo* can be fully or partially compensated by the  
472 other ETO proteins. Because residues that form the Prdm14 binding interface in Mtgr1 are conserved  
473 among the ETO family members, we expect that the other ETO proteins would bind directly to the  
474 Prdm14 pre-SET/SET regions and could therefore compensate for the absence of Mtgr1 function.

475 Our study provides novel insights into how pre-SET and SET regions might mediate high  
476 affinity protein-protein interactions. While many structures of the catalytic SET domains have been  
477 obtained with their substrates (typically, histone tails), to the best of our knowledge our study  
478 represents first structural analysis of the SET domain acting as a module for a high affinity protein-  
479 protein interaction. Interestingly, a comparison of our structure with that of the Prdm9-AdoHcy-  
480 histone peptide complex indicates that the surface of the Prdm14 pre-SET and SET regions engaged in  
481 interaction with Mtgr1 overlaps with surfaces other Prdm proteins use for binding to their histone  
482 peptide substrate (Figure 5—figure supplement 3B; (Wu et al., 2013)). Indeed, Mtgr1 Lys109 is in close  
483 proximity to Tyr355 that corresponds to the catalytic Tyr based on the consensus SET domain sequence  
484 (Smith and Denu, 2009). In Prdm9, the catalytic tyrosine Tyr356 along with Tyr276 and Tyr341 form  
485 the Lys4me2 binding pocket and are critical for catalytic activity (Wu et al., 2013). In Prdm14, the side  
486 chain of Tyr355 flips over (with respect to the conformation of Tyr357 in Prdm9) and interacts with  
487 K109 of Mtgr1. The location of K109 is distinct from that of Lys4me2 in Prdm9 (Figure 5—figure  
488 supplement 3C). In addition, the high affinity of the Prdm14-Mtgr1 interaction strongly suggests that  
489 Mtgr1 would be a poor substrate, given that a substrate needs to be released after catalysis for efficient  
490 enzyme reaction. We also note that His211 and Ala212 in the pre-SET region of Prdm14 occupy the S-  
491 adenosyl-L-methionine (SAM)-binding site (Figure 5—figure supplement 3A) and thus, at least in the  
492 context of the presented structure, preclude the binding of this cofactor necessary for methylation.  
493 Taken together, our results suggest that Mtgr1 Lys109 is unlikely to be an actual substrate for Prdm14-  
494 mediated methylation. Consistent with this notion, radioactive *in vitro* methyltransferase assays with

495 recombinant Prdm14 and Mtgr1 proteins, their respective interaction mutants, as well as histone  
496 substrates, all failed to yield methyltransferase activity (not shown). However, we cannot exclude a  
497 possibility that under presently unknown conditions, such activity can ultimately be found. Regardless,  
498 our comparisons suggest that substrate-binding surface can in some SET domain proteins be coopted  
499 for mediating high affinity protein-protein interactions, which may provide a molecular explanation as  
500 to why such surfaces are typically highly conserved even in SET domain proteins with no apparent  
501 catalytic activity.

502       Lastly, aberrant reactivation of the *PRDM14* locus is associated with a variety of human  
503 cancers, and mice overexpressing Prdm14 in blood cells develop early-onset T-cell acute lymphoblastic  
504 leukemia (T-ALL) (Carofino et al., 2013). We speculate that the oncogenic function of Prdm14 may be  
505 mediated by the formation of a complex with Mtgr1, which is broadly expressed and thus readily  
506 available for association in many different organ systems, or perhaps with other ETO proteins using the  
507 same interface. If this proves to be the case, inhibition of the Mtgr1-interaction surface on Prdm14  
508 could be an attractive target for therapy as Prdm14 expression is restricted under non-pathological  
509 conditions to the preimplantation embryo and PGCs, reducing the risk of off-target effects on normal  
510 somatic tissues. Thus, the monobodies directed to Prdm14 generated in this study will be powerful  
511 tools for testing the "druggability" of the Prdm14-Mtgr1 interaction. Indeed, we have introduced  
512 monobodies into CML cells and demonstrated the potential druggability of a domain interface in Bcr-  
513 Abl (Grebien et al., 2011). In addition, the crystal structure shows that Mb(S4) binds to a distinct  
514 surface of Prdm14. Because monobodies usually bind to functional sites on target proteins, we  
515 hypothesize that the Mb(S4) epitope may also be important for Prdm14 function, which will be a  
516 subject of future research. We emphasize that the genetically encoded monobodies are portable tools, as  
517 they can be readily introduced into different cells via transfection or viral transduction. This attribute  
518 should facilitate investigation of Prdm14 functions in diverse contexts.

519

## 520 **Materials and Methods**

### 521 **Cell culture, line derivation and embryoid body formation**

522 Stable lines expressing tagged Prdm14 were established from single colonies by transducing  
523 LF2 mESCs with FH-Prdm14 pTrip lentivirus for 24 h, followed by selection with neomycin as  
524 described previously (Ma et al., 2011); these lines were used for immunoprecipitations followed by  
525 mass spectrometry experiments. *Stella*:GFP cells were a gift from A. Surani and were used for all other  
526 experiments and further genetic manipulations, unless indicated otherwise. This line contains a  
527 transgene spanning 10kb upstream of the *Stella* transcriptional start site, exon1, intron1, and part of  
528 exon2, followed by *eGFP* fused in-frame and SV40 polyadenylation sequence. *Stella*:GFP line acts as  
529 a transcriptional reporter and has been previously shown to faithfully recapitulate endogenous *Stella*  
530 expression in mouse and mark PGCs as early as E7.5 (Payer et al., 2006). Stable double reporter lines  
531 (*Stella*:GFP and mCherry) were created by transfecting piggyBac Tet-On expression plasmid controlled  
532 by rtTA and doxycycline (mCherry alone or mCherry fused to Mb(S14)) with transposase and selected  
533 with blasticidin for 7 days. Single clones were picked and expanded further.

534 For maintenance all mouse ESC lines were grown in so-called “2i+LIF” medium that is serum-  
535 free N2B27-based medium supplemented with MEK inhibitor PD0325901 (0.8  $\mu$ M) and GSK3 $\beta$   
536 inhibitor CHIR99021 (3.3  $\mu$ M) in tissue culture (TC) dishes pretreated with 7.5  $\mu$ g/ml polyL-ornithine  
537 (Sigma) and 5  $\mu$ g/ml laminine (BD) (Hayashi and Saitou, 2013; Hayashi et al., 2011). For ChIP-seq,  
538 RNA-seq and immunoprecipitation experiments *Stella*:GFP mESC lines and FH-Prdm14 derivatives  
539 were cultured for 5 days in feeder-free conditions in DMEM-high glucose medium (DMEM/high  
540 glucose; HyClone) containing 15% (v/v) FBS, 1 mM glutamine, penicillin/streptomycin, 0.1mM  
541 nonessential amino acids, 0.1 mM 2-mercaptoethanol and supplemented with LIF (serum+LIF  
542 conditions).

543 To form embryoid bodies, *Stella*:GFP mESCs line of interest was maintained in a 15 cm Petri  
544 dish with serum culture medium without LIF, to allow aggregation as a hanging drop with 400 cells per

545 drop (360 drops per mESC line). Cells were cultured for 4d and half were collected for further analysis  
546 by RT-qPCR, while the other half was transferred into non-adherent petri dish and allowed to grow for  
547 another 4d (8d total) before RT-qPCR analysis.

548

#### 549 **EpiLC and PGC-LC differentiation**

550 To induce mEpiLC differentiation, mESC were washed with PBS, trypsinized, and strained. A  
551 total of about 100,000 cells per one well of 12-well plate were plated on TC dishes pretreated with  
552 5 µg/ml Fibronectin (Millipore) in N2B27-based medium supplemented with 1% KSR (Invitrogen) and  
553 12 µg/ml bFGF (Peprotech). The mPGCLCs were induced similarly to what has been described  
554 previously for 6 days (Hayashi and Saitou, 2013; Hayashi et al., 2011). Specifically, 1000-2000  
555 mEpiLC cells were aggregated in a hanging drop in a serum-free medium (GMEM, Invitrogen)  
556 supplemented with 15% KSR, 0.1mM NEAA, 1mM sodium pyruvate, 0.1mM 2-mercaptoethanol,  
557 penicillin/streptomycin, 1mM glutamine and cytokines 500ng/ml BMP4 (R&D systems), 500ng/ml  
558 BMP8a (R&D systems), 100ng/ml SCF (R&D systems), 50ng/ml EGF(R&D systems), LIF. Where  
559 indicated doxycycline was added at the point of mPGC-LCs induction from mEpiLCs.

#### 560 **RT-qPCR expression analysis**

561 Total RNA was isolated with Trizol and afterwards treated with turbo DNase. For reverse  
562 transcription of mRNAs, we used 1 µg of DNase digested RNA, random hexamer primers  
563 (5xTransAmp Buffer, Bioline) and reverse transcriptase (Bioline) in 20-µl reaction volume. qPCR  
564 analyses were carried out with SensiFAST SYBR No-ROX kit (Bioline) on LightCycler 480 II qPCR  
565 machine (Roche).

566

#### 567 **RNA-seq**

568 RNAs from at least two independent biological replicates of indicated cell lines were extracted  
569 with Trizol (Invitrogen), following the manufacturer's recommendations. Ten micrograms of total RNA  
570 were subjected to two rounds purification using Dynaloligo-dT beads (Invitrogen). Purified RNA was  
571 fragmented with 10x fragmentation buffer (Ambion) and used for first-strand cDNA synthesis, using  
572 random hexamer primers (Invitrogen) and SuperScript II enzyme (Invitrogen). Second strand cDNA  
573 was obtained by adding RNaseH (Invitrogen) and DNA Pol I (New England BioLabs). The resulting  
574 double-stranded cDNA was used for Illumina library preparation and sequenced with Illumina Genome  
575 Analyzer. Following library preparation, samples were pooled and sequenced on an Illumina NextSeq  
576 instrument using 76 base-pair single-end reads on a NextSeq high output kit (Illumina) or HiSeq  
577 instrument using 51 base-pair single-end reads.

578

#### 579 **ChIP-seq**

580 ChIP assays were performed from  $10^7$  mESC per experiment, according to previously described  
581 protocol with slight modification (Rada-Iglesias et al., 2011). Briefly, cells were crosslinked with 1%  
582 formaldehyde for 10 min at room temperature and the reaction was quenched by glycine at a final  
583 concentration of 0.125 M. Chromatin was sonicated to an average size of 0.5–2 kb, using Bioruptor  
584 (Diagenode). A total of 5 µg of antibody was added to the sonicated chromatin and incubated overnight  
585 at 4°C. Subsequently, 50 µl of protein G Dynal magnetic beads were added to the ChIP reactions and  
586 incubated for ~4 hr at 4°C. Magnetic beads were washed and chromatin eluted, followed by reversal of  
587 crosslinks and DNA purification. ChIP DNA was dissolved in water. ChIP-seq and input libraries were  
588 prepared according to Illumina protocol and sequenced using Illumina Genome Analyzer. Following  
589 library preparation, samples were pooled and sequenced on an Illumina NextSeq instrument using 76  
590 base-pair single-end reads on a NextSeq high output kit (Illumina).

591

#### 592 **RNA-seq analysis**



593           Quality of FASTQ files was assessed using FastQC software. Raw sequencing reads were  
594 aligned using Tophat against mm9 genomic index and with refseq gene models as available at  
595 illumina.com ftp site. Aligned reads were converted to counts for every gene using HTSeq and gene  
596 counts were further analyzed using R and DESeq2 package (Love et al., 2014). For the scatter plot we  
597 identified significantly affected transcripts and plotted RPKMs of transcripts significantly ( $q < 0.05$ )  
598 affected by Prdm14 or Mtgr1 loss. We compiled from published literature a set of official gene  
599 symbols for representative marker genes characteristic for epiblast, extraembryonic endoderm and  
600 naïve lineages.

601           The heatmap was created by looking at top 100 genes with the highest variance across samples  
602 (topVarGenes). We looked at the amount by which each gene deviates in a specific sample from the  
603 gene's average across all samples, thus we centered and scaled each genes' values across samples, and  
604 then plotted a heatmap. To visualize sample-to-sample distances between different lines we used  
605 principal component analysis (PCA), plotPCA function within DESeq2 package in R on the rlog-  
606 transformed counts.

