A non-catalytic role for MLL2 in controlling chromatin organisation and mobility during the priming of pluripotent cells for differentiation

Authors:

Maike Steindel[1,2§], Oliver Davis[1,3,4§], Katrin Neumann[5], Liviu Pirvan[1], Gökçe Agsu[1,4], Andrea Kranz[5], Dwaipayan Adhya[1,6], Jorg Morf[1], Shuyue Yang[4], Ziwei Zhang[7], Jun Fu[5,11], Melania Barile[1], Annabelle Wurmser[1], Stanley E. Strawbridge[1,2,13], Irina Mohorianu[1], Pradeepa Madapura[8], Brian Huntly[1], Berthold Göttgens[1], David Holcman[9], Shamith A. Samarajiwa[10], David Klenerman[7], Konstantinos Anastassiadis[5], A. Francis Stewart[5,11,12]*, Srinjan Basu[1,2,4]*

[1] Cambridge Stem Cell Institute, University of Cambridge, Cambridge, CB2 0AW, UK.

[2] Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 3DY, UK.

[3] John van Geest Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, Cambridge, CB2 0QQ, UK

[4] Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

[5] Center for Molecular and Cellular Bioengineering, Technische Universität Dresden, Germany

[6] Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, UK

[7] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

[8] The Blizard Institute, School of Medicine and Dentistry, Queen Mary University of London, London, UK

[9] Department of Applied Mathematics and Computational Biology, Ecole Normale Supérieure, 75005 Paris, France

[10] Department of Metabolism, Digestion and Reproduction, Division of Systems Medicine, Imperial College London, London, W12 0NN, UK.

[11] Key State Laboratory for Microbial Technology, Shandong University Qingdao, China

[12] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia

[13]Current Address: Centre for Stem Cell Biology, School of Biosciences, University of Sheffield, Sheffield, S10 2TN, UK.

[§] These authors contributed equally to the work

* Correspondence:
Srinjan Basu srinjan.basu@imperial.ac.uk (lead contact)
Francis Stewart francis.stewart@tu-dresden.de

Author contributions

Conceptualization: SB and AFS. **Data curation**: LP, MS, AFS and SB. **Formal Analysis:** (scRNAseq) OD, JM, DA, SB, MB and BG. (Micro-C) LP, SB and SAS. (Single-molecule localisation microscopy, SMLM) SY, MS and SB. (Quantitative immunofluorescence) MS. **Funding acquisition:** SB, MS and AFS. **Investigation:** (Neuroectoderm differentiation, ChIP-qPCR) KN, AK, KA and AFS. (scRNAseq) JM and MS. (Micro-C) MS with help from AW and GA. (SMLM) MS and GA. **Methodology:** No new methods used. **Project administration:** SB, BH and AFS. **Resources:** (*Mll2* knockout, *Mll2-GFP* and *Mll2* NCat cell lines) KN, AK, JF, KA and AFS. (Cell lines expressing HaloTag-H2B and dCas9-GFP) MS and GA. (SMLM instrument alignment and maintenance) ZZ and DK. **Software:** (SMLM) SES, AW, SY and DH. **Supervision:** SB, BH and AFS. **Validation:** (Quantitative immunofluorescence) MS, OD and SB. **Visualization:** (scRNAseq and Micro-C) DA, OD, JM and LP. (Immunofluorescence) KN, AK and MS. (SMLM) MS and SY. **Writing – original draft:** SB and AFS. **Writing – review & editing:** MS, OD and AW.

Acknowledgements:

We thank all the members of the Basu and Stewart labs, and Michelle Meredyth, for insightful and critical discussions. We thank the Genomics Core Facility, Cancer Research UK Cambridge Research Institute for 10x Chromium single cell library preparation and Illumina sequencing. We thank the Cambridge Stem Cell Institute (CSCI) Genomics Core Facility for assisting with quality control and submission of sequencing libraries. We thank Peter Humphreys, Darran Clements and Louis Elfari in the CSCI Imaging Facility for training students and assisting with experiments. We thank Sally Lees and the CSCI Tissue Culture Core Facility for technical advice and support. We thank Arash Shahsavari and Elze Lauzikaite in the CSCI Bioinformatics Core Facility for help processing/analysis of some of the sequencing datasets. The PBase and dCas9-GFP plasmids were kind gifts from Dr Brian Hendrich and Prof Joanna Wysocka. We also thank Menan Loganathan and Igor Orsine de Almeida for preliminary analyses.

MS and AW were funded by the Wellcome Trust PhD programme in Stem Cell Biology and Medicine (224929/Z/22/Z). OD was funded by the Addenbrooke's Charitable Trust and the UKRI Medical Research Council (MC_PC_17230). LP and GA were funded by the UKRI

Biotechnology and Biological Sciences Research Council (BB/W000423/1, BB/W000423/2). AK, KA and AFS were funded by Deutsche Forschungsgemeinschaft grants (STE 903/12-2 and STE 903/13-1). DA was funded by the Simons Foundation Autism Research Initiative. JM and SB were funded at the Cambridge Stem Cell Institute through a starting grant from the Wellcome Trust (203151/Z/16/Z) and UKRI Medical Research Council (MC_PC_17230). SB was also funded by the Trinity College Stem-Cell Medicine Senior Postdoctoral Researcher Fund and Imperial College London. SES was funded by the Joint Research Grant (Isaac Newton Trust, Wellcome Trust ISSF, University of Cambridge), the CSCI Wellcome Trust seed fund and the Wellcome Trust (Sir Henry Wellcome Fellowship).

Abstract

The chromatin regulator MLL2 (KMT2B) is the primary histone 3 lysine 4 (H3K4) trimethyltransferase acting at bivalent promoters in embryonic stem cells (ESCs) and is required for differentiation toward neuroectoderm. Here, we demonstrate that this requirement occurs during exit from naïve pluripotency, days before neuroectoderm differentiation is impaired. During exit, the effect of MLL2 on transcription is subtle, increasing the expression of a few important neuroectodermal transcription factors. In contrast, MLL2's effect on chromatin architecture is substantial, stabilising loops associated with bivalent promoters in primed ESCs. MLL2 H3K4 catalytic activity is dispensable for stabilising these loops during ESC exit and for neuroectoderm differentiation. We therefore identify a non-catalytic function for MLL2 in stabilising 3D chromatin architecture, which has implications for lineage specification. Because MLL2 shares features with all four MLLs, we propose that chromatin tethering, rather than H3K4 methylation, represents a primary function for MLLs during lineage commitment decisions.

Introduction

Multicellularity is anchored by the capacities of stem cells to utilize the information encoded in the genome in alternative ways. These capacities involve the hierarchical regulation of gene expression by transcription factors coupled with epigenetic hierarchies, including alterations in 3D chromatin organisation (Hubner, Eckersley-Maslin, and Spector 2013; Bompadre and Andrey 2019; Lackner et al. 2021; Jaenisch and Bird 2003; Marks et al. 2012). Together these mechanisms elicit the multiple utilizations of the genome required for multicellularity. Unravelling these molecular mechanisms is key to understanding development, disease, and the therapeutic potential of stem cells for regenerative medicine.

The differentiation of mammalian pluripotent cells (PCs), including embryonic stem cells (ESCs), both *in vitro* and *in vivo*, has revealed a multi-step process in which naïve PCs generate primed PCs before differentiation toward primary cell lineages such as neuroectoderm. *In vivo*, naïve PCs emerge in the epiblast of pre-implantation embryos (between embryonic day E3.75 and E4.75 in the mouse (Boroviak et al. 2014)) with the potential to differentiate into all cell lineages of the mammalian body (Silva and Smith 2008). PCs transition from a naïve to a primed pluripotent state as the embryo implants (between E4.5 and E6.5) (Nichols and Smith 2009; Hackett and Surani 2014). *In vitro*, ESCs derived from the inner cell mass of pre-implantation blastocysts are equivalent to naïve PCs when cultured under defined conditions called 2iLIF (Ying et al. 2008). Removal of 2iLIF promotes exit from naïve pluripotency (Kalkan and Smith 2014; Kalkan et al. 2017; Strawbridge et al. 2020). Subsequent addition of FGF2 and activin promotes priming to post-implantation epiblast-like cells (EpiLCs) (Hayashi et al. 2011; Buecker et al. 2014; Yang, Humphrey, et al. 2019). EpiLCs can then be differentiated towards specific lineages such as neuroectoderm following specific inductive cues (Conti et al. 2005).

During the initial priming phase as PCs exit from self-renewal, the enhancers and/or promoters of lineage-specifying genes become accessible, acquire specific chromatin marks (DNA hypomethylation, H3K4me1/3, H3K27ac/me3) (Argelaguet et al. 2019; Eckersley-Maslin et al. 2020; Grosswendt et al. 2020; Ashokkumar et al. 2020) and alter their 3D genome architecture (Joshi et al. 2015; Chovanec et al. 2021). Depleting proteins responsible for these changes perturbs differentiation, supporting the proposition that early changes in the chromatin landscape prime PCs towards specific lineages (Argelaguet et al. 2019; Eckersley-Maslin et al. 2020; Grosswendt et al. 2020; Ashokkumar et al. 2020; Chovanec et al. 2021).

Amongst chromatin regulators, MLL1-4 (also termed KMT2A-D) play central roles in embryonic development, differentiation and stem cell maintenance in a range of mammalian

cell lineages (Crump and Milne 2019). In particular, MLL1 is required for three types of adult stem cell in mammals: hematopoietic stem cells, intestinal stem cells and satellite cells in skeletal muscle (Ernst et al. 2004; Goveas et al. 2021; Addicks et al. 2019). Notably, the MLLs are histone 3 lysine 4 methyltransferases (H3K4 MTs). Despite the relevance of H3K4 methylation to epigenetic regulation (Talbert and Henikoff 2021), the methyltransferase activities of the MLLs are not always required for the critical functions played by these proteins (Mishra et al. 2014; Mas et al. 2018; Dorighi et al. 2017). In all four MLLs, the C-terminal SET methyltransferase domain is only a small part of these large nuclear proteins (MLL2 is ~300 kDa). Notably, MLLs also contain arrays of chromatin binding zinc fingers (PHD and CXXC), intra- (FYRN, FYRC) and intermolecular protein-protein interaction domains including a CBP/p300 transactivation region, interspersed amongst extensive intrinsically disordered regions (IDRs) (Crump and Milne 2019). Indeed, in higher eukaryotes the MLLs and orthologues are all large proteins, even though most of them are IDRs that are highly variable. The MLLs clearly convey functions beyond their role as methyltransferases that warrant further investigation.

Here, we address how MLL2 and its catalytic activity influence neuroectoderm differentiation. Although dispensable for self-renewal of naïve ESCs, MLL2 is required for the coordination and timing of differentiation decisions (Glaser et al. 2006; Lubitz et al. 2007; Glaser et al. 2009). MLL2 is the H3K4 MT primarily responsible for H3K4me3 deposition at bivalent promoters in ESCs. Accordingly, MLL2 is bound at most bivalent promoters. However, it is also bound at most active promoters and is not required for the vast majority of inducible gene expression (Denissov et al. 2014). Notably, MLL2 opposes silencing by Polycomb-Group action and DNA methylation (Douillet et al. 2020; Glaser et al. 2009; Ladopoulos et al. 2013) and also influences DNA accessibility and 3D genome organisation (Mas et al. 2018). But what is the contribution of MLL2 and its catalytic activity to neuroectoderm differentiation? To address this, we combined single-cell transcriptomics, chromosome conformation capture (Micro-C) and single-molecule localisation microscopy (SMLM) to establish that MLL2 functions during the transition of naïve ESCs to primed ESCs and stabilises enhancer-promoter loops, particularly at bivalent genes, with this process not dependent on H3K4 trimethylation. This stabilisation appears to be MLL2's primary function in early neuroectoderm lineage determination.

Results

The requirement for MLL2 in neuroectoderm differentiation is exerted during ESC exit from naïve pluripotency

A requirement for MLL2 in neuroectoderm differentiation has previously been identified (Lubitz et al. 2007). To determine when MLL2 is required, we used ligand-regulated conditional mutagenesis (Logie and Stewart 1995) and *Mll2* conditional knock-out (cKO) ESCs (Denissov et al., 2014) to delete *Mll2* at different timepoints during priming and neuroectoderm differentiation (**Figure 1A-D and S1**). This revealed that the requirement for MLL2 is exerted in the earliest possible window during ESC exit, about three days and >6 cell cycles before the emergence of the differentiation defect (**Figure 1D**). To ensure specificity, ectopic expression of GFP-tagged MLL2 from a BAC (bacterial artificial chromosome) transgene rescued neuroectoderm differentiation of *Mll2* cKO cells (**Figure 1B-E and S1B-C**). Because tamoxifen-induced cKO led to loss of MLL2 protein in ESCs within 48 hours (**Figure S1A**), MLL2 must be acting within the first 4 cell cycles during the ESC exit from self-renewal to influence subsequent neuroectoderm differentiation. Notably, cells deleted for *Mll2* at the EpiLC stage (**Figure 1F**) were unimpaired in neuroectoderm differentiation, again indicating that MLL2 is no longer required after ESC exit and exerts its function either during ESC self-renewal, exit, and/or priming several days before the major defects in neuroectodermal differentiation are observed.