607

## 608 **ChIP-seq analysis**

609           Quality of FASTQ files was assessed using FastQC software. ChIPseq peak calls were done  
610 with MACS2 callpeak with default settings (<https://pypi.python.org/pypi/MACS2>). Superset of  
611 intervals was created by merging summits from all calls using mean shift algorithm with 300bp  
612 bandwidth. The modal peaks were extended  $\pm 300$ bp and read coverage was calculated with bedtools.  
613 Regions with outlier counts in negative controls were excluded from further analysis.

614

## 615 **DNA sequence motif analysis**

616           Motifs enriched in Mtgr1 ChIP peaks were obtained with SeqPos (He et al., 2010), using a set  
617 of 1,884 coordinates for top Mtgr1 peaks with signal higher in Prdm14 over expressing cells than in wt

618 cells (Prdm14-dependent) and 1,721 Mtgr1 peaks with signal higher in wt cell than in the Prdm14  
619 overexpression background (Prdm14-independent). ChIP regions containing the MTGR1/Prdm14 motif  
620 were identified with FIMO with *P* value cutoff set to 0.001. In wt mESCs 64% of Prdm14-dependent  
621 sites contain the Prdm14 motif and 16% of Prdm14-independent sites contain the Prdm14 motif  
622 (p.value cutoff at 0.0003). This corresponds to odds ratio 9.84 and  $p < 10^{-16}$  in Fisher's exact test.

623

## 624 **CRISPR/Cas9 targeting**

625 Wild-type Cas9 plasmid pX330 was obtained from Addgene. The sgRNAs sequences were  
626 designed using Zhang Lab website (<http://crispr.mit.edu/>). Guide for Prdm14 was in exon 2  
627 (CGCCGCCGAGGACCAAATTTTGG, score 95) and guide for Mtgr1 was in exon 3  
628 (GACTCTCGTTCTAGCCTTGGTGG, score 78). Note that guides against exons 1, 2, and 3 within  
629 Mtgr1 were designed as well as nickase version of Cas9 was used, but only aforementioned guide  
630 within exon3 produced mutations that resulted in the loss of protein. *Stella*:GFP mESC line was  
631 transfected with the desired sgRNA in pX330 plasmid together with piggyBac mCherry (transient  
632 transfection) using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instruction  
633 manual. 48hrs post-transfection we did single-cell sorting into 96-well plates on mCherry-positive  
634 cells, assuming that these cells got transfected with both mCherry and pX330 plasmids. The colonies  
635 that arose from single cells were screened for the presence of the deletion. The target sequence was  
636 amplified by PCR with specific primers from genomic DNA. We then picked a restriction enzyme  
637 close to the PAM sequence that upon mutation of the sequence would not be able to cut. For Prdm14  
638 we used PflMI restriction enzyme and for Mtgr1 we used StyI restriction enzyme. Clones that could not  
639 be digested were further analyzed by doing PCR with specific primers from cDNA and subsequently  
640 Sanger sequencing. The results were analyzed with Sequencher 5.1 software and TIDE (Brinkman et  
641 al., 2014). Clones that were confirmed to have a mutation in cDNA were further validated for the  
642 presence of protein using Western blotting.

643

644 **Fluorescence-activated cell sorting (FACS)**

645 Cells carrying *Stella*:GFP reporter were used to monitor the efficiency of mPGC-LCs formation  
646 after 6 days of differentiation. Cells carrying dual reporter constructs (mCherry-S14 Mb and  
647 *Stella*:GFP) were PGC-LCs induced for 6 days with doxycycline after which the cells were analyzed.  
648 Differentiation was carried out in hanging drops as described. The cells were trypsinized, strained  
649 through a 30µm cell strainer and analyzed on an LSR Fortessa Analyzer (BD), data were analyzed  
650 further using FlowJo. For statistical analysis, Student's t test was used to compare two normally  
651 distributed data sets. The analysis was done in R using unpaired t.test and paired t.test for the same cell  
652 population before and after doxycycline treatment.  $p < 0.05$  was considered to be statistically  
653 significant.

654

655 **Immunoprecipitation**

656 Dignam nuclear extracts from mESCs were prepared as previously described (Peng et al.,  
657 2009). For immunoprecipitations, monobodies or antibodies that were used are listed in the antibody  
658 section below. Typically, 50-100 picomoles of monobody and 50 µl of pre-washed M280 dynabeads  
659 were used per immunoprecipitation. For immunoprecipitations performed using antibodies, we used 5  
660 µg of antibody and 75 µl of ProteinG-sepharose (Sigma) beads per immunoprecipitation. In double-step  
661 IP we first used FLAG M2-beads with peptide elution followed by incubation with HA antibody. If the  
662 immunoprecipitation was followed by mass spectrometry peptide identification, then the eluant was ran  
663 on the 1D SDS-PAGE gel, fixed and excised.

664

665 **Western Blotting**

666 Cells were lysed with RIPA buffer (50 mM Tris HCl pH 8, 300 mM NaCl, 1% Triton X-100,  
667 0.5% sodium deoxycholate, 0.1% SDS, 1mM EDTA) containing protease inhibitors (Roche tablet) and

668 1mM DTT. The protein concentration was estimated with Bradford reagent (Biorad) and equal or  
669 indicated amounts of protein were run on 8% SDS-PAGE gels and transferred to nitrocellulose  
670 membranes. Antibodies used in this study are listed in the antibody section.

671

## 672 **Protein identification by Mass Spectrometry**

673 In gel digestion was performed as previously reported (Schevencko, Nature Protocols 1, - 2856  
674 - 2860 (2007)) with the addition of Protease Max for increased peptide and protein solubility. The  
675 extracted peptides were dried using a speed vac and reconstituted in mobile phase A. The UPLC was a  
676 Waters M-class where mobile phase A was 0.2% formic acid, 5% DMSO, 94.8% water and mobile  
677 phase B was 0.2% formic acid, 5% DMSO, 94.8% acetonitrile. The UPLC was run at 300nL/min from  
678 4% mobile phase B to 35% mobile phase B followed by a wash and re-equilibration step. The mass  
679 spectrometer was an Orbitrap Fusion mass spectrometer set to acquire in a data dependent fashion to  
680 optimize cycle time and fragment ion acquisition. The .RAW data was searched using Byonic against  
681 the Uniprot mouse database downloaded on 09/29/2015. The fixed modifications were Cys.  
682 propionamide, an the variable Met. oxidation, Asp. deamidation and N-terminal modifications. The  
683 data was filtered and presented at a 1% false discovery rate (FDR).

684

## 685 **Protein expression and purification**

686 An expression vector for human PRDM14 (residues 238-487) with an N-terminal biotin-  
687 acceptor tag and C-terminal His<sub>6</sub> tag based on the p28BIOH-LIC vector (GenBank accession  
688 EF442785) was kindly provided by Susanne Gräslund and Cheryl Arrowsmith (Structural Genomics  
689 Consortium). The genes for Prdm14 (residues 184-373) and Mtgr1 (residues 98-206) were assembled  
690 using synthetic oligonucleotides and cloned in the pHBT vector that adds an N-terminal His<sub>6</sub> tag  
691 followed biotin-acceptor tag and a TEV cleavage site (Sha et al., 2013). The Mtgr1 construct in this  
692 vector contained a H200K mutation located outside the NHR1/TAFH domain due to a cloning artifact.

693 For binding assays, all proteins were expressed in BL21 (DE3) cells containing pBirACm plasmid  
694 (Avidity) in the presence of 50  $\mu$ M biotin to produce biotinylated proteins. The Prdm14-Mtgr1 fusion  
695 protein was designed to have a GSSGSSGS linker separating Prdm14 (residues 184-373) and Mtgr1  
696 (98-206). The DNA sequences for these genes have been deposited to the GenBank.

697 All proteins were expressed as His<sub>6</sub> tagged proteins as described. Proteins were purified using  
698 Ni-Sepharose gravity flow columns (GE Healthcare) and the monodispersity of these proteins was  
699 assessed by size-exclusion chromatography. For crystallization, the fusion tags were removed using  
700 TEV protease cleavage, and the tags were removed using Ni-Sepharose columns.

701

702 **Phage display and Yeast display based selection**

703 The method of selecting target specific monobodies from phage and yeast display libraries has  
704 been previously described (Koide et al., 2012a, 2012b). Two monobody libraries ('loop' and 'side')  
705 were used to generate monobodies with diverse binding modes (Koide et al., 2012a). Each of these  
706 libraries contains approximately ten billion unique monobody clones in which 16-26 residues are  
707 diversified using highly tailored amino acid combinations (Gilbreth and Koide, 2012; Koide et al.,  
708 2012a). Four rounds of phage display selection were performed using target concentrations of 100 nM,  
709 100 nM, 75 nM and 50 nM. Streptavidin-coated magnetic beads (Streptavidin MagneSphere  
710 Paramagnetic Particles; Promega, Z5481/2) were used for immobilizing the target and captured phages  
711 were eluted with 0.1M Gly-HCl, pH 2.1. After gene shuffling among the selected clones within the  
712 enriched population (Koide et al., 2012a), the monobody-coding genes were transferred into a yeast  
713 display vector. We performed library selection by yeast surface display, using magnetic beads in the  
714 first round followed by two rounds of fluorescence-activated cell sorting (FACS)-based selection.  
715 Binding assay for testing the affinity and specificity of individual monobody clones was performed  
716 using yeast surface display as described previously (Sha et al., 2013).

717

718 **Bead-based binding assays**

719 The general methods for bead-based assays have been described (Nishikori et al., 2012). In the  
720 assay, streptavidin-coated Dynabeads M280 beads (Invitrogen) at 20 µg/mL were incubated with 5 nM  
721 biotinylated target protein diluted in BSS/EDTA/DTT buffer (50 mM Tris–HCl, 150 mM NaCl, pH 8, 1  
722 mg/mL bovine serum albumin, 1 mM EDTA, 0.1 mM DTT) for 30 min. The remaining free biotin-  
723 binding sites of streptavidin on the M280 beads were blocked with 5 µM free biotin for 30 min. Ten µL  
724 of the target-immobilized beads were transferred to the wells of a 96-well filter plate  
725 (MultiScreen® HTS HV, 0.45 µm, Millipore), drained using a vacuum manifold (MultiScreen® HTS  
726 Vacuum Manifold, Millipore), and washed with 100 µL of BSS/EDTA/DTT buffer. Next, a biotinylated  
727 protein (biotinylation of the proteins was checked by their ability to bind to streptavidin beads) to be  
728 tested at various concentrations was added to individual wells and incubated for 30 min with gentle  
729 shaking. Then the wells of the filter plate were washed twice with 150 µL of the buffer, 20 µL of 10  
730 µg/mL SAV-Dylight650 (ThermoFisher) in the buffer was added to the wells, and the plate incubated  
731 with shaking for 30 min. The wells were washed again and the beads resuspended in 140 µL buffer and  
732 analyzed using a Guava EasyCyte 6/L flow cytometer (Millipore).

733

734 **Crystallization of the Prdm14-linker-Mtgr1/Mb(S4) complex**

735 Purified Prdm14-linker-Mtgr1 and Mb(S4) were mixed in the molar ratio of 1.0:1.3 and the  
736 complex was purified using a Superdex 75 16/600 size exclusion chromatography column (GE  
737 healthcare) in 25 mM Tris pH 8.0, 100 mM NaCl, 0.2 mM TCEP. The protein complex was then  
738 concentrated to a final concentration of 15 mg/ml. Initial crystallization screening of ~ 500 conditions  
739 was carried out in 96 well plates using the hanging-drop vapor-diffusion method with a crystallization  
740 robot (Mosquito™, TTPlabtech). Crystals used for data collection were obtained in 17% PEG3350 and  
741 8% Tascimate pH 5.5, and were cryoprotected in 1:1 mix of Paratone and Paraffin oil and flash-cooled  
742 in liquid nitrogen prior to data collection. The Prdm14-linker-Mtgr1 protein was labeled with SeMet as

described (Doublé, 1997), purified as a complex with the monobody and crystallized in a similar manner as the native proteins.