Both wild-type and *Mll2* cKO ESCs transition from naïve to primed pluripotency via an intermediate stage with features of formative pluripotency

The tamoxifen cKO time course indicated that, for neuroectoderm differentiation, MLL2 is required before the ESC transition from naïve to primed pluripotency is completed. To gather further evidence about this transition, we used single-cell RNA-seq (scRNA-seq) to compare transcription in wild-type (control) and *Mll2* cKO cells every 24 hours for 3 days during the priming transition (**Figure 2A**). scRNA-seq was employed because priming occurs through intermediate cell states with some asynchronicity (Smith 2017; Kinoshita et al. 2021; Du et al. 2018; Neagu et al. 2020). In total, 8737 single-cell transcriptomes (4597 *Mll2+/+,* 4140 *Mll2-/-*) passed stringent quality control measures (**Figure S2A**). As replicates varied in sequencing depth/cell, with more unique transcripts and genes detected per cell from replicate 1 than from replicate 2 (**Figure S2B-C**), high-confidence conclusions were drawn only when both replicates showed similar trends.

Using these datasets and by conducting cluster stability tests, we identified 7 major cell clusters (**Figure S2D-G**) (Shahsavari, Munteanu, and Mohorianu 2022). Both wild-type and *Mll2* cKO cells contributed to all 7 clusters (**Figure 2A-B**). Although only one major cell population was detected for naïve ESCs at 0 hours ("0hrs"), there was an increase in co-existing cell clusters during priming with two dominant clusters at 24 and 48 hours ("24hrs A/B" and "48hrs A/B"), and one major cluster at 72 hours ("72hrs") (**Figure 2A-B**). One unassigned

cluster, found at all time points, was excluded from further analysis due to its high levels of apoptotic genes ("unassigned") (**Figure 2B and S3A**). Clustering was not driven by the cell cycle: although subtle cell cycle differences were observed between wild-type and *Mll2* cKO cells (**Figure S3B**) regressing out cell cycle factors did not affect clustering (**data not shown**).

To functionally assign the clusters, we calculated the Spearman correlation of all genes assigned to the second highest expression quartile (**Figure S3C**). Clusters at 0hrs and 48-72hrs were consistent with prior knowledge of the priming transition, expressing naïve pluripotency factors *Nanog* and *Klf4* at 0hrs and primed genes *Otx2* and *Pou3f1* at 48-72hrs (**Figure 2C and S3D-F**) (Chen et al. 2008; Hayashi et al. 2011; Dunn et al. 2014; Kalkan and Smith 2014; Buecker et al. 2014; Kalkan et al. 2017). Due to these similarities, we concluded that the 0hr cluster represented "naïve" ESCs and the 48-72hrs clusters represented "primed" ESCs. The two 24hrs clusters expressed a unique set of genes, showed strong correlation coefficients (>0.8) and correlated similarly to preceding and later time points (**Figure 2C and S3D-F**), suggesting that these 2 cell states were different to both naïve and primed ESCs. Gene expression profiles of the two 24hrs clusters were characterised by increased expression of a unique set of transcription factors (e.g. *Zic3, Tcf15*, *Nr0b1, Dppa2, Dppa4*) and chromatin regulators (e.g. *Rbpj, Jarid2, Dnmt3l*) (**Figure 2C and S3D-F**). *Zic3, Rbpj* and *Jarid2* play a role in the exit from naïve pluripotency (Leeb et al. 2014; Lackner et al. 2020; Yang, Andrabi, et al. 2019), suggesting that cells must transition through this intermediate phase to form primed PCs (48-72hr). Indeed, UMAP plots generated for RNA trajectory analysis (after excluding the "unassigned" cluster) confirmed that this was the case (**Figure S3G**).

To validate our clusters, we compared our dataset to published *in vivo* and *in vitro* datasets. *In vivo* datasets (Mohammed et al. 2017) revealed that our transition mimics the *in vivo* pre-to-post-implantation embryo transition from the E3.5 to E6.5 epiblast (**Figure S3H**). Next, we assessed whether our 24hrs clusters were linked to the hypothesis of a transient state between naïve and primed pluripotency called formative pluripotency, a state with higher multi-lineage potential than both naïve and primed PCs (Smith 2017). Indeed, expression of *Dppa2/*4 and *Tcf15* during this transition have previously been shown to correlate with higher multi-lineage potential (Davies et al. 2013; Eckersley-Maslin et al. 2020), lending support to this hypothesis. We therefore compared our data to cells recently captured *in vitro* that have features of formative pluripotency (Kinoshita et al. 2021) and found that they were a mixture of our 24hrs and 48hrs clusters (**Figure S3I**). These comparisons suggested that our datasets were relevant to *in vivo* cell types and that our 24hr and 48hr clusters had features of formative pluripotency.

MLL2 has little impact on the priming transition of pluripotent cells

Cell cluster identities at 72hrs remained largely unchanged in *Mll2* cKO cells compared to control cells ($p > 0.1$) (**Figure 2B**), indicating that MLL2 had only a slight impact on steady state mRNA expression levels early on, which is consistent with previous studies (Lubitz et al. 2007; Glaser et al. 2009; Mas et al. 2018). It is therefore unlikely that the defects in neuroectoderm differentiation arose from a loss of the primed ESC programme despite the requirement for MLL2 during this period.

Because even small changes in gene expression could have later effects on differentiation, we conducted a detailed analysis of MLL2-dependent changes in gene expression during the transition. With good correlation between $\log_2$FC changes in the two replicates (**Figure S4A**), we identified 1047 differentially expressed genes (DEGs) with 530 up-regulated and 517 down-regulated genes ($p<0.001$) (**Figure 2C-D** and **S4B-C**). To identify direct targets of MLL2, we utilized the published MLL2 ESC ChIP-seq dataset (Denissov et al. 2014) to filter for genes that had MLL2 bound within 1 kb of their transcription start site (**Figure S4D**). This revealed that DEGs had a significantly higher likelihood of being MLL2 bound (84% vs 36%; Chi-squared test: $p = 10^{-267}$) (**Figure S4E**) and that there were a significantly higher number of MLL2 peaks within 1 kb of the gene transcription start site (TSS) of these genes (average peak number of 1.1 peaks vs 0.5 peaks; Mann Whitney Test: $p = 10^{-273}$) (**Figure S4F**). It also showed that 90% of MLL2-dependent genes had MLL2 bound: 947 genes out of the 1047 MLL2-dependent DEGs were MLL2-bound. We then separated MLL2-bound DEGs ($p<0.001$) by cluster and observed cluster-specific changes: more up- than down-DEGs in the 0hr naïve ESC cluster, similar numbers of up- and down-DEGs in the 24hrs clusters and more down- than up-DEGs in the 48-72hrs primed ESC clusters (**Figure 2C and S5B**). To validate these changes, we examined the well-characterised MLL2-target gene *Magohb* (Glaser et al. 2009; Ladopoulos et al. 2013) and found consistent loss of *Magohb* expression in every cluster (**Figure S4F**). Overall, we conclude that MLL2 was acting directly to modulate transcription at only a few genes during the priming transition, consistent with previous ESC studies (Denissov et al. 2014; Glaser et al. 2009; Mas et al. 2018).

Comparison of the *Mll2* cKO DEGs with pathways known to play a role in pluripotency, as defined by PluriNetwork (Som et al. 2010), highlighted the transcription factors SOX2 and OTX2. OTX2 is required for activation of enhancers during the priming transition (Yang et al. 2014; Buecker et al. 2014) and SOX2 must be expressed at high levels for activation of neuroectoderm genes (Blassberg et al. 2022). MLL2 increased the transcriptional amplitude of both genes at 48/72hrs primed ESCs ($p = 10^{-4}$ for *Sox2* at 48hrs A; $p = 10^{-11}$ and $p = 10^{-9}$

for *Otx2* at 48hrs A and at 72hrs respectively) (**Figure 2D**). MLL2 was also bound at the *Sox2* and *Otx2* genes (**Figure S4D**). However, MLL2 only subtly increased transcriptional amplitude at these genes so we took *Sox2* as an example gene to determine whether the small change in *Sox2* mRNA level was enough to cause larger alterations in SOX2 protein expression. Quantitative immunofluorescence revealed that *Mll2* cKO did indeed increase SOX2 expression in naïve ESCs (1.3-fold, $p < 10^{-16}$) and decrease it in 72hrs primed ESCs (0.5-fold, $p = 10^{-14}$) (**Figure 2E-F**), comparable to changes observed at the transcript level (**Figure 2D**). Interestingly, SOX2 continued to decrease over time in *Mll2* cKO cells, suggesting that small changes at the mRNA level accumulated over time and were amplified at the protein level.

MLL2 catalytic activity is not required for neuroectoderm differentiation

Having observed minimal changes in transcription and yet dramatic changes in differentiation, we asked whether MLL2's methyltransferase activity was playing a role during this priming transition by introducing a BAC transgene into the *Mll2* cKO cell line to express a GFP-tagged MLL2 with an N2650A mutation (MLL2-NCat) that inactivates its catalytic activity (**Figure 3A and S5**). After tamoxifen induction to ablate endogenous MLL2 expression, immunoprecipitation using anti-GFP antibodies established that the MLL2-NCat expression level was similar to that of the GFP-tagged wild-type MLL2 BAC transgene used to rescue the loss of *Mll2* (**Figure S5A**). ChIP-qPCR of the *Magohb* promoter confirmed that the N2650A mutation impaired catalytic activity ($p = 10^{-4}$) (**Figure S5B**). The validated catalytically-dead MLL2 cell line was successfully differentiated into neural rosettes (**Figure 3B**). Moreover, quantitative immunofluorescence of SOX2 revealed rescue of decreased SOX2 protein level (**Figure 3C-D**). These data establish that MLL2's catalytic activity is not required for neuroectoderm differentiation.

MLL2 has a minor effect on the compartmentalisation of the 3D genome in primed PCs

Because MLL2 function in neuroectoderm differentiation does not require its methyltransferase activity, we considered other ways MLL2 can influence gene expression, including by decompacting chromatin (through the recruitment of CBP/p300) or by tethering chromatin as suggested by its large size and various chromatin binding domains. To gather evidence of the impact of MLL2 on changes in chromatin other than H3K4 methylation, we monitored 3D genome organisation in wild-type and *Mll2* cKO cells during the priming transition using chromosome conformation capture, which reveals snapshots of the 3D genome through the analysis of DNA sequences that are in physical proximity (Tolhuis et al. 2002; Dekker et al. 2002; Dostie et al. 2006; Zhao et al. 2006; Lieberman-Aiden et al. 2009; Dixon et al. 2012; Rao et al. 2014; Hsieh et al. 2020; Krietenstein et al. 2020; Hsieh et al. 2015). We first considered a previous Hi-C chromosome conformation capture study that used

the same *Mll2* cKO cells. However, this study (Mas et al. 2018) utilized a culture condition that presents an artificial epigenetic state, termed cycling ESCs, which does not distinguish between naïve and primed states (Marks et al. 2012). We therefore carried out chromosome conformation capture of both naïve and primed ESCs separately. Rather than Hi-C, we applied chromosome conformation capture using Micro-C (Hsieh et al. 2020; Krietenstein et al. 2020; Hsieh et al. 2015) to achieve a higher resolution of the 3D genome. Micro-C was carried out in duplicate for wild-type and *Mll2* cKO cells in both naïve and primed ESCs (0 hr and 72 hr), which generated >270 million uniquely mapped reads/sample (**Figure S6A**).

Analysis of genome A-B compartments (Lieberman-Aiden et al. 2009) revealed relatively large changes during the priming transition (**Figure 4A**) but smaller changes after the loss of MLL2 (**Figure 4B-C**). Although most of the genome (64%) remained in their original A or B compartments, priming provoked 27% of the genome to switch from A to B compartments with only 9% of the genome switching the other way (**Figure 4A**). There was also a 3.1-fold increase in B compartment size ($p = 10^{-8}$) relative to the smaller 1.2-fold increase in A compartment size ($p = 10^{-12}$) (**Figure 4D**). In contrast, the loss of MLL2 provoked more subtle changes with few A-B compartment changes in naïve PCs (**Figure 4B**) but with more changes in primed PCs with a 12% shift from B to A compartments and only 5% of genomic regions shifting in the opposite direction (**Figure 4C**). This was consistent with a 1.4-fold increase in A compartment size in primed PCs ($p = 10^{-4}$) with no change in B compartment size ($p = 0.4$) (**Figure 4D**). We conclude that priming involves a general shift from the A to the B compartment, consistent with DNA methylation and heterochromatinisation known to occur during the priming transition (Tosolini et al. 2018). MLL2 makes a modest contribution to this shift mainly limited to the expansion of certain B compartments in primed ESCs.

Next, we analysed changes at the scale of topologically associating domains (TADs; 500 kb to 3 Mb). Based on the TAD insulation score, which reflects the likelihood of interactions within a TAD versus outside of it, we found that priming provoked larger changes than the loss of MLL2 (**Figure 4E-G**). Specifically, we observed a decrease in TAD size during the priming transition ($p = 10^{-4}$), a decrease that was intriguingly not observed in primed PCs upon loss of MLL2 ($p = 10^{-3}$) (**Figure 4E**). We also observed an increase in TAD formation during priming, with insulation scores increasing from 0.553 to 0.733 ($p = 10^{-76}$) (**Figure 4F**). Loss of MLL2 related to only minor changes in this score, with a slight increase from 0.553 to 0.599 in naïve PCs ($p = 10^{-8}$) and a slight decrease from 0.733 to 0.706 in primed PCs ($p = 0.02$). Therefore, in agreement with the conclusion drawn by Mas et al. (2018) using cycling ESCs, MLL2 has only minor effects on TAD organization in naïve and primed PCs, although MLL2 was required to decrease TAD size during the priming transition.

MLL2 influences 3D chromatin looping

Global comparisons of A-B compartments and TADs in the replicates were highly correlated (**Figure S6B**). Therefore, we compiled replicates to increase the chromatin looping resolution, which allowed robust analysis at 40kb resolution. Previous studies indicated that priming involves changes in long-range interactions (> 1Mb) (Joshi et al. 2015; Chovanec et al. 2021) so 40kb resolution is more than sufficient to detect these changes. High-confidence 3D chromatin loops were mapped using FAN-C ($p < 0.02$), which implements HICCUPS for loop detection (Rao et al. 2014; Kruse, Hug, and Vaquerizas 2020). This revealed changes in chromatin looping at our previously identified DEGs *Sox2, Otx2* and *Magohb*, with *Sox2* and *Otx2* ranking among the top 10 genes with the greatest number of loop changes within 1 Mb of the gene promoter (**Figure 5A-C** and **S6C-I**). In wild-type conditions, similar numbers of genome-wide high-confidence loops in naïve and primed chromatin were mapped (13 425 vs 13 380 loops) (**Figure S6A**). However, priming involved changes in >40% of loops, with 5775 unique loops identified in naïve PCs and 5609 unique loops in primed PCs (**Figure 5D-F**). Loss of *Mll2* also led to a small but significant increase in overall loop size ($p = 10^{-24}$ and 10^{-48} for naïve and primed ESCs respectively) (**Figure 5G**).

To identify changes related to MLL2 binding, we used a published MLL2 ChIP-seq dataset (Denissov et al. 2014) to subset Micro-C loops with MLL2 binding sites at either end of the loop (within our 40kb resolution). This revealed 10468 MLL2-associated loops in naïve PCs (79% of total loops) and 8190 loops in primed PCs (63% of total loops) (**Figure 5D**), indicating a 22% decrease in the number of MLL2-associated loops during the priming transition. Priming also involved a change in MLL2-associated loops with 4757 unique loops identified in naïve PCs and 2479 unique loops in primed PCs (**Figure 5D**). Furthermore, WT loops with MLL2 binding sites at both ends were overall smaller than loops that had MLL2 at only one end or at neither end (**Figure 5H**), consistent with smaller loops being lost and the observed gain in loop size in the absence of *Mll2* (**Figure 5G**). We conclude that (i) MLL2 is associated with a large proportion of loops in both naïve and primed PCs; (ii) the number of MLL2-associated loops decreases during the priming transition; and (iii) MLL2 can influence loop size.

MLL2 contributes to loops involving bivalent promoters

MLL2 is the predominant H3K4 methyltransferase at bivalent promoters in ESCs (Denissov et al. 2014). Indeed, Denissov et al (2014) reported that loss of *Mll2* resulted in reduced H3K4me3 at bivalent promoters even though MLL2 was also bound to most active promoters where MLL2 loss did not alter H3K4me3. This evidence for a special relationship between MLL2 and bivalent promoters led us to first focus on loops involving bivalent promoters. As

noted previously, H3K27me3 is virtually absent in naïve ESCs (Marks et al. 2012). Hence there are virtually no bivalent promoters in naïve ESCs with bivalent promoters arising during the transition to primed PCs (**Figure 5I**). Therefore, we focused our analysis on loops and bivalent promoters in primed PCs. Using published ChIP-seq datasets (Yang et al. 2014), we identified the correspondence between loop anchors and active enhancers, active promoters or bivalent promoters to classify loops into 5 main categories (bivalent promoter-bivalent promoter; enhancer-bivalent promoter; enhancer-enhancer; enhancer-promoter; promoter-promoter). In primed PCs, we observed 467 loops between bivalent promoters and most of these were lost in the absence of *Mll2* (3.2-fold decrease; $p = 10^{-26}$) (**Figure 5J and S7A**). Intriguingly, Micro-C of the *Mll2-NCat* cell line (**Figure S6A**) revealed that many of these loops were restored by *Mll2-NCat* (**Figure 5J and S7A**), albeit with some variance, since the specific promoter-promoter loops being restored were not always the same as in wild-type conditions (**Figure S7B**). In contrast, loops in primed PCs between active elements (enhancer-enhancer, enhancer-promoter and promoter-promoter) were more modestly affected by the loss of *Mll2* (1.5-fold, 1.5-fold and 1.6-fold decrease respectively; $p = 10^{-26}$ for all 3 comparisons) and these losses were not rescued by *Mll2-NCat* (**Figure 5J and S7A**). These findings at active elements were also replicated in naïve PCs (**Figure S7A**). Notably, loops between active enhancers and bivalent promoters showed an intermediate impact of the loss of *Mll2* (2.2-fold decrease; $p = 10^{-74}$), which tended to be rescued by *Mll2-NCat* (**Figure 5J and S7A**). Together with the impact of MLL2 on loop numbers and size, the relationship between MLL2 and loops involving bivalent promoters validates our working classification of loops anchored by MLL2 binding sites and strengthens the conclusion that MLL2 directly contributes to looping, with gene ontology showing that these loops are linked to neuroectoderm-specifying genes (**Figure S7C**).

Notably, loss of *Mll2* promoted a slight increase in loops between enhancers and non-specific/undefined genomic regions (1.2-fold/1.8-fold in naïve/primed PCs; $p = 10^{-16}$/$p = 10^{-8}$) and between active promoters and non-specific/undefined genomic regions (1.3-fold/1.8-fold in naïve/primed PCs; $p = 10^{-17}$/$p = 10^{-7}$) but not between bivalent promoters and non-specific/undefined genomic regions ($p = 0.24$/$p = 0.29$ in naïve/primed PCs) (**Figure S7D**). These observations suggest that MLL2 also blocks the formation of non-specific/undefined looping to active enhancers and promoters.

MLL2 enhances chromatin loops associated with gene activation while inhibiting loops associated with DNA methylation and heterochromatin

Because loss of *Mll2* led to a decrease in looping at enhancers/promoters but also an increase in non-specific looping, we further characterised loops mediated/inhibited by MLL2 by taking

advantage of published ChIP-seq datasets (Hammal et al. 2022) to identify chromatin binding proteins enriched at the loop anchor regions gained/lost in *Mll2* cKO cells. To validate this approach, we compared loop anchors that change during the naïve-to-primed transition and found an increase in loops enriched for Polycomb components (CBX2/6/7), the DNA methyltransferase DNMT3B and OTX2, factors known to become active during this transition (**Figure S8A**).

Next, we evaluated proteins enriched at loops lost in *Mll2* cKO cells (**Figure S8B-D**). In naïve PCs, TOP2A, the KAT8-MOF complex, the SPT16-FACT complex and the MED12-Mediator complex predominated (**Figure S8B**). In primed PCs, *Mll2* cKO cells again had fewer loops for TOP2A and the SPT16-FACT complex (**Figure S8C**). Interestingly, *Mll2* cKO also led to a loss of Polycomb-associated loops in primed PCs (**Figure S8C**). This initially appeared surprising given the antagonism between MLL2 and Polycomb (Douillet et al. 2020; Mas et al. 2018). However, because MLL2 and Polycomb co-bind, for example at bivalent promoters, it is likely that they act together to increase looping at these co-bound sites.

To identify loops inhibited by MLL2, we analysed MLL2-associated loops gained in *Mll2* cKO cells (**Figure S8B-D**). In naïve PCs, *Mll2* cKO cells had more loops enriched for the cohesin complex (RAD21), heterochromatin protein HP1β and DNMT3B (**Figure S8B**). In primed PCs, *Mll2* cKO again led to more loops enriched for DNMT3B and cohesin (**data not shown**). These observations indicate that MLL2 blocks the formation of loops associated with DNMT3B and cohesin, consistent with the increase in TAD insulation score observed in *Mll2* cKO naïve PCs (**Figure 4E-G**). Interestingly, the loops without MLL2 in wild-type naïve PCs are enriched for DNMT3B and cohesin (**data not shown**), further suggesting that MLL2 blocks the formation of DNMT3B/SMC3-associated loops in naïve PCs.

MLL2 tethers chromatin during the priming transition

Because MLL2 substantially influences looping in PCs, we asked whether it modulates chromatin tethering on a genome-wide scale by using 3D single-molecule localisation microscopy (SMLM) to track the mobility of individual nucleosomes within the nucleus +/- MLL2 (**Figure 6A**). SMLM has previously revealed that chromatin mobility decreases in the presence of cohesin loops and increases upon histone acetylation or in the presence of specific chromatin remodellers (Nagashima et al. 2019; Nozaki et al. 2017; Basu et al. 2023; Ashwin et al. 2019). It has also shown that chromatin mobility depends on cell type, with differences observed between ESCs and more differentiated cell types (Gomez-Garcia et al. 2021).

To conduct SMLM, we introduced Halo-tagged histone H2B into the *Mll2* cKO cell line, validating expression and nuclear localisation through labelling of cells with a JFX_{549} dye (**Figure 6B and S9A**). We then labelled H2B for SMLM using the photo-activatable PA-JF_{549} dye (Grimm et al. 2016). Cells were imaged using highly inclined and laminated optical (HILO) illumination and double-helix point spread function microscopy (Carr et al. 2017; Basu et al. 2023). This allowed us to track individual nucleosomes in 3D through a 4 μm slice of the nucleus (**Figure 6B and S9A**). Individual PA-JF_{549}-HaloTag-tagged histone H2B molecules were detected as evidenced by single-step photobleaching profiles (**Figure S9B-C**). We collected videos at 500 ms temporal resolution (5 fields of view per replicate, 3 replicates, 2000 frames) because 500 ms exposures detect nucleosomal motion while motion blurring any freely diffusing H2B (Etheridge et al. 2014; Elf, Li, and Xie 2007; Basu et al. 2023). Using fixed cell datasets, we determined the precision at which we could localise a single H2B molecule (54 nm $\pm$ 10 nm in x, 68 nm $\pm$ 8 nm in y and 134 nm $\pm$ 8 nm in z) and found that H2B displacements were significantly higher in all live cell datasets when compared to fixed cell datasets (**Figure S9D**).

To analyse these H2B trajectories, we extracted biophysical parameters from sub-trajectories of 7 frames (3.5 s) to determine how fast chromatin diffuses – the apparent diffusion coefficient D_{app}. Chromatin has viscoelastic properties, so its mobility depends on the time and length scale used for measurements (Hansen, Maeshima, and Hendzel 2021; Zidovska 2020). We chose this timescale because it corresponds to a length scale of ~250 nm, which corresponds to the size of TADs (Nozaki et al. 2017; Miron et al. 2020; Cremer et al. 2020) and to the scale of TAD/loop size changes we had observed (**Figure 4E and 5G**). Also, at this time and length scale, chromatin mobility does not change during the cell cycle (Iida et al. 2022), whereas cell cycle effects have been observed at timescales of a minute or longer (Ma et al. 2019; Seeber, Hauer, and Gasser 2018). As previously described, this revealed slow- and fast-moving sub-diffusive chromatin states (**Figure 6C-E**) (Basu et al. 2023; Gu et al. 2018; Forte et al. 2023; Ashwin et al. 2019). Although the nature of slow- and fast-moving chromatin remains unresolved, meaningful information can still be extracted regarding whether perturbations increase tethering or chromatin compaction (lower D_{app}) (Nagashima et al. 2019; Ashwin et al. 2019; Basu et al. 2023). We therefore took advantage of our published 4-parameter classification algorithm to classify individual trajectories as either slow- or fast-moving prior to extraction of D_{app} (**Figure 6C-E**) (Basu et al. 2023), and then used these parameters to determine how MLL2 influences slow- or fast-moving chromatin during priming.