## **Data Collection, Structure Determination and Refinement**

X-ray diffraction data were collected at Beamline 19ID of the Advanced Photon Source (Argonne National Laboratory, Chicago, IL, USA) (Table 1). The data were indexed and integrated using HKL3000 (Minor et al., 2006). Molecular replacement using PHASER (McCoy et al., 2007) and MOLREP (Vagin and Teplyakov, 1997) with a hPrdm12 structure (PDB ID 3EP0) and the monobody backbone (PDB 3UYO) did not have sufficient scattering power to generate a solution with a signal-to-noise ratio that is high enough to be identified. Thus, we determined the structure through single-wavelength Se anomalous dispersion experiment. A total of eight Se sites were identified and refined using Autosol (Terwilliger et al., 2009), resulting in an overall figure of merit of 0.45 and Z-score of 43.1. These phases were then used against the SAD data for model building in phenix.autobuild (Adams et al., 2010). Iterative model building and refinement were done using the programs COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010). The structure refined from SAD data was later refined against the higher resolution native data at a 3.05 Å resolution. The final structures were analyzed using Procheck and Molprobit (Davis et al., 2004). Figures were made using Pymol (DeLano, 2002). The structure has 100% residues in the allowed regions of the Ramachandran plot with no outliers. The Molprobit score (2.16) is above average for structures refined at comparable resolutions.

## **NMR Spectroscopy**

Uniformly  $^{15}\text{N}$ -labeled Prdm14-linker-Mtgr1 and  $^{15}\text{N}$ -labeled Prdm14 were prepared by growing bacterial cells in M9 minimal media supplemented with  $^{15}\text{N}$ -labeled ammonium sulphate (0.8 g l<sup>-1</sup>, Cambridge Isotope Laboratories, Inc.). The labeled proteins were purified in the same manner as

768 the unlabeled proteins described above. <sup>15</sup>N-labeled Prdm14 in complex with unlabeled <sup>14</sup>N-Mtgr1 was  
769 purified by gel filtration chromatography. NMR data was collected at 30°C on a 600 MHz Bruker  
770 AVANCE III Spectrometer. The samples used for data collection contained 50-200 µM protein in 50  
771 mM Tris-Cl buffer pH 8.0 containing 150 mM NaCl and 0.2 mM TCEP supplemented with 10% D<sub>2</sub>O.  
772 All spectra were processed by the NMRPipe software (Delaglio et al., 1995) and analyzed using  
773 SPARKY (Goddard and Kneller).

774

## 775 **Antibodies**

776 Antibodies for Mtgr1 (Western and IP, ab53363, lot GR56108-2,4; ChIP, ab96161), V5tag  
777 (Western and IP, ab27671, lot GR186433-4), HA (ChIP, ab9110, lot GR146572-8) were from Abcam,  
778 and Suz12 (IP, 04-046) from Millipore. HA antibody (Western and IP, H3663), anti-Flag M2 agarose  
779 beads (A2220), M2 Flag antibody (Western, F1804) were from Sigma and dynabeads M280  
780 streptavidin (11205D) were from Life Technologies.

781

## 782 **Author Contributions**

783 N.N. designed and performed biochemical and cellular studies of Prdm14-Mtgr1 interaction presented  
784 in Figs. 1, 2, 3, 4E and 6, with exceptions of Fig.1A, E-G; A.G. performed biophysical, structural and  
785 protein-engineering studies presented in Figs. 1E-G, 4A-D and 5; Z.M. performed IP-MS experiments  
786 in Fig. 1A and done preliminary work on Mtgr1-Prdm14 functional association; T.S. contributed to  
787 analyses of RNA-seq and ChIP-seq datasets and provided experimental advice; A.K. provided input on  
788 experiments in Fig. 4. S.K. and J.W. guided experimental design and data interpretation. N.N., A.G.,  
789 S.K. and J.W. wrote the manuscript.

790

## 791 **Acknowledgments**

792 We would like to thank members of the Wysocka lab for insightful comments and critical editing of the



manuscript. We thank A. Surani for *Stella*:GFP mESCs, J. Collier, D. Wagh, and X. Ji at Stanford Functional Genomics Facility (SFGF) for assistance with high-throughput sequencing, L. Picton for help with NMR data collection, S. Ginell and Y. Kim at the Advanced Photon Source (APS) Sector 19ID for beamline access and assistance during data collection and processing. This work was supported by the Howard Hughes Medical Institute funds, National Institutes of Health grants R01GM112720 (J.W.), R01DA036887 and R01GM090324 (SK), P30 CA014599 (the University of Chicago Comprehensive Cancer Center, S.K.), and Canadian Institutes for Health Research (CIHR) post-doctoral fellowship (N.N.). This research used resources of the APS, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

803

804 **References**

805 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-  
806 W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based  
807 system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 213–221.

808 Amann, J.M., Chyla, B.J.I., Ellis, T.C., Martinez, A., Moore, A.C., Franklin, J.L., McGhee, L., Meyers,  
809 S., Ohm, J.E., Luce, K.S., et al. (2005). *Mtgr1* is a transcriptional corepressor that is required for  
810 maintenance of the secretory cell lineage in the small intestine. *Mol. Cell. Biol.* *25*, 9576–9585.

811 Amano, T., Hirata, T., Falco, G., Monti, M., Sharova, L.V., Amano, M., Sheer, S., Hoang, H.G., Piao,  
812 Y., Stagg, C.A., et al. (2013). *Zscan4* restores the developmental potency of embryonic stem cells. *Nat.*  
813 *Commun.* *4*, 1966.

814 Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of  
815 genome editing by sequence trace decomposition. *Nucleic Acids Res.* *42*, e168.

816 Burton, A., Muller, J., Tu, S., Padilla-Longoria, P., Guccione, E., and Torres-Padilla, M.-E. (2013).  
817 Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the  
818 mammalian embryo. *Cell Rep.* *5*, 687–701.

819 Carofino, B.L., Ayanga, B., and Justice, M.J. (2013). A mouse model for inducible overexpression of  
820 *Prdm14* results in rapid-onset and highly penetrant T-cell acute lymphoblastic leukemia (T-ALL). *Dis.*  
821 *Model. Mech.* *6*, 1494–1506.

822 Chan, Y.-S., Göke, J., Lu, X., Venkatesan, N., Feng, B., Su, I.-H., and Ng, H.-H. (2013). A PRC2-  
823 dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem

824 cell reprogramming. *Stem Cells Dayt. Ohio* 31, 682–692.

825 Dan, J., Li, M., Yang, J., Li, J., Okuka, M., Ye, X., and Liu, L. (2013). Roles for Tbx3 in regulation of  
826 two-cell state and telomere elongation in mouse ES cells. *Sci. Rep.* 3, 3492.

827 Davis, C.A., Haberland, M., Arnold, M.A., Sutherland, L.B., McDonald, O.G., Richardson, J.A.,  
828 Childs, G., Harris, S., Owens, G.K., and Olson, E.N. (2006). PRISM/PRDM6, a transcriptional  
829 repressor that promotes the proliferative gene program in smooth muscle cells. *Mol. Cell. Biol.* 26,  
830 2626–2636.

831 Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure  
832 validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* 32,  
833 W615–W619.

834 Davis, J.N., McGhee, L., and Meyers, S. (2003). The ETO (MTG8) gene family. *Gene* 303, 1–10.

835 Del Rizzo, P.A., and Trievel, R.C. (2011). Substrate and product specificities of SET domain  
836 methyltransferases. *Epigenetics Off. J. DNA Methylation Soc.* 6, 1059–1067.

837 Dettman, E.J., Simko, S.J., Ayanga, B., Carofino, B.L., Margolin, J.F., Morse, H.C., and Justice, M.J.  
838 (2011). Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling  
839 common lymphoid progenitors. *Oncogene* 30, 2859–2873.

840 Doublé, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods*  
841 *Enzymol.* 276, 523–530.

842 Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta*  
843 *Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.

844 Ernst, A., Appleton, B.A., Ivarsson, Y., Zhang, Y., Gfeller, D., Wiesmann, C., and Sidhu, S.S. (2014). A  
845 structural portrait of the PDZ domain family. *J. Mol. Biol.* 426, 3509–3519.

846 Fog, C.K., Galli, G.G., and Lund, A.H. (2012). PRDM proteins: important players in differentiation and  
847 disease. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 34, 50–60.

848 Gilbreth, R.N., and Koide, S. (2012). Structural insights for engineering binding proteins based on non-  
849 antibody scaffolds. *Curr. Opin. Struct. Biol.* 22, 413–420.

850 Gilbreth, R.N., Esaki, K., Koide, A., Sidhu, S.S., and Koide, S. (2008). A dominant conformational role  
851 for amino acid diversity in minimalist protein-protein interfaces. *J. Mol. Biol.* 381, 407–418.

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

855 Grebien, F., Hantschel, O., Wojcik, J., Kaupe, I., Kovacic, B., Wyrzucki, A.M., Gish, G.D., Cerny-  
856 Reiterer, S., Koide, A., Beug, H., et al. (2011). Targeting the SH2-kinase interface in Bcr-Abl inhibits  
857 leukemogenesis. *Cell* 147, 306–319.

858 Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013).  
859 Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem*

860 Cell Rep. *1*, 518–531.

861 Hatlen, M.A., Wang, L., and Nimer, S.D. (2012). AML1-ETO driven acute leukemia: insights into  
862 pathogenesis and potential therapeutic approaches. *Front. Med.* *6*, 248–262.

863 Hayashi, K., and Saitou, M. (2013). Stepwise differentiation from naïve state pluripotent stem cells to  
864 functional primordial germ cells through an epiblast-like state. *Methods Mol. Biol. Clifton NJ* *1074*,  
865 175–183.

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

868 He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien,  
869 M., et al. (2010). Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* *42*, 343–347.

870 Hohenauer, T., and Moore, A.W. (2012). The Prdm family: expanding roles in stem cells and  
871 development. *Dev. Camb. Engl.* *139*, 2267–2282.

872 Kobe, B., Ve, T., and Williams, S.J. (2015). Fusion-protein-assisted protein crystallization. *Acta*  
873 *Crystallogr. Sect. F Struct. Biol. Commun.* *71*, 861–869.

874 Koide, S. (2009). Engineering of recombinant crystallization chaperones. *Curr. Opin. Struct. Biol.* *19*,  
875 449–457.

876 Koide, A., Bailey, C.W., Huang, X., and Koide, S. (1998). The fibronectin type III domain as a scaffold  
877 for novel binding proteins. *J. Mol. Biol.* *284*, 1141–1151.

878 Koide, A., Wojcik, J., Gilbreth, R.N., Hoey, R.J., and Koide, S. (2012a). Teaching an old scaffold new  
879 tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J. Mol. Biol.* *415*, 393–  
880 405.

881 Koide, S., Koide, A., and Lipovšek, D. (2012b). Target-binding proteins based on the 10th human  
882 fibronectin type III domain (<sup>10</sup>Fn3). *Methods Enzymol.* *503*, 135–156.

883 Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun,  
884 B., Knezovich, J.G., Smith, A., et al. (2013). Naïve pluripotency is associated with global DNA  
885 hypomethylation. *Nat. Struct. Mol. Biol.* *20*, 311–316.

886 Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition  
887 sites. *J. Mol. Biol.* *285*, 2177–2198.

888 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for  
889 RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.

890 Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. (2011). Sequence-specific regulator  
891 Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat. Struct. Mol. Biol.*  
892 *18*, 120–127.

893 Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer,  
894 O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous  
895 retrovirus activity. *Nature* *487*, 57–63.

896 Magnúsdóttir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Dev. Camb. Engl.*  
897 *141*, 245–252.

898 Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao,  
899 K., Gottgens, B., and Azim Surani, M. (2013). A tripartite transcription factor network regulates  
900 primordial germ cell specification in mice. *Nat. Cell Biol.* *15*, 905–915.

901 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007).  
902 Phaser crystallographic software. *J. Appl. Crystallogr.* *40(Pt 4)*, 658–674.