We first determined how chromatin mobility changes as cells transition from naïve to primed ESCs. There was no change in the proportion of H2B molecules observed in slow- and fast-

moving chromatin states (p = 0.8) (**Figure 6F and S9E**). However, an increase in D_{app} during the priming transition was apparent for both the slow- and fast-moving chromatin states (p = 10^{-3} for both) (**Figure 6G**), indicating that chromatin mobility was increased. We next asked whether similar changes were occurring in the absence of *Mll2*. Although no change in the proportion of H2B molecules in slow- and fast-moving chromatin states was apparent (p = 0.3 and 0.09 for naïve and primed ESCs) (**Figure 6F**), loss of *Mll2* increased D_{app} for slow- and fast-moving chromatin states in both naïve and primed ESCs (p = 10^{-6} and 10^{-3} for naïve ESCs; p = 10^{-3} and 0.02 for primed ESCs) (**Figure 6G**). We concluded from this that MLL2 restricts chromatin mobility in both naïve and primed ESCs, consistent with a role in chromatin tethering, although the effect was larger in primed ESCs possibly because there was an overall increase in chromatin mobility during the naïve to primed transition. This was consistent with the observed increase in TAD and loop size observed upon loss of *Mll2*, particularly in primed PCs (**Figure 4E and 5G**).

To confirm that this effect on chromatin mobility was occurring at MLL2-bound regions, we assessed local chromatin mobility at the MLL2-bound DEG *Sox2* (**Figure 6, S4D and S9F**). We introduced GFP-tagged catalytically-dead dCas9 into the *Mll2* cKO cell line alongside an Atto647-labelled guide RNA sequence that was designed to target a 14-repeat genomic region near the promoter of *Sox2.* We then repeated 3D SMLM and classified trajectories as slow- or fast-moving as described above for H2B (**Figure 6I-K and S9G**). This revealed that the *Sox2* gene was more likely to be observed in the fast-moving chromatin state than previously observed genome-wide using H2B (**Figure 6L**). We also found that loss of *Mll2* reduced the proportion of *Sox2* gene trajectories classified as fast-moving but only in primed ESCs (p = 0.03) (**Figure 6L**). Interestingly, consistent with our H2B findings, loss of *Mll2* again increased D_{app} for both slow- and fast-moving chromatin states but this time only in primed ESCs (p = 0.2 and 0.2 for naïve ESCs; p = 10^{-5} and 10^{-4} for primed ESCs) (**Figure 6M**). We concluded from this that, although MLL2 broadly restricts chromatin mobility across the genome, its impact on specific genes is likely influenced by cell state-specific chromatin loops present at these genes.

Finally, since enhancer-promoter looping is rescued by *Mll2-NCat*, we introduced Halo-tagged histone H2B into the *Mll2-NCat* cell line (**Figure S5D**) We again saw no change in the proportion of H2B molecules observed in slow- and fast-moving chromatin states (p = 0.14 and 1.0 for naïve and primed ESCs) (**Figure S9H**). However, in contrast to the *Mll2* cKO cell line, we no longer observed an increased D_{app} for the fast-moving chromatin state (p = 0.5 and 0.7 for naïve and primed ESCs) (**Figure S9I**). For slow-moving chromatin, we observed a

minor increase in D_{app} in primed ESCs ($p = 0.01$) but no change in naïve ESCs ($p = 0.5$). Consequently, we concluded that *Mll2-NCat* partially, although incompletely, restored the chromatin mobility changes observed after loss of endogenous *Mll2*.

Discussion

H3K4 methylation is important for eukaryotic promoter activity, and epigenetic regulation. This has focused attention on the enzymatic activity of the MLLs to the distraction from their other remarkable characteristics. Indeed, evidence that the H3K4 methyltransferase activities are not required for MLL functions was unexpected (Mishra et al. 2014; Dorighi et al. 2017; Xie et al. 2023) and is still largely ignored. Like the *Drosophila* orthologs Trithorax and Trr/Lpt (Trithorax-related/Lost PHDs of Trr), the MLLs are among the largest nucleoplasmic proteins, but their highly conserved SET H3K4 methyltransferase domains comprise less than 10% of the protein. Here, we describe how other regions of MLL2, which include extensive IDRs and multiple chromatin interaction modules, convey an essential function during neuroectodermal differentiation.

Using conditional mutagenesis, we found that the requirement for MLL2 in neuroectodermal differentiation is exerted many cell cycles before differentiation is impaired and is independent of its H3K4 methyltransferase activity (**Figure 1 and 3)**. Consistent with findings that none of the MLLs are required in ESCs or for mouse development before implantation (Ernst et al., 2004, Glaser et al., 2009; Ashokkumar et al., 2020), scRNA-seq analysis confirmed that MLL2 is not required for the naïve to primed PC transition (**Figure 2**). During the transition, the impact of subtle reductions in gene expression, including reduced expression of the important neuroectodermal transcription factors, *Sox2* and *Otx2*, could explain the downstream defects, because high SOX2 expression is required for differentiation (Blassberg et al. 2022).

To link the non-enzymatic functions of MLL2 to changes in gene expression, we used Micro-C to map 3D genome organisation in the naïve to primed PC transition. Only limited associations between MLL2 and A/B compartments or TADs were observed. In contrast, there was substantial loss of 3D enhancer-promoter looping, particularly at bivalent genes related to neuroectoderm differentiation (**Figure 4-5**). Notably, we found that MLL2 did not require its catalytic activity to stabilise enhancer-promoter loops near these bivalent genes. We therefore tracked chromatin mobility using SMLM to test whether MLL2 stabilises 3D chromatin loops. This revealed that it does indeed restrain chromatin mobility both globally and at the MLL2-bound *Sox2* gene locus (**Figure 6**), consistent with this function. Consequently, we propose that MLL2 stabilises 3D enhancer-promoter loops through its chromatin binding domains

rather than its enzymatic activity, and that this function is required for neuroectoderm differentiation (**Figure 7 and S7C**).

Our proposition relies in part on correlating loops mapped by Micro-C, which like other Hi-C methods is far from quantitative, with genome and chromatin features. The strongest relationship that emerged from these correlations identified loops that have MLL2 binding sites nearby to both anchor regions (which we term 'MLL2 loops') with nearby bivalent promoters (**Figure 5J**). The concordance between MLL2 and bivalent promoters had previously been identified based on H3K4me3 ChIP-sequencing after MLL2 conditional mutagenesis and RNAi knock-down (Denissov et al. 2014; Hu et al. 2013). Here, the independent identification of a relationship between MLL2 and bivalent promoters by Micro-C analysis lends confidence to the qualities of our data and analysis. Further confidence derives from the prominent relationship between MLL2 loops and Polycomb binding sites (**Figure S8C**), which resonates with the well-established interactions between Trithorax and the Polycomb-Group (Kassis, Kennison, and Tamkun 2017; Klymenko and Muller 2004; Schuettengruber et al. 2017). Other associations between MLL2 loops and cohesin, TOP2A or SPT16-FACT warrant further investigation.

Understanding how MLL2-dependent changes in chromatin organization or mobility influence transcription is not yet clear from our studies because MLL2 apparently secures enhancer-promoter loops prior to transcriptional activation. Nevertheless, some insight can be gained from specific cases such as *Magohb* or *Sox2*. Removal of MLL2 results in loss of *Magohb* expression, acquisition of H3K27me3 and then DNA methylation of its CpG island promoter (Douillet et al. 2020; Glaser et al. 2009; Ladopoulos et al. 2013). Here, we showed that MLL2 loss also results in the formation of a new loop to the *Magohb* gene in both naïve and primed ESCs (**Figure S6E-F**). In this case, MLL2 bound at the *Magohb* promoter did not appear to secure a loop anchor but rather prevented the formation of a loop associated with repression. For *Sox2*, MLL2 secures a loop in naïve ESCs, where there is no change in expression or chromatin mobility, but then prevents the formation of a repressive loop in primed ESCs, where there is a loss of expression and an increase in chromatin mobility. The transcriptional amplitude of *Sox2* is associated with dynamic interactions between the gene and super-enhancer condensates (Du et al. 2024) so it is possible that MLL2's ability to stabilise enhancer-promoter loops plays a role, increasing the encounter frequency between the gene and nearby condensates and blocking interactions with repressive condensates. However, these dynamic interactions are likely to be different in naïve and primed ESCs for the *Sox2* gene, leading to the observed cell state-specific effects at this gene.

The finding that MLL2 is required for a developmental transition towards neuroectoderm, but not for ESC self-renewal, resonates with findings for the architectural organisation mediated by CTCF and cohesin, which also plays a limited role in transcription during self-renewal (Hsieh et al. 2022; Nora et al. 2017) but is required for cellular transitions and embryo development (Arzate-Mejia, Recillas-Targa, and Corces 2018). Moreover, like MLL2, CTCF looping increases the transcriptional amplitude of *Sox2* (Du et al. 2024). The MLLs may therefore participate alongside the CTCF/cohesin management of the 3D genome or may present an overlapping web that also contributes to stabilising lineage commitment decisions. Interestingly, the *Sox2* gene exhibits more restrained chromatin mobility during transcription than would be expected from just CTCF/cohesin looping and this additional restriction of chromatin mobility is thought to homogenise and increase the frequency of enhancer-promoter encounters during transcription (Platania et al. 2024). It is therefore possible that the smaller MLL2 loops restrict chromatin mobility and that it is this restriction that drives the higher enhancer-promoter encounter frequency we observe, with loss of *Mll2* leading to formation of only the larger CTCF/cohesin loops (**Figure S8B**). It would therefore be interesting to investigate double CTCF-MLL2 knockouts in future studies.

Our findings underscore the need to look beyond histone marks when studying enhancer-promoter communication, as histone marks can change independently of 3D genome organization. Moreover, they reveal a non-canonical role for MLL2 in stabilizing enhancer-promoter loops that warrants further exploration. Future studies could explore how MLL2's chromatin binding domains and protein-protein interactions influence chromatin mobility and differentiation. MLL2 is also one of four MLLs (KMT2A-D), all of which play roles in cell differentiation. Because all four MLLs and the *Drosophila* orthologs Trithorax and Trr/Lpt have a similar protein architecture, we anticipate that the findings for MLL2 will be relevant for the other MLLs. In particular, MLL1, the closest paralogue of MLL2, is similarly required for lineage commitment in several somatic stem cells. In intestinal stem cells, MLL1 is required for maintenance and for controlling lineage specification of intestinal secretory Paneth and goblet cells (Grinat et al. 2020; Goveas et al. 2021; Grinat et al. 2022). From this perspective, it is likely that the other MLLs also restrain chromatin mobility and collectively may substantially account for enhancer-promoter loop stabilisation during differentiation in many stem cells and lineages. Recent evidence from ESCs enhances the proposition for MLL co-operativity (Kubo et al. 2024).

Limitations

Apoptosis is increased about 2-fold in *Mll2-/-* ESCs (Lubitz et al. 2007). Consequently, we selected single-cell RNA-seq and quantitative immunofluorescence thresholds to remove

smaller apoptotic cells and the SMLM experiments were designed to avoid fields of view containing apoptotic cells. However, it is possible that elevated apoptosis accounts for some of our observed changes. Our SMLM analysis was carried out over a 5-second timescale to monitor changes at the level of TADs and to avoid cell cycle-dependent effects on chromatin mobility. It therefore remains possible that MLL2 also affects chromatin at different timescales. Alternatively, MLL2-dependent effects may be more pronounced at specific cell cycle phases since MLL2 acts primarily in G1 cells (Singh et al. 2015). We also cannot easily exclude transcription-dependent effects, although this seems unlikely because MLL2 affects chromatin mobility similarly in naïve and primed PCs while having an opposing effect on transcription in these cell types. We also cannot exclude the possibility that MLL2 loss influences chromatin mobility indirectly. For example, MLL2-dependent loss of FOXD3 may lead to an increase in H3K27ac (chromatin decompaction), in SS18-BAF-induced chromatin remodelling or in recruitment of DNA methyltransferases (Douillet et al. 2020).