903 Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the integration of  
904 data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta*  
905 *Crystallogr. D Biol. Crystallogr.* *62*, 859–866.

906 Nakaki, F., and Saitou, M. (2014). PRDM14: a unique regulator for pluripotency and epigenetic  
907 reprogramming. *Trends Biochem. Sci.* *39*, 289–298.

908 Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y., and Saitou, M. (2013). Induction of mouse  
909 germ-cell fate by transcription factors in vitro. *Nature* *501*, 222–226.

910 Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* *4*, 487–492.

911 Nishikawa, N., Toyota, M., Suzuki, H., Honma, T., Fujikane, T., Ohmura, T., Nishidate, T., Ohe-Toyota,  
912 M., Maruyama, R., Sonoda, T., et al. (2007). Gene amplification and overexpression of PRDM14 in  
913 breast cancers. *Cancer Res.* *67*, 9649–9657.

914 Nishikori, S., Hattori, T., Fuchs, S.M., Yasui, N., Wojcik, J., Koide, A., Strahl, B.D., and Koide, S.  
915 (2012). Broad ranges of affinity and specificity of anti-histone antibodies revealed by a quantitative  
916 peptide immunoprecipitation assay. *J. Mol. Biol.* *424*, 391–399.

917 Okashita, N., Kumaki, Y., Ebi, K., Nishi, M., Okamoto, Y., Nakayama, M., Hashimoto, S., Nakamura,  
918 T., Sugawara, K., Kojima, N., et al. (2014). PRDM14 promotes active DNA demethylation through the  
919 ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Dev.*  
920 *Camb. Engl.* *141*, 269–280.

921 Park, S., Chen, W., Cierpicki, T., Tonelli, M., Cai, X., Speck, N.A., and Bushweller, J.H. (2009).  
922 Structure of the AML1-ETO eTAFH domain-HEB peptide complex and its contribution to AML1-ETO  
923 activity. *Blood* *113*, 3558–3567.

924 Payer, B., Chuva de Sousa Lopes, S.M., Barton, S.C., Lee, C., Saitou, M., and Surani, M.A. (2006).  
925 Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. *Genes. N. Y. N*  
926 *2000* *44*, 75–83.

927 Payer, B., Rosenberg, M., Yamaji, M., Yabuta, Y., Koyanagi-Aoi, M., Hayashi, K., Yamanaka, S.,  
928 Saitou, M., and Lee, J.T. (2013). Tsix RNA and the germline factor, PRDM14, link X reactivation and  
929 stem cell reprogramming. *Mol. Cell* *52*, 805–818.

930 Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009).  
931 Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in  
932 pluripotent cells. *Cell* *139*, 1290–1302.

933 Plevin, M.J., Zhang, J., Guo, C., Roeder, R.G., and Ikura, M. (2006). The acute myeloid leukemia  
934 fusion protein AML1-ETO targets E proteins via a paired amphipathic helix-like TBP-associated factor  
935 homology domain. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 10242–10247.

936 Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A  
937 unique chromatin signature uncovers early developmental enhancers in humans. *Nature* *470*, 279–283.

938 Reddy Chichili, V.P., Kumar, V., and Sivaraman, J. (2013). Linkers in the structural biology of protein-  
939 protein interactions. *Protein Sci. Publ. Protein Soc.* *22*, 153–167.

940 Rossetti, S., Hoogetveen, A.T., and Sacchi, N. (2004). The MTG proteins: chromatin repression players  
941 with a passion for networking. *Genomics* *84*, 1–9.

942 Rossetti, S., van Unen, L., Sacchi, N., and Hoogetveen, A.T. (2008). Novel RNA-binding properties of  
943 the MTG chromatin regulatory proteins. *BMC Mol. Biol.* *9*, 93.

944 Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., Perdeaux, E., Rapley, E., Eeles, R.,  
945 Peto, J., et al. (2013). Identification of nine new susceptibility loci for testicular cancer, including  
946 variants near DAZL and PRDM14. *Nat. Genet.* *45*, 686–689.

947 Saitou, M., Kagiwada, S., and Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-  
948 implantation development and primordial germ cells. *Dev. Camb. Engl.* *139*, 15–31.

949 Sha, F., Gencer, E.B., Georgeon, S., Koide, A., Yasui, N., Koide, S., and Hantschel, O. (2013).  
950 Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2  
951 SH2 domains. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 14924–14929.

952 Smith, B.C., and Denu, J.M. (2009). Chemical mechanisms of histone lysine and arginine  
953 modifications. *Biochim. Biophys. Acta* *1789*, 45–57.

954 Stockbridge, R.B., Kolmakova-Partensky, L., Shane, T., Koide, A., Koide, S., Miller, C., and Newstead,  
955 S. (2015). Crystal structures of a dual-topology, double-barrelled fluoride ion channel. *Nature*.

956 Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W.,  
957 Afonine, P.V., Zwart, P.H., and Hung, L.W. (2009). Decision-making in structure solution using  
958 Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.*  
959 *65*, 582–601.

960 Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J.*  
961 *Appl. Crystallogr.* *30*, 1022–1025.

962 Wei, Y., Liu, S., Lausen, J., Woodrell, C., Cho, S., Biris, N., Kobayashi, N., Wei, Y., Yokoyama, S., and  
963 Werner, M.H. (2007). A TAF4-homology domain from the corepressor ETO is a docking platform for  
964 positive and negative regulators of transcription. *Nat. Struct. Mol. Biol.* *14*, 653–661.

965 Wojcik, J., Hantschel, O., Grebien, F., Kaupe, I., Bennett, K.L., Barkinge, J., Jones, R.B., Koide, A.,  
966 Superti-Furga, G., and Koide, S. (2010). A potent and highly specific FN3 monobody inhibitor of the  
967 Abl SH2 domain. *Nat. Struct. Mol. Biol.* *17*, 519–527.

968 Wu, H., Mathioudakis, N., Diagouraga, B., Dong, A., Dombrovski, L., Baudat, F., Cusack, S., de  
969 Massy, B., and Kadlec, J. (2013). Molecular basis for the regulation of the H3K4 methyltransferase

970 activity of PRDM9. *Cell Rep.* 5, 13–20.

971 Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and  
972 Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice.  
973 *Nat. Genet.* 40, 1016–1022.

974 Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y.,  
975 Shirahige, K., and Saitou, M. (2013). PRDM14 ensures naive pluripotency through dual regulation of  
976 signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* 12, 368–382.

977 Yu, J., Angelin-Duclos, C., Greenwood, J., Liao, J., and Calame, K. (2000). Transcriptional repression  
978 by blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. *Mol. Cell. Biol.* 20, 2592–2603.

979 Zhang, T., Meng, L., Dong, W., Shen, H., Zhang, S., Liu, Q., and Du, J. (2013). High expression of  
980 PRDM14 correlates with cell differentiation and is a novel prognostic marker in resected non-small cell  
981 lung cancer. *Med. Oncol. Northwood Lond. Engl.* 30, 605.

982 Zhou, Q., Lai, Y., Bacaj, T., Zhao, M., Lyubimov, A.Y., Uervirojnangkoon, M., Zeldin, O.B., Brewster,  
983 A.S., Sauter, N.K., Cohen, A.E., et al. (2015). Architecture of the synaptotagmin-SNARE machinery  
984 for neuronal exocytosis. *Nature* 525, 62–67.

985

986

987 **Figure Legends**

988 **Figure 1. Prdm14 Directly Binds ETO Family Protein, Mtgr1.** **A.** Two-step immunoaffinity  
989 purification of Prdm14-associated proteins. FLAG-HA immunoprecipitations were performed from  
990 wild type or FH-Prdm14 mESC extracts, followed by visualization of polypeptides by SDS-PAGE-  
991 silver stain and mass spectrometry identification. Polypeptides corresponding to Mtgr1 and Prdm14 are  
992 highlighted. MW, molecular weight marker. **B.** Reciprocal Mtgr1 and Prdm14 co-immunoprecipitations  
993 from FH-Prdm14 mESCs. **C-D.** Identification of the Prdm14-Mtgr1 interaction regions. Co-  
994 immunoprecipitations were performed in HEK293 cells transfected with full-length V5-tagged Mtgr1  
995 and distinct FH-Prdm14 constructs (C) or full-length FH-Prdm14 and distinct V5-Mtgr1 constructs (D),  
996 as indicated in the top diagrams; co-immunoprecipitated proteins were visualized by immunoblotting  
997 with  $\alpha$ -HA (Prdm14) and  $\alpha$ -V5 (Mtgr1) antibodies (top left and right panels, respectively). Tagged  
998 Prdm14 and Mtgr1 levels in the input extracts are shown in the bottom panels. \*indicates non-specific  
999 bands. **E.** Recombinant biotinylated Prdm14 (pre-SET+SET) was immobilized on streptavidin coated  
1000 beads and incubated with recombinant Mtgr1 (NHR1), and changes in fluorescence (SAV-Dylight650)  
1001 were measured. **F-G.** Recombinant biotinylated Mtgr1 (NHR1) was immobilized on streptavidin  
1002 coated beads and incubated with (F) Prdm14 (pre-SET+SET) or (G) Prdm14 (SET) and changes in  
1003 fluorescence (SAV-Dylight650) were measured. The errors bars and the errors for the  $K_D$  values are the  
1004 s.d. (n = 3). The curves show the best fit of the 1:1 binding model using the GraphPad software.

1005  
1006 **Figure 2. Prdm14 and Mtgr1 complex co-occupy genomic targets.** **A.** Prdm14 and Mtgr1 ChIP-seq  
1007 enrichments at selected gene loci. Tracks represent sequence tag enrichments as determined by Quest  
1008 software. **B.** Scatter plot of Prdm14 and Mtgr1 genomic occupancies in FH-Prdm14 mESC line. **C.** Top  
1009 sequence motif recovered in Mtgr1 ChIP-seq corresponds to Prdm14 motif, as defined previously (Ma  
1010 et al. 2011). Logos for the consensus motifs were generated using SeqPos. **D.** Scatter plot of Mtgr1  
1011 genomic occupancy in wild type and FH-Prdm14 mESC lines. The plot is colored based on the

1012 presence of Prdm14 motif (red, motif is present p.value < 10<sup>-3</sup>; black, motif is absent).

1013

1014 **Figure 3. Loss of Mtgr1 Phenocopies Requirement for Prdm14 in Safeguarding Pluripotency and**

1015 **PGC Induction. A.** RNA-seq from *Prdm14*<sup>-/-</sup> cells or *Mtgr1*<sup>-/-</sup> cells (y axis) were compared to wild

1016 type cells (x axis) and expression values (RPKM) of all significantly changed transcripts were plotted.

1017 Select transcripts corresponding to those enriched in the post-implantation epiblast, extraembryonic

1018 endoderm or naïve pluripotent mESC are highlighted in red, green or blue, respectively; shaded colors

1019 indicate no significant difference. **B.** Heatmap displaying top 100 variable genes between wild type

1020 mESCs grown under naïve 2i+LIF or serum+LIF conditions, *Prdm14*<sup>-/-</sup> (2 clones) or *Mtgr1*<sup>-/-</sup> (3

1021 clones) cells, and their respective rescue clones. Clustering represents sample divergence. **C.** Principal

1022 component analysis on the same populations as in B. **D.** mESC to mEpiLC transition followed by

1023 PGC-LC induction using defined media in cells containing *Stella*:GFP reporter. Schematic of the

1024 *Stella*:GFP transgene reporter that contains a 10kb 5' upstream sequence and includes exon 1 and part

1025 of exon 2 fused in-frame with eGFP, followed by the SV40 polyadenylation sequence (Payer et al.

1026 2006). The reporter is active in mPGC-LCs when *Stella* expression is activated. **E.** Quantification of

1027 the GFP signal in wild type cells during the mESC to mEpiLC and further to mPGC-LC transition. **F.**

1028 FACS plots and gated quantification of GFP signal as a measure of mPGC-LC induction from wild

1029 type cells, *Prdm14*<sup>-/-</sup> cells or Prdm14 rescue clones. **G.** FACS plots and quantification of GFP signal as

1030 a measure of PGC-LC induction from wild type cells, *Mtgr1*<sup>-/-</sup> cells or Mtgr1 rescue clones.

1031

1032 **Figure 4. Generation of PRDM14-binding monobodies. A.** Schematic of the monobody scaffold.

1033 The β strands and loops are labeled, and the diversified residues are marked as red spheres. The amino

1034 acid sequences of the monobody libraries and monobody clones. In the library designs, “X” denotes a

1035 mixture of 30% Tyr, 15% Ser, 10% Gly, 5% Phe, 5% Trp, and 2.5% each of all the other amino acids

1036 except for Cys; “B”, a mixture of Gly, Ser, and Tyr; “J”, a mixture of Ser and Tyr; “O”, a mixture of

1037 Asn, Asp, His, Ile, Leu, Phe, Tyr, and Val; “U”, a mixture of His, Leu, Phe, and Tyr; “Z”, a mixture of



Ala, Glu, Lys, and Thr (Koide et al. 2012). A hyphen indicates a deletion. **B.** Titration curves of Mb(hPRDM14\_S4) and Mb(hPRDM14\_S14) to human PRDM14 and mouse Prdm14. The error bars are the s.d. (n = 3). The curves show the best fit of the 1:1 binding model. **C.** Binding of Mb(S4) and Mb(S14) expressed on yeast surface to 50 nM of hPRDM14 and its homologues, mouse Prdm14, human PRDM12 and human PRDM6. **D.** Competitive binding assay for Mtgr1 and monobodies. Binding of 10 nM Mtgr1 to biotinylated Prdm14 immobilized on streptavidin coated M280 beads in the absence and presence of 500 nM purified monobodies, Mb(S4) and Mb(S14). **E.** Co-immunoprecipitation of FLAG-HA tagged Prdm14 expressed in mESC using Mb(S4), Mb(S14),  $\alpha$ -FLAG M2 antibody or a negative control antibody ("IgG"). Antibodies used for Western blotting are indicated with the blots. FT, flow through; E, Elution.

**Figure 5. Crystal structure of the Prdm14-Mtgr1 complex.** **A.** The overall structure of the Prdm14-link-Mtgr1/Mb(S4) complex. Missing residues are shown with dotted lines. The pre-SET region (in orange), SET domain (in yellow) and post-SET region (in green) of Prdm14 are shown. The Mtgr1 helices are marked for clarity ( $\alpha$ A- $\alpha$ D). **B.** Superposition of the Prdm14 crystal structure with the crystal structure of hPRDM12 (PDB ID 3EP0). **C.** Superposition of the Mtgr1 crystal structure with the NMR structure of the MTG8 NHR1 (eTAFH) domain (PDB ID 2KNH) **D.** Prdm14-Mtgr1 interface. Prdm14 is shown in white surface representation with the interacting residues in yellow (Top). Mtgr1 is shown in cartoon representation in pink color. In the detailed view, Prdm14 residues are marked in red and Mtgr1 residues are marked in black. Salt bridges and hydrogen bonds between Prdm14 and Mtgr1 are shown in dotted lines. . Residues that were mutated based on the structure are underlined. **E.** Binding of wild type, E294K and Y339R Prdm14 (residues 184-373) to immobilized wild-type Mtgr1 (top) and Mtgr1(K109E) (bottom) in a bead-based assay. **F.** Mtgr1 is shown in white surface representation with interacting residues in pink (left). Non-identical residues between Mtgr1 and the other ETO proteins are shown in blue (left). The detailed view shows the interaction of Prdm14

1063 residues N-terminal of SET domain (pre-SET) with Mtgr1 (right). Prdm14 residues are labeled in red  
1064 and Mtgr1 residues in black.

1065  
1066 **Figure 6. Inhibition of Prdm14-Mtgr1 interaction affects stem cell maintenance and PGC-LC**  
1067 **induction.** **A.** Single amino-acid substitutions at the interaction surface abrogate Prdm14-Mtgr1  
1068 association in cells. Indicated full length V5-Mtgr1 wild type or single point mutant proteins were  
1069 introduced to HEK293 cells and co-immunoprecipitated with full length FH-Prdm14 wild type or  
1070 single mutant protein as indicated in the diagram. Please note rescue of the association when the  
1071 combination of Mtgr1 K109E and Prdm14 E294K mutants is tested. **B.** Western blot showing protein  
1072 expression levels of wild type, E294K and Y339R FH-Prdm14 protein in lysates from Prdm14<sup>-/-</sup>  
1073 mESCs cells reconstituted with the respective transgenes. WCL, whole cell lysate. **C.** RNA-seq from  
1074 *Prdm14*<sup>-/-</sup> cells reconstituted with wild type Prdm14 protein (x axis) were compared to *Prdm14*<sup>-/-</sup> cells  
1075 reconstituted with E294K Prdm14 mutant (y axis) and expression values (RPKM) of all significantly  
1076 changed transcripts were plotted. The transcripts of specific genes are highlighted in red, green, blue or  
1077 black as indicated; shaded colors indicate no significant difference. **D.** Quantification of GFP signal as  
1078 a measure of mPGC-LC induction from *Prdm14*<sup>-/-</sup> cells and *Prdm14*<sup>-/-</sup> cells reconstituted with  
1079 transgenes encoding wild type, E294K, or Y339R Prdm14 protein. **E.** Schematics of the piggyBac  
1080 transposon-based reporter system used to create dual reporter lines. mESC line was transfected with  
1081 either dox-inducible mCherry construct or dox-inducible mCherry-Mb(S14) fusion protein. The lines  
1082 were selected using blasticidin and three populations were tested further for their competency to form  
1083 mPGC-LCs. mESC to mEpiLC transition followed by mPGC-LC transition using defined media in  
1084 cells containing *Stella*:GFP reporter (lower panel). Doxycycline was added after mEpiLC stage. **F.**  
1085 Quantification of GFP signal as a measure of mPGC-LC induction from mCherry population of cells  
1086 and mCherry-Mb(S14) population of cells with and without addition of doxycycline.

1087  
1088

1089

1090

1091

1092

1093

**Table Legends**  
**Table 1.** Data collection, phasing and refinement statistics for Prdm14-linker-Mtgr1/Mb(S4) complex crystals.  
  
**Table 1.** Data collection, phasing and refinement statistics for Prdm14-linker-Mtgr1/Mb(S4) complex crystals.

	Native	SeMet SAD
<b>Data collection</b>		
Beamline	APS 19ID	APS 19ID
Space group	P4 <sub>3</sub> 2 <sub>1</sub> 2	P4 <sub>3</sub> 2 <sub>1</sub> 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	106.8,106.8,180.7	106.9, 106.9,180.9
$\alpha$ , $\beta$ , $\gamma$ (°)	90,90,90	90,90,90
		<i>Peak</i>
<u>Wavelength</u>	0.97918 Å	0.97918 Å
Resolution (Å)	37.7-3.05 (3.16-3.05)	50-3.18 (3.23-3.18)
<i>R</i> <sub>pin</sub>	0.024 (0.482)	0.022 (0.315)
<i>I</i> / $\sigma$ <i>I</i>	30.0 (1.4)	52.4 (2.0)
Completeness (%)	100 (100)	100 (100)
Redundancy	20.5 (19.5)	86.2 (45.2)
<b>Refinement</b>		
Resolution (Å)	37.7-3.05 (3.16-3.05)	
No. of unique reflections	20631 (2013)	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.192/0.253	
No. atoms	5476	
Protein	5476	
Ligand/ion	0	
Water	0	
<i>B</i> -factors	114.3	
Protein	114.3	
Ligand/ion	0	
Water	0	
R.m.s deviations		
Bond lengths (Å)	0.006	
Bond angles (°)	1.13	
Ramachandran Statistics		
Favorable	96.0	
Allowed	4.0	
Outliers	0.0	

1094

1095

1096 **Figure Supplement Legends**

1097 **Figure 1—figure supplement 1. Expression levels of mRNAs encoding ETO proteins.** Plotted are  
1098 RPKM values for Mtgr1, Mtg1 and Mtg16 obtained in RNA-seq experiments from mESC grown under  
1099 serum+LIF or 2i+LIF conditions or from epiblast-like cells (EpiLC).

1100  
1101 **Figure 2—figure supplement 1. Genomic occupancy of Mtgr1.** (A). Correlation coefficient, R,  
1102 between Mtgr1 and Prdm14 (anti-HA) ChIP-seq genomic occupancy levels in two different clonal FH-  
1103 Prdm14 mESC lines. (B). Functional annotation categories of the Mtgr1 binding sites in FH-Prdm14  
1104 mESCs as determined by GREAT.

1105  
1106 **Figure 2—figure supplement 2. Genomic occupancy of Mtgr1 at Prdm14 motif-lacking sites.** (A)  
1107 Genomic tracks of at select Prdm14 motif-lacking Mtgr1 binding sites. Note that these sites lack  
1108 Prdm14 binding and Mtgr1 peaks are present in wt cells, but not in the FH-Prdm14-overexpressing  
1109 line. (B). Aggregate plots of Mtgr1 ChIP-seq enrichments at Prdm14 motif-containing (left) or Prdm14  
1110 motif-lacking (right) Mtgr1 sites from indicated ESC cell lines. (C). Top five motifs recovered from the  
1111 Mtgr1-binding sites where Prdm14 motif was absent.

1112  
1113 **Figure 3—figure supplement 1. Generation of Mtgr1-null line in *Stella*:GFP mESCs using**  
1114 **CRISPR-Cas9 system.** (A). Schematic of the *Mtgr1* gene and sequence of the guide RNA  
1115 (<http://crispr.mit.edu/>). (B). Genomic and cDNA were extracted and chromatograms of the wild type  
1116 (top), mutant A3 line (clone 1), mutant C1 line (clone 2), and mutant C3 line (clone 3) are shown. (C).  
1117 Indel spectrum centered around the PAM sequence was performed using software TIDE  
1118 (<http://tide.nki.nl/>). Clones A3, C1, and C3 are shown, they all produce aberrant transcript. (D).  
1119 Western blotting of the whole cell lysate from the wt, A3, C1 and C3 mESC *Stella*:GFP line.

1120

1121 **Figure 3—figure supplement 2. Generation of Prdm14-null line in *Stella*:GFP mESCs using**  
1122 **CRISPR-Cas9 system.** (A). Schematic of the *Prdm14* gene and sequence of the guide RNA  
1123 (<http://crispr.mit.edu/>). (B). Genomic and cDNA were extracted and chromatograms of the wild type  
1124 (top), mutant G6 line (clone 1), and mutant F10 line (clone 2) are shown. (C). Indel spectrum centered  
1125 around the PAM sequence was performed using software TIDE (<http://tide.nki.nl/>). Clone G6 is shown  
1126 on top and clone F10 below. Clones G6 and F10 produce aberrant transcript.

1128 **Figure 3—figure supplement 3. Morphological changes associated with loss of Prdm14 or Mtgr1.**  
1129 Brighfield images of cell morphology from wt ESCs, *Prdm14*<sup>-/-</sup> or *Mtgr1*<sup>-/-</sup> ESCs or those  
1130 reconstituted with wild type FH-Prdm14 or FH-Mtgr1, respectively, all grown under serum+LIF  
1131 conditions; wt ESC maintained in 2i+LIF are included for relative comparisons (upper left image).

1133 **Figure 3—figure supplement 4. Additional analyses of RNA-seq datasets.** (A). RNA-seq from  
1134 *Prdm14*<sup>-/-</sup> cells or *Mtgr1*<sup>-/-</sup> cells (y axis) were compared to respective rescue lines (x axis) and  
1135 expression values (RPKM) of all significantly changed transcripts were plotted. The transcripts of  
1136 specific genes are highlighted in red, green or blue as indicated; shaded colors indicate no significant  
1137 difference. (B). Scatter plot of genes upregulated at least twofold upon loss of either Prdm14 or Mtgr1.  
1138 Highlighted in blue are genes significant for Mtgr1-null cells only, red – Prdm14-null only, magenta –  
1139 genes significantly upregulated in both cell lines.