Methods

Primary antibody for Western blot:

Antigen	Dilution	Host	Isotype	Manufacturer
MLL2	1:1000	Rabbit	IgG1	(Glaser et al. 2006)

Secondary antibody for Western blot:

Fluorophore	Dilution	Host species	Reactivity	Manufacturer	Cat #
HRP	1:10000	goat	anti-rabbit	Thermo Scientific	31460

Plasmids used:

Plasmid Name	Description	Source
LZ10 PBREBAC-H2BHalo	Construct expressing the human H2b sequence with a C-terminal HaloTag under the HF1 promoter. Includes an ampicillin and a neomycin resistance cassette.	(Li et al. 2016) Addgene #91564
pCyL43 (PBase)	Construct expressing PBase, the PiggyBac transposase. Includes an ampicillin resistance cassette.	(Wang et al. 2008)
PB-TRE3G-dCas9-eGFP-WPRE-ubcp-rtTA-IRES-puroR	Construct with dual promoter backbone, with a TRE3G (Tet-on) promoter expressing GFP-tagged inactive dCas9 and the ubiquitin C promoter expressing the reverse tetracycline-controlled transactivator, rtTA, and a puromycin cassette via an IRES sequence.	(Gu et al. 2018)

Mouse embryonic stem cell (mESC) lines:

Cell line	Description	Source

Mll2F/F;CreERT2/+	ESCs established *de novo* from blastocysts with LoxP sites either side of exon 2 of the endogenous *MLL2* locus and expressing CreERT2 recombinase from the Rosa26 locus. MLL2 cKO induced by addition of 4-OHT (4-hydroxy-tamoxifen) for 48h and validated by genomic PCR and Western blot (**Figure S1A**)	(Denissov et al. 2014)
Mll2F/F;CreERT2/+ MLL2-GFP	*Mll2F/F;CreERT2/+* ESCs described above with stably integrated BAC expressing GFP-tagged MLL2. Addition of 4-OHT leads to depletion of endogenous but not GFP-tagged MLL2.	(Denissov et al. 2014)
Mll2F/F;CreERT2/+ MLL2(N2650A)-GFP	*Mll2F/F;CreERT2/+* ESCs described above with stably integrated BAC expressing GFP-tagged MLL2 where MLL2 contains a N→A point mutation at position 2650, resulting in loss of catalytic activity. Addition of 4-OHT leads to depletion of endogenous but not GFP-tagged MLL2 (N2650A).	This study
Mll2F/F;CreERT2/+ H2b-HaloTag	same as *Mll2F/F;CreERT2/+* ESCs described above but with an additional stably integrated piggyBac construct carrying *H2b* with a C-terminal HaloTag. MLL2 cKO again induced by addition of 4-OHT	This study
Mll2F/F;CreERT2/+ MLL2(N2650A)-GFP H2b-HaloTag	*Mll2F/F;CreERT2/+ MLL2(N2650A)-GFP* ESCs described above but with an additional stably integrated piggyBac construct carrying *H2b* with a C-terminal HaloTag.	This study

Mll2F/F;CreERT2/+ dCas9-GFP	*same as Mll2F/F;CreERT2/+ ESCs described above but with an additional stably integrated piggyBac construct carrying dCas9 with a C-terminal GFP tag.*	This study

Sequencing datasets:

GSE100597	(Mohammed et al. 2017)
GSE156588	(Kinoshita et al. 2021)
GSE276814	10x scRNA-seq (this study)
GSE279709	Micro-C (this study)

Software:

Data collection

Microscope image acquisition	Micro-manager (Edelstein et al. 2014) (https://www.micro-manager.org)
	ImageJ software (Edelstein et al. 2010; Schneider, Rasband, and Eliceiri 2012)

Data analysis - sequencing

Bowtie2	(Langmead and Salzberg 2012)
Samtools	(Li et al. 2009)
MACS2 (version 2.2.7.1)	(Zhang et al. 2008)
featureCounts	(Liao, Smyth, and Shi 2014)
limma	(Bolstad et al. 2003)
DESeq2	(Love, Huber, and Anders 2014)
Vegan (version 2.4)	https://github.com/vegandevs/vegan
R software (version 4.1.0)	
Cell Ranger (version 3.1)	(Zheng et al. 2017)
Seurat	(Hao et al. 2021) https://github.com/satijalab/seurat
ClustAssess	(Shahsavari, Munteanu, and Mohorianu 2022) https://core-bioinformatics.github.io/ClustAssess/articles/stability-based-parameter-assessment.html

Scanpy	(Wolf, Angerer, and Theis 2018) https://scanpy.readthedocs.io/en/stable/
CellRank	(Lange et al. 2022) https://github.com/theislab/cellrank
Tradeseq	(Van den Berge et al. 2020) https://github.com/statOmics/tradeSeq
MultiQC	(Ewels et al. 2016)
Nuc_processing	(Stevens et al. 2017) https://github.com/tjs23/nuc_processing
BWA	(Li and Durbin 2009) https://github.com/lh3/bwa
Pairtools	(Open2C et al. 2023) https://github.com/open2c/pairtools
Pairix	(Lee et al. 2022) https://github.com/mdozmorov/HiC_tools
HiCRep	(Yang et al. 2017) https://github.com/TaoYang-dev/hicrep
Juicer	(Durand et al. 2016) https://github.com/aidenlab/juicer
HiCCUPS	(Kruse, Hug, and Vaquerizas 2020) https://github.com/vaquerizaslab/fanc
CoolBox (version 0.3.8)	(Xu et al. 2021) https://github.com/GangCaoLab/CoolBox
Jupyter Notebook (version 6.4.12)	https://github.com/jupyter/notebook

Data analysis - imaging

CellProfiler	(Stirling et al. 2021) https://cellprofiler.org/
3D SMLM peak fitting	(Lew 2013) https://sourceforge.net/projects/easy-dhpsf/
Trajectory analysis	(Basu et al. 2018) https://github.com/wb104/trajectory-analysis
SMLM Analysis Toolbox	(this study) https://github.com/stanleystrawbridge/singleMoleculeLocalizationMicroscopyAnalysisToolBox

Trajectory classification	(Basu et al. 2023) https://zenodo.org/record/6497411#.YmlGFy8w3q0

1 – Cell culture

Embryonic stem cell (ESC) line generation and validation

MLL2 Cre-inducible conditional knockout (cKO) ESCs were generated as previously described (Denissov et al. 2014). Tamoxifen was added for 2 days to generate knockout ESCs. Genomic PCR and Western blots ensured gene knockout and protein knock-down respectively. Cell numbers were doubled in knockout conditions to ensure similar numbers of cells when compared to wild-type conditions after 2 days.

MLL2 cKO cells expressing GFP-tagged MLL2 or GFP-tagged MLL2 N2650A (catalytically-dead) ectopically were generated by first using recombineering to generate a bacterial artificial chromosome (BAC) containing the *Mll2* gene (Source Biosciences, bMQ-389B2) tagged with GFP internally inserted on the N-terminal part 5' upstream of the Taspase cleavage sites (Hofemeister et al. 2011). An internal ribosomal entry site (IRES) followed by a neomycin resistance gene was placed 3' downstream of the open reading frame. The N2650A mutation was then introduced to this BAC by site-directed mutagenesis. 1 µg of purified BAC DNA was transfected into ESCs using Lipofectamine LTX and Plus reagent (Invitrogen, 15338-100). Cells were selected with 175 µg/ml G418 starting 1 day after lipofection and picked clones were validated by Western blot and microscopy.

MLL2 cKO ESCs expressing HaloTag-tagged histone H2B (+/- GFP-tagged MLL2 N2650A) were generated by transfecting cells with the LZ10 PBREBAC-H2BHalo (Addgene, Plasmid #91564) (Li et al. 2016) and PiggyBAC transposase PBase plasmids (Wang et al. 2008). 1 µg of the purified plasmids were added to freshly passaged MLL2 cKO ESCs using Lipofectamine-3000 (Thermo Fisher, #L3000001) following the standard protocol. G418-resistant mESCs were selected by growing cells in the presence of 1 µg/ml G418. Single colonies were picked after 7 days, plated individually, and grown to confluence. Clones were chosen for experiments if they had positive nuclear signal after labelling with JFX_{549} or with PA-JF_{549} and if immunofluorescence experiments confirmed that they proceeded through priming as expected.

MLL2 cKO ESCs expressing GFP-tagged inactive dCas9 were generated by transfecting cells with the PiggyBAC transposase PBase plasmid (Wang et al. 2008) and the PB-TRE3G-

dCas9-eGFP-WPRE-ubcp-rtTA-IRES-puroR vector containing a dual promoter backbone, with a TRE3G (Tet-on) promoter expressing GFP-tagged inactive dCas9 and the ubiquitin C promoter expressing the reverse tetracycline-controlled transactivator, rtTA, and a puromycin cassette via an IRES sequence (Gu et al. 2018). Puromycin-resistant ESCs were selected for 7 days and doxycycline added for 24 hours to induce expression of dCas9-GFP (through activation of the rtTA). Stable transfectants were FACS sorted for GFP expression.

ESC culture conditions

Naïve PCs were grown in 2iLIF medium at 37 °C, 7% CO_2 and 80% humidity as previously described (Ying et al. 2008). Phenol red-free 2iLIF medium was prepared fresh every week by adding 1 μM PD0325901 (ABCR, AB 253775), 3 μM CHIR99021 (ABCR, AB 253776) and 10 ng/ml murine leukaemia inhibitory factor (mLIF, provided by the Department of Biochemistry, University of Cambridge) to N2B27 basal medium. N2B27 medium is composed of 50% phenol red-free DMEM/F-12 (Gibco #21041025) and 50% phenol red-free and L-Glutamine-free Neurobasal (Gibco #12348017) supplemented with 0.5x N2 (made and batch tested in-house by Cambridge Stem Cell Institute), 1x B-27™ Supplement (Gibco #17504044), 2mM L-Glutamine (Life tech, #25030024), and 0.1 mM 2-mercaptoethanol (in-house/Life Tech, #1985023). Cells were passaged every two days by washing in PBS (Sigma-Aldrich #D8537), followed by incubation with accutase (Sigma-Aldrich #A6964) for 2 min at 37 C to detach. Cell clumps in suspension were then washed and pelleted in PBS before resuspending and re-plating as single-cell suspension in fresh medium. To help cells attach to the surface, plates were incubated for 15 minutes at 37 C in PBS containing 0.2% gelatin (Sigma-Aldrich #G1890). All cell lines were routinely screened for mycoplasma contamination at least twice yearly and tested negative.

Priming of naïve ESCs to primed epiblast-like cells (EpiLCs)

Naïve ESCs were converted to primed EpiLCs as previously described. Naïve mESCs were converted to primed EpiLCs as previously described (Hayashi et al. 2011). Briefly, plates were coated with 20 μg/ml human plasma fibronectin (Sigma-Aldrich #FC010-10MG) for 5 min, then incubated to dry for 30 min at 37°C and 7% CO_2 prior to plating cells at a density of $3x10^5$ cells per 6-well (~$3x10^4$ cells/cm^2). 2i inhibitors and mLIF were removed and replaced with N2B27 basal medium supplemented with 12.5 ng/ml FGF2 and 20 ng/ml Activin A (both provided by the Department of Biochemistry, University of Cambridge). Cells were cultured in FGF2 plus Activin A medium for three days with daily media changes.

Differentiation of EpiLCs to neuroectoderm

To continue differentiating EpiLCs to neuroectoderm, wells were washed with DMEM/F12 and cells were cultured on coverslips for 7 days in N2B27 without growth factors. The medium was changed every other day. After 7 days in N2B27, coverslips were fixed and permeabilized for immunofluorescence. Neural rosettes from 6-well plates were dissociated with 0.5 ml accutase for 5 min at 37ºC, diluted and resuspended with 4.5 ml DMEM/F12, collected by centrifugation for 5 min at 215 g and cultured on non-adhesive 6 cm suspension culture dishes in 4 ml NSA with 0.5x B27 for 2-4 days to allow formation of neurospheres. These were then transferred to centrifugation tubes, allowed to settle down for 5 min and the old medium was carefully aspirated. Spheres were resuspended in 3-4 ml fresh NSA with B27 and seeded to 6-well plates (Nunc) or 6 cm dishes (VWR) that were beforehand coated with 0.1% gelatin for at least 1 hr at RT, aspirated and dried under the laminar flow hood with open lids for 10-20 min. Spheres were allowed to attach and grow out for 2 days without moving the dishes around before still unattached spheres were aspirated and medium was changed. When reaching near confluency (95%) at day 3-5 on gelatine, neural stem (NS) cells were passaged for the first time by aspirating the medium and detaching the cells with 0.5 ml Accutase for 5 min at 37ºC. Accutase was diluted with 4.5 ml DMEM/F12, cells were collected by centrifugation for 5 min at 215 g,resuspended in fresh NSA with B27 and seeded to gelatin-coated dishes (Nunc or WVR) in a splitting ratio of 1:3 to 1:8. B27 was removed from the medium one day after the first passage and NS cells were regularly passaged every 3-5 days shortly before confluency was reached (95%) as described above. EpiSCs were differentiated similarly but after 21-35 days of culture in FGF2 plus Activin A medium, passaging cells every 3-5 days around 24 h after reaching full confluency.