1141 **Figure 3—figure supplement 5. Expression of differentiation markers in embryoid bodies derived**  
1142 **from wt or Mtgr1<sup>-/-</sup> ESCs.** (A). RT-qPCRs analyses of indicated markers in EBs induced from wt or  
1143 *Mtgr1*<sup>-/-</sup> ESCs and cultured for four days in suspension after LIF withdrawal; shown relative to wild  
1144 type mESCs cultured under serum+LIF conditions. (B). RT-qPCRs analyses of select markers in EBs  
1145 induced from wt or *Mtgr1*<sup>-/-</sup> ESCs and cultured for four days in suspension after LIF withdrawal;

1146 shown relative to wild type mESCs cultured under serum+LIF conditions.

1147

1148 **Figure 3–figure supplement 6. Loss of *Mtgr1* results in defect in PGC-LC induction.** (A). Imaging  
1149 of fixed mPGC-LCs on Day6 of differentiation using bright field, DAPI (stained with Hoescht) and  
1150 GFP channels. Imaging was done using z-stacks and maximum projection was taken afterwards. (B).  
1151 FACS on the *Stella*:GFP line in ES state or PGC-LC state as indicated. Wild type cells or *Prdm14*-null,  
1152 *Mtgr1*-null or respective rescue lines were used.

1153

1154 **Figure 4–figure supplement 1. Sequences of monobodies selected against human PRDM14 and**  
1155 **their  $K_D$  values to hPRDM14 measured in yeast display format.** Wild type monobody sequence is  
1156 shown on top with residues that were permuted highlighted in red. Two types of libraries were used,  
1157 loop and side library.

1158

1159 **Figure 4–figure supplement 2. Human PRDM14 can substitute for the mouse *Prdm14* in mESCs.**  
1160 (A) Morphological changes in *Prdm14*<sup>-/-</sup> mESCs and those reconstituted with mouse or human FH-  
1161 PRDM14 cDNA, respectively. Note that all lines were grown under serum+LIF conditions. (B)  
1162 Western Blot showing that both human and mouse *Prdm14* proteins were expressed at similar levels in  
1163 the reconstituted lines. (C). RT-qPCRs measuring changes in gene expression in lines reconstituted with  
1164 human or mouse *Prdm14* relative to the *Prdm14*<sup>-/-</sup> line. Genes marked with an asterisk denote direct  
1165 mouse *Prdm14* targets from ChIP-seq.

1166

1167 **Figure 4–figure supplement 3. Monobody affinity pulldown of the endogenous *Prdm14* protein.**  
1168 (A). Co-precipitation of *Mtgr1* with endogenous *Prdm14* from wt mESC extracts via affinity pulldown  
1169 with Mb(S4), Mb(S14) or, as a negative control, beads only followed by detection of *Mtgr1* by Western  
1170 blotting. FT, flow through; W, wash; E, elution. (B). Silver staining of polypeptides recovered in the

1171 affinity pulldown with Mb(S4) from wt mESC extracts. (C). Mass spectrometry analysis of proteins  
1172 recovered in the affinity pulldown with Mb(S4) from wt mESC extracts. Results of the top recovered  
1173 and correctly matched proteins (prob > 5.00; prob – log probability of the correct fit to the data) are  
1174 shown. Also shown are total intensity, number of spectra and number of unique peptides recovered,  
1175 percent coverage and number of amino acids in the recovered protein.

1176

1177 **Figure 5–figure supplement 1. Inclusion of linker does not affect properties of Prdm14-Mtgr1**  
1178 **complex.** (A) Elution profiles of Prdm14-linker-Mtgr1 (in black) and Prdm14 mixed with Mtgr1 in 1:2  
1179 molar ratio (in red) from a Superdex75 size-exclusion column. (B) Overlay of <sup>15</sup>N-<sup>1</sup>H HSQC spectra of  
1180 <sup>15</sup>N-Prdm14-linker-Mtgr1 complex (black) with <sup>15</sup>N-labeled Prdm14 in complex with unlabeled Mtgr1  
1181 (red). (C) Titration curves for binding of Prdm14/Mtgr1 complex and Prdm14-linker-Mtgr1 to Mb(S4).  
1182 The error bars shown on each data points are the s.d. from triplicate measurements. The curves show  
1183 the best fit of the 1:1 binding model. The errors for the  $K_D$  values are the s.d. from triplicate  
1184 measurements.

1185

1186 **Figure 5–figure supplement 2. Structural features of the Prdm14-Mtgr1-Mb complex.** (A)  
1187 Crystals of Prdm14-linker-Mtgr1/Mb(S4) complex obtained in 16% PEG3350, 8% Tascimate pH 5.3.  
1188 (B) The Prdm14-monobody interface. Prdm14 is shown in white surface representation with the  
1189 interface residues on Prdm14 shown in yellow and are labeled in black. Mb(S4) is shown in cartoon  
1190 representation in cyan and are labeled in red. The upper box illustrates prominent features in the  
1191 interface involving residues in the FG loop of the monobody. Trp81 of Mb(S4) is in the pocket formed  
1192 by Pro250, Phe270, Val268, Asp262 and Ala265 of Prdm14. Gln75, Tyr77 and Trp81 of Mb(S4) form  
1193 hydrogen bonds with Prdm14 Arg269, Asp262 and Glu251, respectively. Hydrogen bonds and salt  
1194 bridges are shown in black dotted lines. The bottom box shows a salt bridge formed between Prdm14  
1195 Glu340 and Lys47 in the  $\beta$ -strand C of Mb(S4). (C) Unbiased feature enhanced map showing electron

density at 1 sigma level for residues involved in Prdm14-Mtgr1 interaction (D) Crystal contacts between monobodies and Prdm14 from neighboring molecules. The symmetry related molecules are shown with Prdm14-linker-Mtgr1 shown in yellow and monobody shown in cyan (E) Binding of Mb(S4) and its point mutant, Trp81 to Ala, to biotinylated Prdm14 immobilized on streptavidin M280 beads, as measured using the bead-based binding assay. Mutation of Trp81 results in substantial loss of Prdm14 binding, confirming the interface between Prdm14 and Mb(S4) shown in panel B.

**Figure 5—figure supplement 3. Comparison of the structure of the Prdm14-Mtgr1 complex with that of the Prdm9-histone H3 peptide-S-adenosyl-L-homocysteine (AdoHcy) complex.** (A) Left, the crystal structure of the Prdm9 complex (PDB ID 4C1Q). Prdm9 is shown in cartoon representation in gray and the histone peptide is shown in sticks in blue. AdoHcy is shown in green sticks. Right, cartoon representation of Prdm14 in gray with Mtgr1 shown in pink ribbon representation. The residues in pre-SET region of Prdm14 that occupy the region equivalent to the AdoHcy binding site in Prdm9 are shown in orange. (B) Left, surface representation of Prdm9 in which the surfaces for the atoms within 4.5 Å from the peptide are marked in blue. Right, Mtgr1-interacting residues of Prdm14 are shown in pink. (C) Left, residues in the Prdm9 catalytic site (labeled in black) and the Lys4me2 of the histone peptide (labeled in red), shown in sticks. Tyr357 is directly involved in the catalytic activity of Prdm9. Right, the same region of Prdm14. Phe273 and Tyr339 in Prdm14 are respectively equivalent to Tyr276 and Tyr341 in Prdm9. Lys109 of Mtgr1 present in proximity to the Prdm14-Tyr355 (equivalent to Tyr357 in Prdm9) is shown. Also shown are Pro366 and Glu294 that have atoms within 4.5 Å of Mtgr1-Lys109. Prdm14 residues are labeled in black. K109 of Mtgr1 is labeled in red.

**Figure 6—figure supplement 1. Gene expression analyses of Prdm14-/- cells reconstituted with wt or E294K Prdm14** (A). Principal component analysis of RNA-seq expression comparisons of the top 898 genes with the highest median average deviation in indicated ESC lines. (B). Heatmap displaying



1221 top 50 variable genes between *Mtgr1*<sup>-/-</sup> (3 clones), *Prdm14*<sup>-/-</sup> (2 clones) mES cells or *Prdm14*<sup>-/-</sup> clones  
1222 reconstituted with wild type or mutant (E294K) Prdm14 protein.

1223

1224 **Figure 6—figure supplement 2. Inhibition of Prdm14-Mtgr1 Interaction Affects Germ Cell**  
1225 **Development.** (A). FACS plots of the gated quantification of GFP signal shown in Figure 6D as a  
1226 measure of mPGC-LC induction from wild type cells, *Prdm14*<sup>-/-</sup> cells or *Prdm14*<sup>-/-</sup> clones  
1227 reconstituted with wt or mutant (E294K, Y339R) Prdm14. (B). FACS plots of the gated quantification  
1228 of GFP signal shown in Figure 6F as a measure of mPGC-LC induction from three independent  
1229 populations of mCherry cells or mCherry-Mb(S14) cells with and without addition of doxycycline.