2 - Assays

Immunofluorescence

Immunofluorescence staining of neural rosettes: Cells on glass cover slips were washed with PBS, fixed 10 min in 4% formaldehyde/PBS (Merck Millipore, 104003), washed with PBS, permeabilized 15 min in 0.5% Triton X-100/PBS (Sigma-Aldrich, T8787), washed again with PBS and blocked 30 min with 3% BSA/PBS (Sigma-Aldrich, A7906). Cover slips were either stored in 3% BSA/PBS (Sigma-Aldrich, A7906) for up to 3 months at 4 ºC or processed directly for immunofluorescence analysis as follows. Each cover slip was incubated in 25 µl drops of primary antibodies diluted in 10% goat serum/PBS (Sigma-Aldrich, G9023) over night at 4 ºC in a humidified atmosphere. The following day cover slips were washed in PBS and incubated 1-4 hours in 25 µl drops of appropriate secondary antibodies diluted in 10% goat serum/PBS (Sigma-Aldrich, G9023) at 4ºC in the dark. Cover slips were washed again in PBS and stained with 100 ng/ml 4,6-diamidino-2-phenylindole (DAPI)/PBS for 2 min (Sigma-Aldrich, D8417). After washing in PBS and then in ddH2O each cover slip was mounted to glass slides with

7 µl Mowiol (Sigma-Aldrich, 81381) and dried overnight at RT before imaging with the TCS SP5 confocal system.

Immunofluorescence staining of priming transition: Cells were cultured in Fibronectin-coated Ibidi 8-well micro-chambers (Ibidi, IB-80826). Cells were washed in PBS and fixed with 4 % PFA for 10 mins at room temperature. Fixed cells were rinsed three times in PBS and then blocked for two hours at room temperature in blocking buffer: PBS containing 0.1 % TritonX-100 (Sigma-Aldrich, T8787-100M) and 3% normal Donkey Serum (Sigma-Aldrich, D9663-10ML). Cells were incubated with primary antibodies diluted in blocking buffer (see concentrations below) at 4 ºC overnight, followed by three 10 min washes in PBST (PBS containing 0.1% TritonX-100). Incubation with secondary antibodies and 1000 ng/ml DAPI (Sigma-Aldrich, MBD0015-1ML) diluted in blocking buffer (see concentrations below) was performed for two hours at room temperature in the dark. Stained samples were then rinsed in PBST three times for 10 min before being stored in PBS at 4 ºC overnight prior to imaging on a Nikon Spinning Disk Ti inverted microscope using a 40x oil immersion objective and an Andor iXon + camera.

Primary Antibodies for immunofluorescence:

Antigen	Dilution	Host	Isotype	Manufacturer	Cat #	LOT #
OTX2	1:50	goat	IgG polyclonal	R&D Systems	AF1979	KNO0918 121
NANOG	1:200	rat	IgG2a, kappa	eBioscience	14-5761-80	2003079, 2403803, 2269925
SOX2	1:200	rat	IgG2a	eBioscience	14-9811-82	2213316
PAX6	1:20	mouse	IgG1	DSHB	supernatant	
NESTIN	1:20	rat	IgG1	DSHB	Rat-401	
ZO-1	1:100	rabbit	IgG1	BD Biosciences	610966	

Secondary Antibodies for immunofluorescence:

Fluorophore	Dilution	Host species	Reactivity	Manufacturer	Cat #	LOT
DyLight488	1:500	donkey	anti-goat	DyLight	SA5-10086	WE3275103
AF555	1:500	donkey	anti-rat	Abcam	ab150154	GR3262493-3
CF647	1:1000	donkey	anti-rabbit	Biotium	20047-1mg	17C0123

Microscope Settings:

Fluorophore	Excitation Wavelength	Laser Power	Exposure Time
DAPI	405 nm	50.0%	80 ms
DyLight488	488 nm	20.0%	200 ms
AF555	561 nm	20.0%	400 ms

CF647	638 nm	68.1%	300 ms

Single-cell Transcriptomics (10X Genomics)

Cell Harvesting: MLL2 cKO and control cells were transitioned from naive to primed pluripotency as described above. Cells harvested at different time points during the protocol (0 (2iLIF), 24, 48 or 72 hours) were prepared for 10X genomic sequencing, according to the manufacturer's protocol. The experiment was repeated twice on separate days. Briefly, cells were washed twice with PBS, detached and collected with PBS supplemented with 0.4 mg/mL BSA (Sigma-Aldrich, A2153) to neutralise excess accutase activity. After pelleting and resuspension in PBS/BSA, live cells were counted using a TC20™ automated cell counter (BioRad Cat.# 1450102), BioRad Cell Counting Slides (BioRad #1450011) and 0.4% Trypan Blue Solution (Gibco™ #15250061). Cells were washed twice with PBS/BSA and passed through a 40 µm cell strainer (Greiner, 542040), then counted, pelleted and resuspended in PBS/BSA at a final concentration of 40 cells/µl. After harvesting, cells were kept on ice between handling and all pipetting steps except the final resuspension step were carried out using wide-bore pipette tips to minimise shearing forces/cell damage.

Single-cell RNA-seq library preparation and sequencing: Libraries were prepared and sequenced in the Cancer Research UK Cambridge Institute Genomics Core Facility using Chromium Next GEM Single Cell 3′ GEM, Library & Gel Bead Kit v3.1 (10X Genomics, #PN-1000121), Chromium Next GEM Chip G Single Cell Kit (10X Genomics, #PN-1000120) and Chromium Next GEM Single Cell 3' Reagent Kits User Guide v3.1 (Manual Section CG000204 Rev D; 10X Genomics). Cell suspensions were loaded on the Chromium instrument with the expectation of collecting gel-beads emulsions containing single cells. RNA from the barcoded cells for each sample was subsequently reverse-transcribed in a C1000 Touch Thermal cycler (Bio-Rad) and all subsequent steps to generate single-cell libraries were performed according to the manufacturer's protocol with no modifications. cDNA quality and quantity were measured with Agilent TapeStation 4200 (High Sensitivity D5000 ScreenTape, Reagents and Ladder, Agilent #5067-5592, -5593 and -5594) after which 25% of material was used for gene expression library preparation. Library size distribution was confirmed with Agilent TapeStation 4200 (High Sensitivity D1000 ScreenTape, Reagents and Ladder, Agilent #5067-5584, -5585, -5587 and -5603) and Qubit 4.0 Fluorometer (ThermoFisher Qubit™, dsDNA HS Assay Kit, #Q32853) to quantify dsDNA. Each sample was normalised and pooled in equal molar concentration. To confirm concentration, pools were qPCRed using the KAPA Library Quantification Kit (Roche # KK4824) on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems, #4485691) before sequencing. Pools were sequenced on an Illumina

NovaSeq6000 sequencer with the following parameters: 28 bp, read 1; 8 bp, i7 index; and 91 bp, read 2.

Micro-C

Micro-C was performed using the Dovetail Micro-C Kit (Dovetail Genomics, #21006) following the manufacturer's instructions with slight modifications, as described below. The analysis consisted of two independent experiments, SLX-22775 and SLX-23477/SLX-23479, each with technical duplicates, and with the latter including the catalytic activity mutant (referred to as N2650A).

Cell Harvesting: Briefly, 1 million naïve or primed pluripotent cells were harvested, washed in 1x PBS, pelleted and the pellets placed at -80°C for at least 30 min before proceeding with the protocol.

Crosslinking and MNase digestion: Long-range chromatin crosslinking was achieved by addition of DSG (disuccinimidyl glutarate; ThermoFisher Scientific #A35392), followed by 37% formaldehyde (Sigma-Aldrich, #F8775) to stabilise shorter-range DNA-protein and protein-protein interactions. After two wash steps, cells were subjected to *in situ* micrococcal nuclease (MNase) digestion. We used 2 uL of MNase enzyme in all experiments. EGTA (ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid) was added to terminate the reaction after exactly 15 min. Cells were subsequently lysed using SDS (sodium dodecyl sulphate) and the lysate stored at -80°C until the next step.

Quality control of MNase digestion profile (QC 1): For the first quality control step, we assessed the size distribution of the MNase-digested DNA to ensure it was the recommended 40-70%. A small sample taken from the lysate was subjected to crosslink reversal and Proteinase K digestion to remove protein contaminants. DNA was purified using the Zymo Research DNA Clean and Concentrator-5 Kit (Zymo Research, #D4013) and analysed using an Agilent D5000 ScreenTape (Agilent, #5067-5588) on the Agilent TapeStation 4200 (JCBC NGS facility; Analysis Software Version 5.1). DNA concentration was determined using a Qubit 2.0 Fluorometer (JCBC NGS facility; Qubit™, ThermoFisher Scientific).

Proximity ligation, end polishing and crosslink reversal: To carry out proximity ligation, samples were thawed, and 1000 ng of each lysate immobilised on chromatin capture beads. Following two washes, an end polishing step was performed to fill 'sticky ends' generated by MNase digestion. After another wash, crosslinked fragments were ligated in a two-step reaction using the provided bridge, bridge ligase and the intra-aggregate ligation reagents. Consequently, a combined crosslink reversal and Proteinase K digestion step was carried out and resulting ligated fragments were eluted from the beads. This step also introduced a biotin-label into the ligated fragments, crucial for enrichment of ligated fragments using streptavidin beads, as

described below. DNA was subsequently purified using SPRIselect beads (Beckman Coulter # B23319) at a 1.8-fold bead-to-sample ratio to ensure maximum recovery of all DNA fragments ≥100 bp and stored overnight at -20 °C.

End repair, USER digest and library indexing: Purified and proximity-ligated fragments were subjected to End Repair, followed by Ligation with Illumina Adapters, USER (uracil-specific excision reagent) digest and another SPRIselect bead-based purification (2.2 bead-to-sample ratio). Ligated sample-adapter fragments were then immobilised on streptavidin beads (via the biotinylated bridge introduced before) and washed for the indicated times with several provided buffers (LWB, NWB, 1x Wash Buffer). Samples were indexed using a set of eight unique UDI Primer Pairs in a standard PCR reaction (12 cycles) as recommended. Primer Pairs were allocated as follows:

Index Name	Sequence (i7/i5)	Repeat 1-2	Repeat 3-4 Pool A	Repeat 3-4 Pool B
xGenUDI1	CTGATCGT-ATATGCGC	0h WT1	0h WT1	72h WT1
xGenUDI2	ACTCTCGA-TGGTACAG	0h WT2	0h WT2	72h WT2
xGenUDI3	TGAGCTAG-AACCGTTC	0h KO1	0h KO1	72h KO1
xGenUDI4	GAGACGAT-TAACCGGT	0h KO2	0h KO2	72h KO2
xGenUDI5	CTTGTCGA-GAACATCG	72h WT1	0h *Mll2-NCat* +WT1	72h *Mll2-NCat* +WT1
xGenUDI6	TTCCAAGG-CCTTGTAG	72h WT2	0h *Mll2-NCat* +WT2	72h *Mll2-NCat* +WT2
xGenUDI7	CGCATGAT-TCAGGCTT	72h KO1	0h *Mll2-NCat* +KO1	72h *Mll2-NCat* +KO1
xGenUDI8	ACGGAACA-GTTCTCGT	72h KO2	0h *Mll2-NCat* +KO2	72h *Mll2-NCat* +KO2

The resulting libraries were subjected to a double-sided selection using SPRIselect beads (right-sided selection at 0.5x, followed by left-sided selection of supernatant at 0.7x bead-to-sample ratio), which excludes any fragments sized <350 bp or >1000 bp.

Quality control for library size distribution (QC2): A small sample of each library was run on an Agilent TapeStation 4200 to validate that we had the required fragment size distribution and a Qubit 2.0 Fluorometer to determine DNA concentration.

Library pooling: Eight libraries (corresponding to UDI primer pairs 1-8) were pooled at equal concentration to yield a final concentration of 5 nM and the average fragment size of the pooled library was determined using the Agilent TapeStation 4200. Average pooled library lengths are listed below:

Library Name	Library Barcode	Samples	Average Length [bp]
A	SLX-23477	0h WT, 0h KO, 0h *Mll2-NCat*+WT, 0h *Mll2-NCat* +KO	460
B	SLX-23479	72h WT, 72h KO, 72h *Mll2-NCat* +WT, 72h *Mll2-NCat* +KO	468

Submission for sequencing: Samples were submitted for 200 bp paired-end sequencing at the CRUK-CI Genomics Core Sequencing Facility and sequenced on a NovaSeq S4 Flow Cell using a standard Illumina sequencing workflow.

Live-cell Single Molecule Localisation Microscopy (SMLM)

Cell preparation: Prior to imaging, cells were passaged onto fibronectin-coated (16.7 µg/ml; Sigma-Aldrich, #FC010; later #F0895) 35 mm glass bottom dishes No 1.0 (MatTek Corporation, #P35G-1.0-14-C) in phenol red-free 2i/LIF or FA medium, respectively, and grown for the indicated duration, according to the protocols described above.

Labelling of H2B-Halo expressing cells: On the day of imaging, H2B-Halo expressing PCs were incubated with 50 nM PA-JF$_{549}$ Halo ligand dyes (Grimm et al. 2016) (dyes were a kind gift from Luke Lavis's Lab, Janelia Research Campus; reconstituted in DMSO) for 15 min at 37 °C, followed by three PBS washes and incubation for 30 min at 37 °C to release excess dye, and then imaged in fresh medium. Fixed cells were prepared by washing cells with PBS, then incubating for 10 min in PBS containing 4% formaldehyde (Thermo Fisher Scientific, 28908) and then washing in PBS three times.

Labelling of *Sox2* gene in dCas9-GFP expressing cells: To label the *Sox2* gene, we first induced expression of dCas9-GFP by adding 1 µg/ml doxycycline a day before cells were plated onto imaging dishes containing 2i/LIF or FA medium. Cells were then transfected using lipofectamine CRISPRMAX™ Cas9 Transfection Reagent (Thermo Fisher Scientific, CMAX00008) with Atto647-labelled gRNAs as cells were plated onto imaging dishes. We targeted the dCas9 to a genomic repeat (15 repeats) that is 1 kb upstream of the *Sox2* gene promoter using 20 µM of a custom crRNA (GGGAGGGAGGAAGUGGGAGG; Sigma, TRACRRNAMOD) assembled with Alt-R™ CRISPR-Cas9 tracrRNA, ATTO™ 647 (IDT,

10007853). crRNA and trcRNA were annealed for 5 min at 95 ºC, then on ice for 20 min, prior to transfection.

Microscope setup: A custom-built double helix point spread function (DHPSF) microscope was used for 3D single-molecule tracking as previously described (Carr et al. 2017; Steindel et al. 2022). Briefly, lasers were used at appropriate wavelengths for fluorophore activation (405 nm: 100 mW, 06-MLD, Cobolt) and excitation (561 nm: 200 mW, 06-DPL, Cobolt) of the PA-JF_{549} fluorophore. Atto647-gRNAs were excited using a 638 nm laser (180 mW, 06-MLD, Cobolt). Laser beams were circularly polarised, expanded, collimated and focused to the back focal plane of an index-matched water immersion objective lens (CFI Plan Apo IR 60XC WI 1.27 NA, MRD07650, Nikon) mounted on an inverted fluorescence microscope (Eclipse Ti2, Nikon). The microscope was equipped with a built-in Perfect Focus System, providing automated focusing over extended imaging periods and precise axial movement. This allows for accurate data acquisition and calibration of the DHPSF microscope by adjusting the axial position of the objective lens. A high-precision motorised linear XY stage (HLD117NN, Prior) was used to position the imaging sample. The system was also equipped with an incubator chamber for live-cell imaging (Digital Pixel), enclosing the stage and objective in a sealed plastic chamber with two heaters to consistently maintain the temperature at 37°C.

Sample illumination: At the beginning of each day of experiments, the laser beam was aligned to the Highly Inclined and Laminated Optical sheet (HILO) illumination mode for better optical sectioning, resulting in reduced fluorescence background, improved signal-to-noise ratio (SNR), and reduced photobleaching of the sample. The thickness of the HILO illumination slice was similar to that of the DH axial range (~4 µm). Once optimal HILO alignment was achieved, constant illumination was maintained throughout all experimental conditions.

Sample emission: The emission signal was redirected from the excitation path into the emission path by a quad-band dichroic mirror (Di01-R405/488/561/635-25x36, Semrock). The detection arm included a 4f system with two 200-mm tube lenses (TTL200-A, Thorlabs) and a 580nm or 687 nm double-helix phase mask (Double Helix Optics) positioned at the Fourier plane to generate the axially-modified PSF containing 3D information. The resulting signals were then focused by a third tube lens (TTL200-A, Thorlabs) onto an EMCCD detector (Evolve Delta 512, Photometrics). Long-pass and band-pass filters (BLP02-561R-25 and FF01-580/14-25 for PA-JF_{549} and FF01-676/29-25 for Atto647, Semrock) were placed immediately before the camera to selectively transmit the emitted fluorescence while blocking unwanted wavelengths of light.

Data acquisition: Micro-Manager (version 1.4) was used to control the shutters, sample stage, camera, and imaging parameters during data acquisition (Edelstein et al. 2014). Calibration series for each DHPSF channel were acquired using TetraSpeck 0.1 µm Microspheres (Thermo Fisher Scientific, #T7279) via a custom Micro-Manager script (10 frames per 40 nm

axial increment spanning a total axial range of 4 µm) before any imaging experiments. For each experiment, 2000 frames of fluorescence images were collected with an exposure time of 500 ms. Although the 405 nm laser was used to validate cells, it was not required for the experiment itself because spontaneous activation of the PA-JF_{549} fluorophore provided sufficient density of single H2B molecules. 561 nm and 638 nm lasers were set to power densities of 40 and 0.7 W/cm^2 respectively. Each experiment included at least 6 biological replicates, with each replicate covering 5 fields of view containing approximately 3 cells (total of 90 cells per condition).

4 – Quantification and statistical analysis

Identification of MLL2-dependent cell fractions and genes using 10x scRNAseq

Raw sequencing reads were aligned to the mouse reference genome (mm10, 10x reference: refdata-gex-mm10-2020-A). Cell Ranger version 3.1 was used to generate unique molecular identified (UMI) count matrices for single cells using default settings (Zheng et al. 2017). Seurat R package version 4.3.0 on R version 4.2.1 was used for data analysis for all conditions and both biological replicates (Hao et al. 2021). The UMI counts were normalised and scaled for each cell using a negative binomial regression-based heterogeneity removal method, *sctransform*. A principal component analysis (PCA) was performed on the normalised and scaled data. The Harmony algorithm was then used on the top 50 principal components, to integrate data from two biological replicates and reduce batch effects. After the integration step, cells were separated into clusters using Seurat's shared nearest neighbour method, *FindNeighbors* and *FindClusters*, with resolution = 0.3. For visualisation and data exploration of cell-to-cell clustering and variation, the top 50 principal components with overlayed cluster information were plotted using a Uniform Manifold Approximation and Projection (UMAP) plot. The fraction of cells/cluster in wild-type and MLL2 cKO cells were compared using Fisher's exact test.

<u>Quality control:</u> Parameter selection for the Seurat pipeline was optimised with the *ClustAssess* package version 0.3.1 (Shahsavari, Munteanu, and Mohorianu 2022). The *plot_feature_stability_boxplot* and *get_nn_importance* functions were used to assess UMAP clustering stability and consistency. 10 computational iterations (*n_repetitions*) were carried out with 50 principal components (*npcs*) and 50 nearest neighbours (*n_neighbours*).

<u>Differential gene expression</u> per cluster was performed using a negative binomial regression model. To identify MLL2-dependent genes per cluster, we used standard Seurat functions to

conduct the analysis, first applying *PrepSCTFindMarkers* function to correct counts, then *FoldChange* to obtain log2-fold change for all genes in each cluster and *FindMarkers* to identify those that were differentially expressed. Differentially expressed genes were selected if they showed a p-value < 0.001.

Comparison of scRNAseq with published scRNAseq datasets for validation

To compare our data to published single-cell datasets taken from mouse embryos *in vivo* and formative-like PCs *in vitro*, we extracted the QC-processed, single cell count matrix data from GSE100597 (Mohammed et al. 2017) and GSE156588 (Kinoshita et al. 2021) respectively. For each comparison, a list of common genes expressed in both datasets was identified and genes not expressed in both datasets were removed. For the *in vivo* comparison, we identified a list of the 3000 most-variable genes in the *in vivo* dataset using the *SCTransform* function from Seurat. For the *in vitro* comparison, we used the 3000 most-variable genes identified in our dataset (see above). We then re-applied our Seurat pipeline using the same functions and parameters as described above.

Trajectory analysis using Cell Rank

To visualize the time evolution of our system, we used the Python module *scanpy* (Wolf, Angerer, and Theis 2018), where we recomputed neighbours and UMAP coordinates with default parameters, excluding cells from the cluster of primarily apoptotic cells found at all timepoints. As a result, we obtained a continuous landscape. We then ran *cellrank* to produce a directional stream plot (Lange et al. 2022). Cellrank was initialised with the diffusion pseudotime kernel, also computed in *scanpy* (Haghverdi et al. 2018). To test for differences in gene dynamics among MLL2 cKO and control cell conditions, we used the R package *tradeSeq* (Van den Berge et al. 2020). Upon running the function *conditionTest*, we identified genes that have overall differential trajectories. Finally, we fed our 200 top differentially expressed genes to the function *clusterExpressionPatterns* and identified 15 clusters of genes that have similar patterns for both MLL2 cKO and control cell conditions.

Quantitative immunofluorescence data analysis

Data inspection and processing were performed in Fiji (is just ImageJ, version 2.9.0/1.53t; Java 1.8.0_322 [64-bit]) and involved adjustment of pixel display range (identical settings applied across conditions for each marker, except DAPI, which was auto-adjusted for best signal), cropping, conversion to RGB format and addition of scalebars.

For single-nucleus segmentation and quantification of signal, images were loaded into CellProfiler (version 4.2.5) and single nuclei were identified by DAPI signal (size range: 70-

150 pixels; thresholding method: Otsu, thresholding strategy: global, lower and upper thresholding bounds: 0-1.0, thresholding smoothing scale: 1.3488). Binarised segmentation files were inspected manually and any erroneous conditions were excluded from further analysis. Average signal intensity was then measured by object, normalised by nuclear area and plotted using ggplot2 in RStudio (version 2022.02.3, R version 4.1.1). For calculation of fold-change values, the median intensity of each condition was chosen since data was not normally distributed. Technical background intensity for each channel was determined in a DAPI-only sample and processed in the same manner.

Fluorophore	**Mean background intensity**	**SD background intensity**	**n cells**
DyLight488	0.001437	8.212416e-6	1982
AF555	0.0007859	1.054071e-5	1982
CF647	0.003491	1.710086e-4	1982

Mean background per area (normally distributed) was then subtracted from median fluorescence intensity for each condition before calculation of fold-change.

Micro-C data analysis

Quality control of raw files was performed using *multiqc* (Ewels et al. 2016). Barcode demultiplexing was performed using *nuc_processing* (Stevens et al. 2017). For alignment, the Dovetail genomics pipeline was used (https://micro-c.readthedocs.io/en/latest/). This included sequence alignment to the mm10 mouse reference genome using *bwa* in MEM mode (Li and Durbin 2009). *pairtools* (Open2C et al. 2023) and *pairix* (Lee et al. 2022) were used to record valid ligation events and for indexing the pairs of genomic coordinates respectively. Then, *juicer* was used to convert the pairs files into contact matrices (hic files) (Durand et al. 2016). A-B compartments and TAD boundaries were called using *FAN-C* (0.9.17) (Kruse, Hug, and Vaquerizas 2020), applying default settings. HiCRep highlighted an acceptable level of consistency between replicates, and so replicates were merged (Yang et al. 2017). FAN-C was then used on the merged replicate contact matrices for more detailed identification of significant loops at 50 kb resolution, again applying default settings. Differential loops were detected by considering significant loops that were present in only one of the conditions compared.

To identify proteins enriched at loop anchor regions, we used the Remap2022 public database (non-redundant dataset) (Hammal et al. 2022) to obtain ChIP-seq binding data for ESCs. To identify chromatin binding proteins enriched in one condition over another, we used a rank-based algorithm as follows: for all loops in each condition and for each chromatin binding protein, we counted the number of binding peaks present at either end of a significant loop. Then, instead of using the raw counts, which is confounded by the number of experiments available for each chromatin binding protein, we only considered the rank of each protein. Then, for comparing data between conditions, we compared the difference in rank rather than the difference in the number of binding peaks.

To relate these loops to enhancers and promoters, we took advantage of published H3K4me3, H3K4me1, H3K27me3 and H3K27ac ChIP-seq datasets at 0 hr and 72 hr of the priming transition (Yang et al. 2014). FASTQ datasets were uploaded to the Galaxy web platform, for which the public server at usegalaxy.org was used to analyse the data (Afgan et al. 2016). Briefly, the analytical pipeline started with mapping FASTQ files to the mm10 genome using Bowtie2 and single 51 bp reads (Langmead and Salzberg 2012). Peaks were subsequently called on the BAM files with the MACS2 Narrowpeaks function (Zhang et al. 2008). Lastly, overlapping peaks were identified using the bedtools Multiple Intersect function (Quinlan and Hall 2010) to define active enhancers (H3K4me1 and H3K27ac), active promoters (H3K4me3 and H3K27ac) and bivalent promoters (H3K4me3 and H3K27me3).