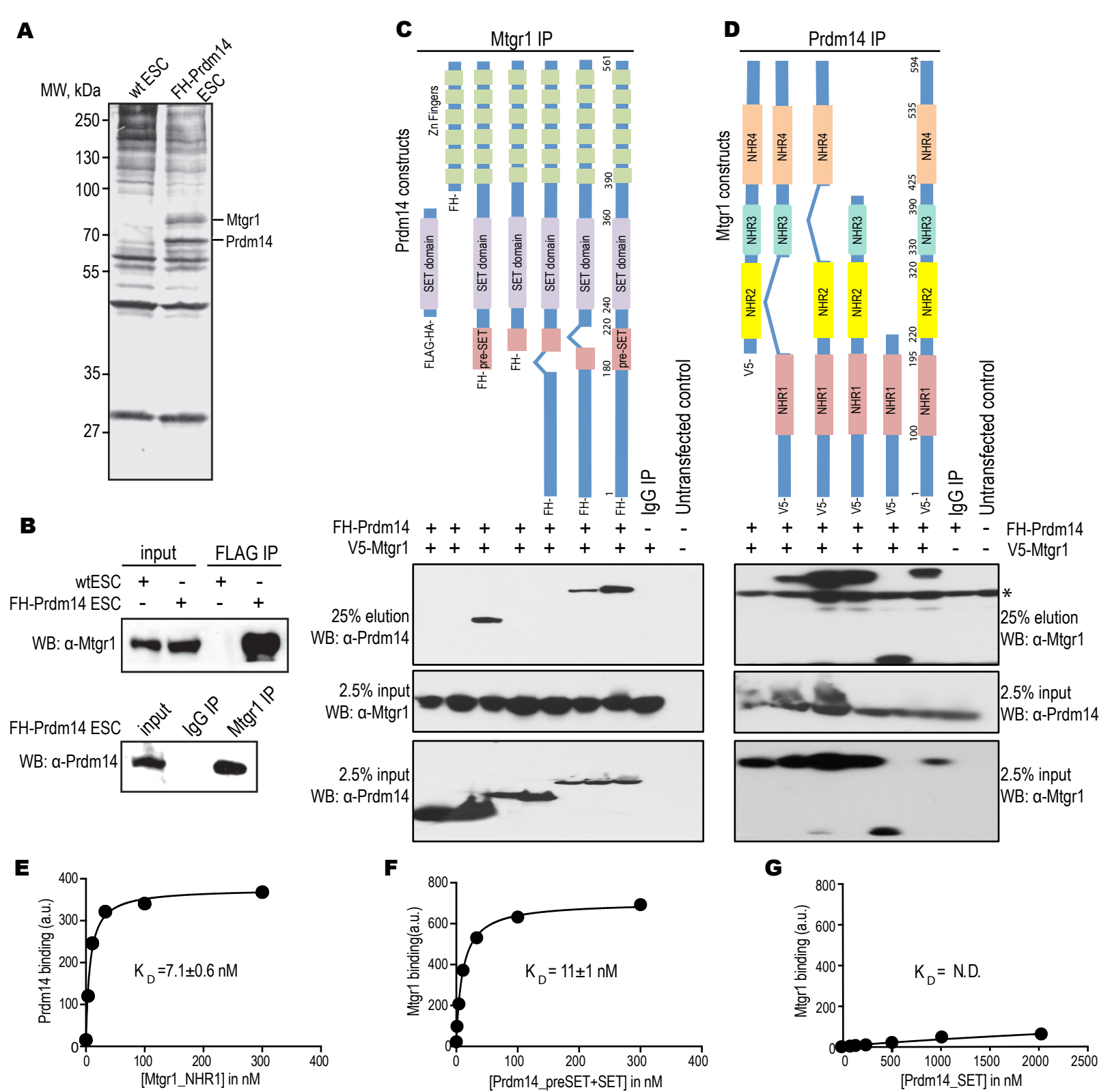
1230

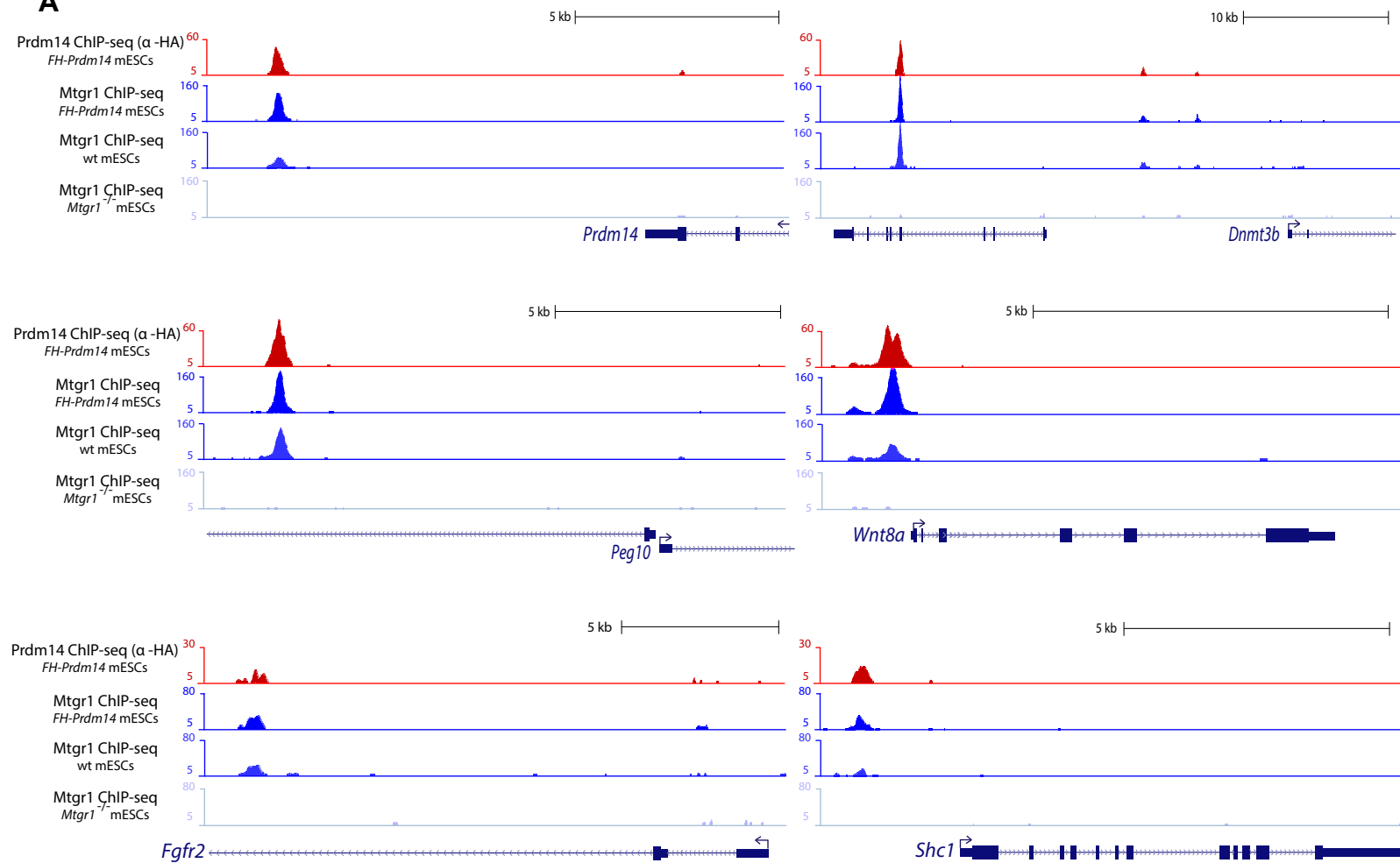
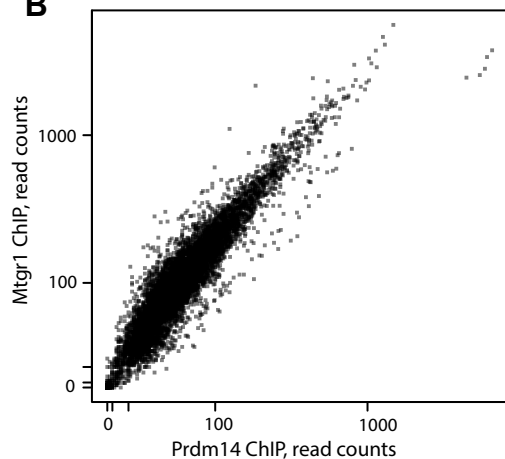
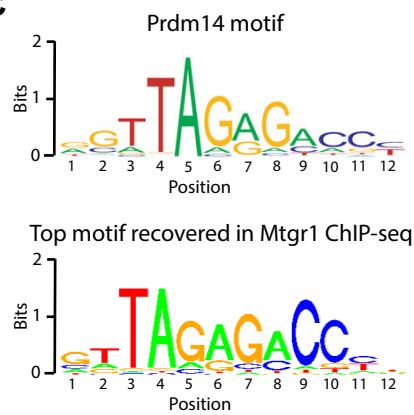
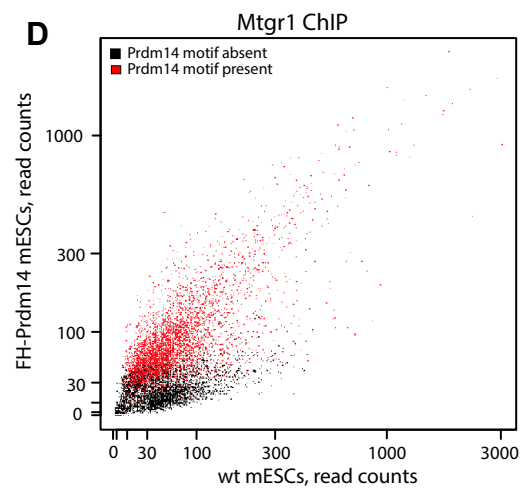
1231

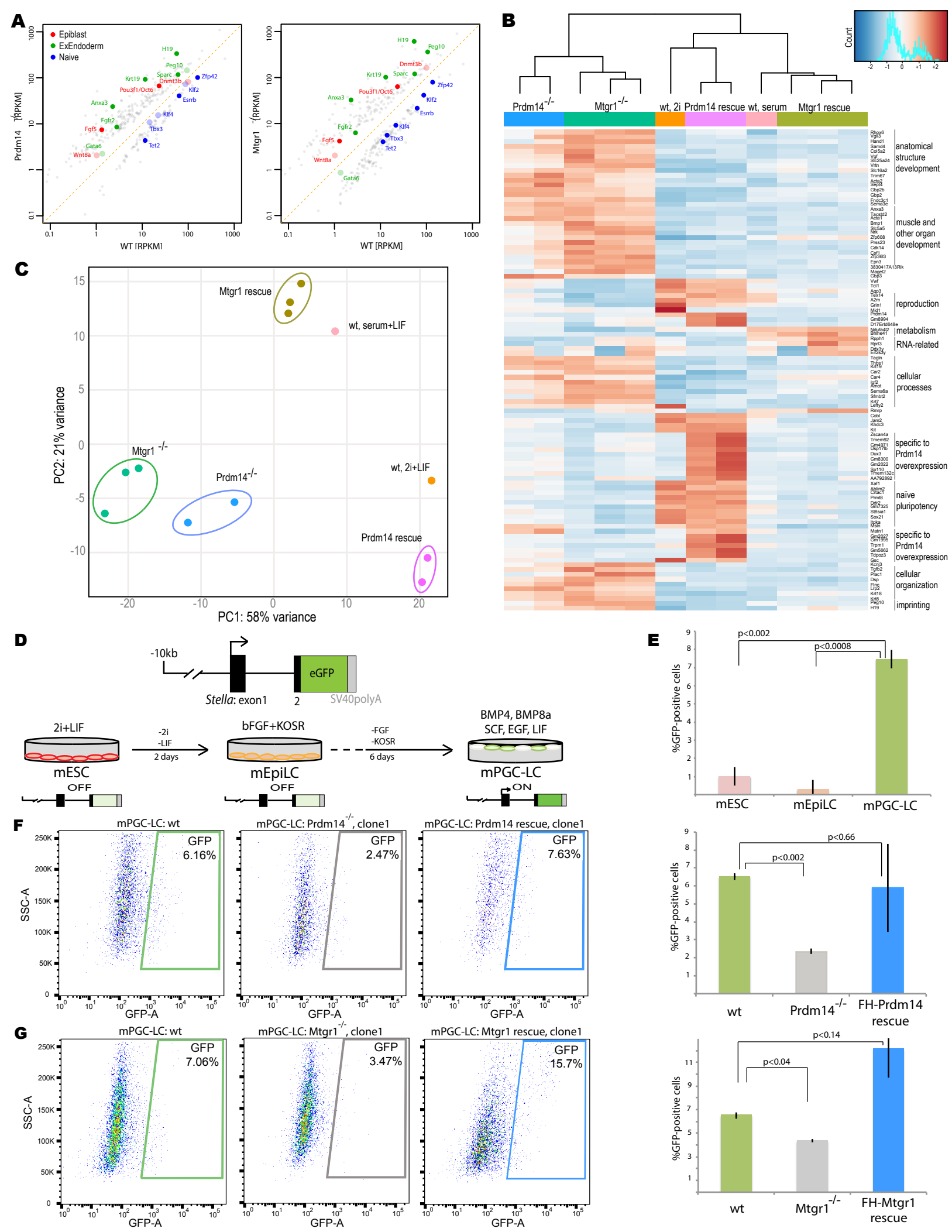
1232 **Figure Source data**

1233

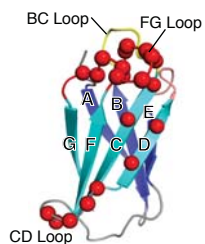
1234 Figure 1 – source data 1. Supplementary Table1. List of proteins recovered from the Prdm14 IP-MS  
1235 experiment.



**A****B****C****D**

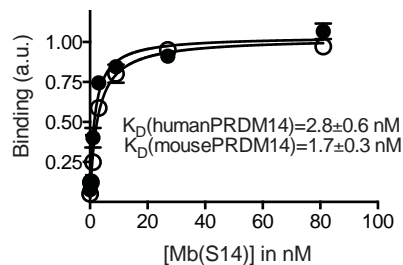
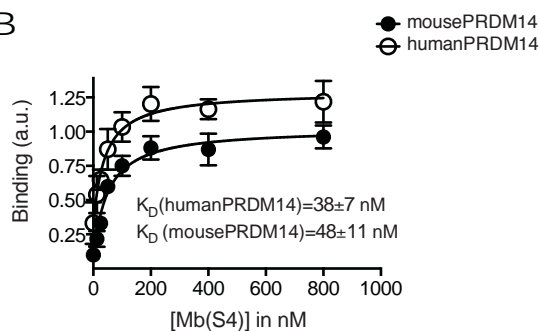


A

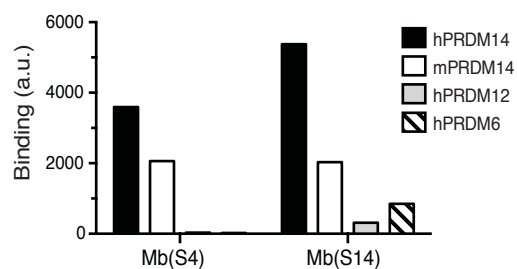


Clone	Amino acid sequence									
	10	20	βC	40	CD	βD	50	60	70	FG
library	VSSVPTKLEVVAATPTSLLISWDAPAVTVOUYOITYGETG(X <sub>5-6</sub> )-QZFZVPGSKSTATISGLSPGVDYTITVYA(X <sub>7-13</sub> )-----SPISINYRT									
Mb(hPRDM14_S14)	VSSVPTKLEVVAATPTSLLISWDAPAVTV <b>DHYR</b> ITYGETG <b>VGSYYH</b> Q <b>EF</b> TVPGSKSTATISGLSPGVDYTITVYA <b>VGWYYPGW</b> HGYTRSPISINYRT									
Mb(hPRDM14_S4)	VSSVPTKLEVVAATPTSLLISWDAPAVTV <b>VLYH</b> ITYGETG <b>GNSPV</b> -Q <b>EF</b> TVPGSKSTATISGLSPGVDYTITVYA <b>TYRLWGSWQYY</b> SSPISINYRT									
Mb(hPRDM14_S10)	VSSVPTKLEVVAATPTSLLISWDAPAVTV <b>VYYH</b> ITYGETG <b>GNSPV</b> -Q <b>AFK</b> VPGSKSTATISGLSPGVDYTITVYA <b>TYRLWGSWQYY</b> SSPISINYRT									
Mb(hPRDM14_S5)	VSSVPTKLEVVAATPTSLLISWDAPAVTV <b>IFYV</b> ITYGETG <b>GNSPV</b> -Q <b>AFK</b> VPGSKSTATISGLSPGVDYTITVYA <b>QGYQSRWYRN</b> ---SPISINYRT									
Mb(hPRDM14_S2)	VSSVPTKLEVVAATPTSLLISWDAPAVTV <b>DHYV</b> ITYGETG <b>GNSPV</b> -Q <b>EF</b> TVPGSKSTATISGLSPGVDYTITVYA <b>GHFWYGYQYW</b> ---SPISINYRT									

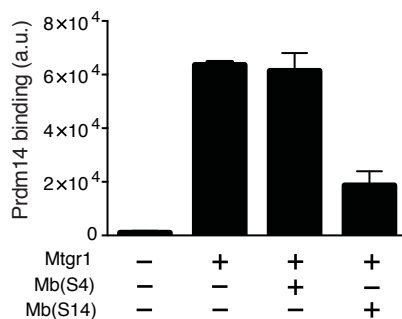
B



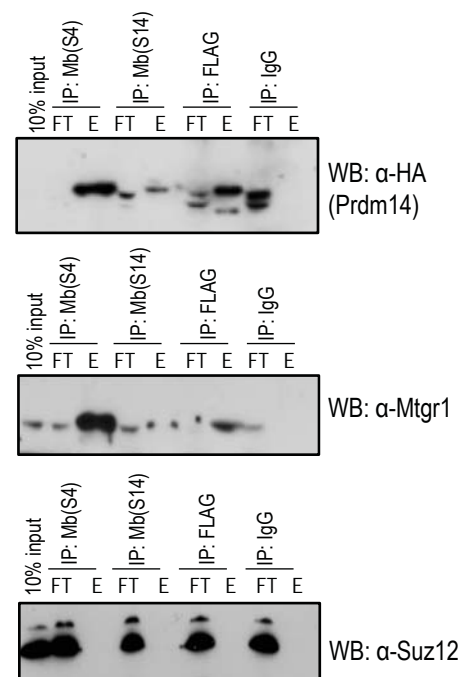
C

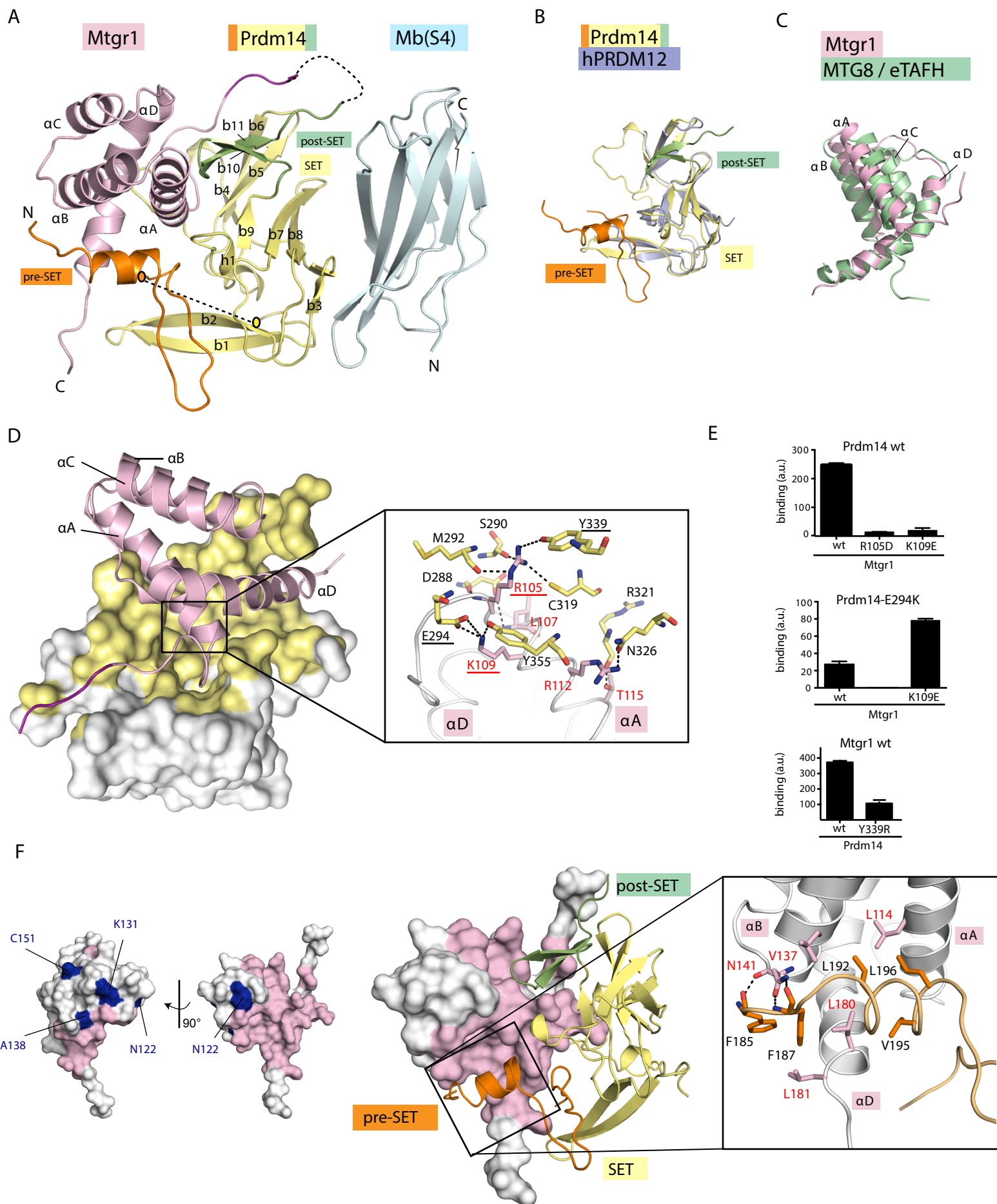


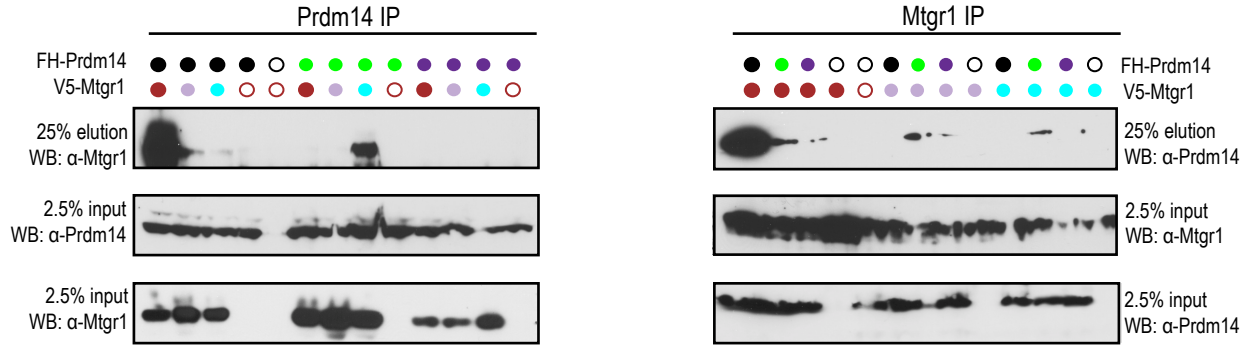
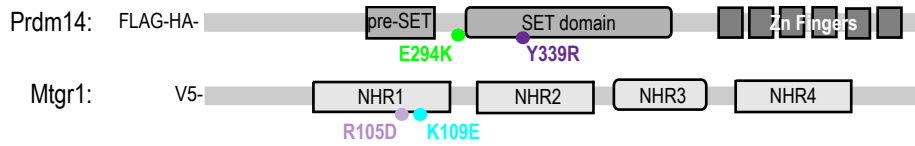
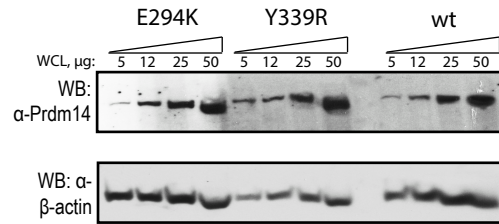
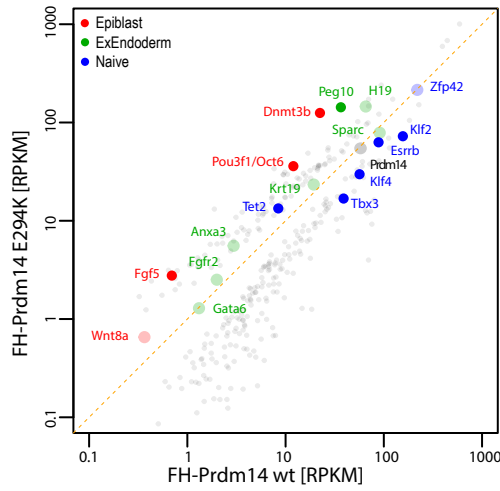
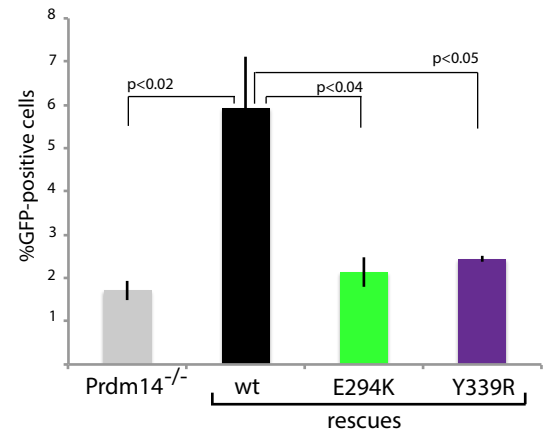
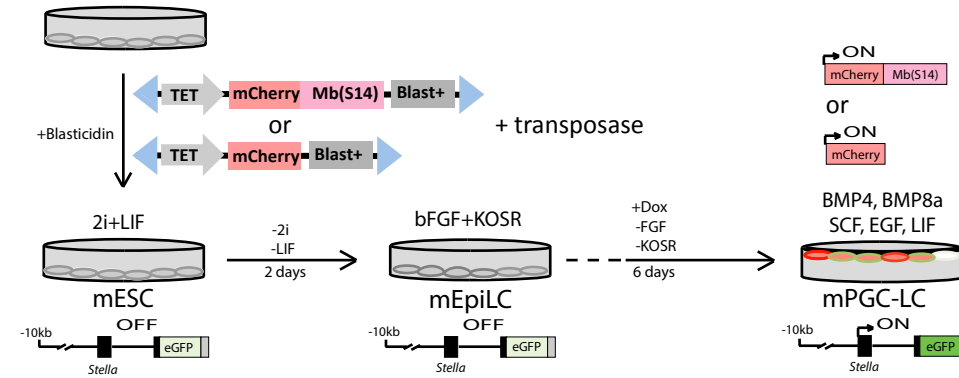
D



E





**A****B****C****D****E****F**