Live-cell Single Molecule Localisation Microscopy data analysis

Data analysis of single-molecule localisations was carried out as previously described (Basu et al. 2023) with minor changes. Briefly, single-molecule localisations were extracted from 3D DHPSF movies using easyDHPSF software (Lew 2013) in MATLAB (version 2016b) and the above-mentioned calibration file. After manual inspection of different thresholds, a relative localisation threshold of 80 and 120 was set for all 6 angles of H2B and *Sox2* datasets respectively. Individual localisations were assembled into trajectories using 'trajectory-analysis' (https://github.com/TheLaueLab/trajectory-analysis). Molecules in subsequent frames within a 500 nm radius were considered part of the same trajectory and hence connected, with a frame gap tolerance of 2 frames. To calculate the precision of the SMLM data, we conducted displacement analysis of fixed cells using the SMLM Analysis Toolbox (https://github.com/stanleystrawbridge/singleMoleculeLocalizationMicroscopyAnalysisToolBox).

A custom algorithm developed in our lab called 'analyze-tracking-data' was then deployed, which uses a sliding window along each trajectory to output four biophysical parameters for each sub-trajectory – anomalous exponent (α), diffusion coefficient (D_{eff}), length of confinement (L_c) and drift magnitude (Basu et al. 2023). Analysis was carried out using the following variable settings in the code: a sliding window of 7 frames for H2B and 11 frames for *Sox2*-loci, 0.5 s exposure time, a minimum trajectory length of 7 frames. Since chromatin mobility can be classified into two subpopulations, slow- or fast-moving chromatin, we also set this algorithm to use a 4-parameter Gaussian mixture model, considering all four biophysical parameters (default), to classify sub-trajectories as either slow- or fast-moving and to then output biophysical parameters separately for these two types of chromatin mobility. The average parameter value for each replicate were plotted using custom scripts in RStudio (version 4.3.0). To assess normality, we conducted the Shapiro-Wilk test and analyzed Q-Q plots before performing any comparisons. Based on the normality results, we used the Mann-Whitney-Wilcoxon test for non-parametric comparisons, applying the Bonferroni-Holm adjustment method for multiple testing corrections.

Materials availability

All plasmids and cell lines generated in this study are available from Addgene or from the lead contact with a completed materials transfer agreement.

Data and code availability

All imaging datasets are available from the lead contact. The XYZT single molecule trajectory data files are uploaded to Zenodo and will be made available upon publication. The scRNAseq and Micro-C datasets reported in this study are available from the Gene Expression Omnibus (GEO) repository under accession codes GSE276814 and GSE279709. All code used is already open-source and links have been provided to the relevant websites. Original code developed for this study, the SMLM Analysis Toolbox, has been made available on Github: (https://github.com/stanleystrawbridge/singleMoleculeLocalizationMicroscopyAnalysisToolBox).

References

Addicks, G. C., C. E. Brun, M. C. Sincennes, J. Saber, C. J. Porter, A. Francis Stewart, P. Ernst, and M. A. Rudnicki. 2019. 'MLL1 is required for PAX7 expression and satellite cell self-renewal in mice', *Nat Commun*, 10: 4256.

Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Gruning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, and J. Goecks. 2016. 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update', *Nucleic Acids Res*, 44: W3-W10.

Argelaguet, R., S. J. Clark, H. Mohammed, L. C. Stapel, C. Krueger, C. A. Kapourani, I. Imaz-Rosshandler, T. Lohoff, Y. Xiang, C. W. Hanna, S. Smallwood, X. Ibarra-Soria, F. Buettner, G. Sanguinetti, W. Xie, F. Krueger, B. Gottgens, P. J. Rugg-Gunn, G. Kelsey, W. Dean, J. Nichols, O. Stegle, J. C. Marioni, and W. Reik. 2019. 'Multi-omics profiling of mouse gastrulation at single-cell resolution', *Nature*, 576: 487-91.

Arzate-Mejia, R. G., F. Recillas-Targa, and V. G. Corces. 2018. 'Developing in 3D: the role of CTCF in cell differentiation', *Development*, 145.

Ashokkumar, D., Q. Zhang, C. Much, A. S. Bledau, R. Naumann, D. Alexopoulou, A. Dahl, N. Goveas, J. Fu, K. Anastassiadis, A. F. Stewart, and A. Kranz. 2020. 'MLL4 is required after implantation, whereas MLL3 becomes essential during late gestation', *Development*, 147.

Ashwin, S. S., T. Nozaki, K. Maeshima, and M. Sasai. 2019. 'Organization of fast and slow chromatin revealed by single-nucleosome dynamics', *Proc Natl Acad Sci U S A*, 116: 19939-44.

Basu, S., L. M. Needham, D. Lando, E. J. R. Taylor, K. J. Wohlfahrt, D. Shah, W. Boucher, Y. L. Tan, L. E. Bates, O. Tkachenko, J. Cramard, B. C. Lagerholm, C. Eggeling, B. Hendrich, D. Klenerman, S. F. Lee, and E. D. Laue. 2018. 'FRET-enhanced photostability allows improved single-molecule tracking of proteins and protein complexes in live mammalian cells', *Nat Commun*, 9: 2520.

Basu, S., O. Shukron, D. Hall, P. Parutto, A. Ponjavic, D. Shah, W. Boucher, D. Lando, W. Zhang, N. Reynolds, L. H. Sober, A. Jartseva, R. Ragheb, X. Ma, J. Cramard, R. Floyd, J. Balmer, T. A. Drury, A. R. Carr, L. M. Needham, A. Aubert, G. Communie, K. Gor, M. Steindel, L. Morey, E. Blanco, T. Bartke, L. Di Croce, I. Berger, C. Schaffitzel, S. F. Lee, T. J. Stevens, D. Klenerman, B. D. Hendrich, D. Holcman, and E. D. Laue. 2023. 'Live-cell three-dimensional single-molecule tracking reveals modulation of enhancer dynamics by NuRD', *Nat Struct Mol Biol*, 30: 1628-39.

Blassberg, R., H. Patel, T. Watson, M. Gouti, V. Metzis, M. J. Delas, and J. Briscoe. 2022. 'Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation', *Nat Cell Biol*, 24: 633-44.

Bolstad, B. M., R. A. Irizarry, M. Astrand, and T. P. Speed. 2003. 'A comparison of normalization methods for high density oligonucleotide array data based on variance and bias', *Bioinformatics*, 19: 185-93.

Bompadre, O., and G. Andrey. 2019. 'Chromatin topology in development and disease', *Curr Opin Genet Dev*, 55: 32-38.

Boroviak, T., R. Loos, P. Bertone, A. Smith, and J. Nichols. 2014. 'The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification', *Nat Cell Biol*, 16: 516-28.

Buecker, C., R. Srinivasan, Z. Wu, E. Calo, D. Acampora, T. Faial, A. Simeone, M. Tan, T. Swigut, and J. Wysocka. 2014. 'Reorganization of enhancer patterns in transition from naive to primed pluripotency', *Cell Stem Cell*, 14: 838-53.

Carr, A. R., A. Ponjavic, S. Basu, J. McColl, A. M. Santos, S. Davis, E. D. Laue, D. Klenerman, and S. F. Lee. 2017. 'Three-Dimensional Super-Resolution in Eukaryotic Cells Using the Double-Helix Point Spread Function', *Biophys J*, 112: 1444-54.

Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y. H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W. K. Sung, N. D. Clarke, C. L. Wei, and H. H. Ng. 2008. 'Integration of external signaling pathways with the core transcriptional network in embryonic stem cells', *Cell*, 133: 1106-17.

Chovanec, P., A. J. Collier, C. Krueger, C. Varnai, C. I. Semprich, S. Schoenfelder, A. E. Corcoran, and P. J. Rugg-Gunn. 2021. 'Widespread reorganisation of pluripotent factor binding and gene regulatory interactions between human pluripotent states', *Nat Commun*, 12: 2098.

Conti, L., S. M. Pollard, T. Gorba, E. Reitano, M. Toselli, G. Biella, Y. Sun, S. Sanzone, Q. L. Ying, E. Cattaneo, and A. Smith. 2005. 'Niche-independent symmetrical self-renewal of a mammalian tissue stem cell', *PLoS Biol*, 3: e283.

Cremer, T., M. Cremer, B. Hubner, A. Silahtaroglu, M. Hendzel, C. Lanctot, H. Strickfaden, and C. Cremer. 2020. 'The Interchromatin Compartment Participates in the Structural and Functional Organization of the Cell Nucleus', *Bioessays*, 42: e1900132.

Crump, N. T., and T. A. Milne. 2019. 'Why are so many MLL lysine methyltransferases required for normal mammalian development?', *Cell Mol Life Sci*, 76: 2885-98.

Davies, O. R., C. Y. Lin, A. Radzisheuskaya, X. Zhou, J. Taube, G. Blin, A. Waterhouse, A. J. Smith, and S. Lowell. 2013. 'Tcf15 primes pluripotent cells for differentiation', *Cell Rep*, 3: 472-84.

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. 'Capturing chromosome conformation', *Science*, 295: 1306-11.

Denissov, S., H. Hofemeister, H. Marks, A. Kranz, G. Ciotta, S. Singh, K. Anastassiadis, H. G. Stunnenberg, and A. F. Stewart. 2014. 'Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant', *Development*, 141: 526-37.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. 2012. 'Topological domains in mammalian genomes identified by analysis of chromatin interactions', *Nature*, 485: 376-80.

Dorighi, K. M., T. Swigut, T. Henriques, N. V. Bhanu, B. S. Scruggs, N. Nady, C. D. Still, 2nd, B. A. Garcia, K. Adelman, and J. Wysocka. 2017. 'Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation', *Mol Cell*, 66: 568-76 e4.

Dostie, J., T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, T. A. Honan, E. D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R. D. Green, and J. Dekker. 2006. 'Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements', *Genome Res*, 16: 1299-309.

Douillet, D., C. C. Sze, C. Ryan, A. Piunti, A. P. Shah, M. Ugarenko, S. A. Marshall, E. J. Rendleman, D. Zha, K. A. Helmin, Z. Zhao, K. Cao, M. A. Morgan, B. D. Singer, E. T. Bartom, E. R. Smith, and A. Shilatifard. 2020. 'Uncoupling histone H3K4

trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation', *Nat Genet*, 52: 615-25.

Du, M., S. H. Stitzinger, J. H. Spille, W. K. Cho, C. Lee, M. Hijaz, A. Quintana, and Cisse, II. 2024. 'Direct observation of a condensate effect on super-enhancer controlled gene bursting', *Cell*, 187: 331-44 e17.

Du, P., M. Pirouz, J. Choi, A. J. Huebner, K. Clement, A. Meissner, K. Hochedlinger, and R. I. Gregory. 2018. 'An Intermediate Pluripotent State Controlled by MicroRNAs Is Required for the Naive-to-Primed Stem Cell Transition', *Cell Stem Cell*, 22: 851-64 e5.

Dunn, S. J., G. Martello, B. Yordanov, S. Emmott, and A. G. Smith. 2014. 'Defining an essential transcription factor program for naive pluripotency', *Science*, 344: 1156-60.

Durand, N. C., M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley, E. S. Lander, and E. L. Aiden. 2016. 'Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments', *Cell Syst*, 3: 95-8.

Eckersley-Maslin, M. A., A. Parry, M. Blotenburg, C. Krueger, Y. Ito, V. N. R. Franklin, M. Narita, C. S. D'Santos, and W. Reik. 2020. 'Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment', *Nat Struct Mol Biol*.

Edelstein, A., N. Amodaj, K. Hoover, R. Vale, and N. Stuurman. 2010. 'Computer control of microscopes using microManager', *Curr Protoc Mol Biol*, Chapter 14: Unit14 20.

Edelstein, A. D., M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman. 2014. 'Advanced methods of microscope control using muManager software', *J Biol Methods*, 1.

Elf, J., G. W. Li, and X. S. Xie. 2007. 'Probing transcription factor dynamics at the single-molecule level in a living cell', *Science*, 316: 1191-4.

Ernst, P., J. K. Fisher, W. Avery, S. Wade, D. Foy, and S. J. Korsmeyer. 2004. 'Definitive hematopoiesis requires the mixed-lineage leukemia gene', *Dev Cell*, 6: 437-43.

Etheridge, T. J., R. L. Boulineau, A. Herbert, A. T. Watson, Y. Daigaku, J. Tucker, S. George, P. Jonsson, M. Palayret, D. Lando, E. Laue, M. A. Osborne, D. Klenerman, S. F. Lee, and A. M. Carr. 2014. 'Quantification of DNA-associated proteins inside eukaryotic cells using single-molecule localization microscopy', *Nucleic Acids Res*, 42: e146.

Ewels, P., M. Magnusson, S. Lundin, and M. Kaller. 2016. 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32: 3047-8.

Forte, G., A. Buckle, S. Boyle, D. Marenduzzo, N. Gilbert, and C. A. Brackley. 2023. 'Transcription modulates chromatin dynamics and locus configuration sampling', *Nat Struct Mol Biol*, 30: 1275-85.

Glaser, S., S. Lubitz, K. L. Loveland, K. Ohbo, L. Robb, F. Schwenk, J. Seibler, D. Roellig, A. Kranz, K. Anastassiadis, and A. F. Stewart. 2009. 'The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis', *Epigenetics Chromatin*, 2: 5.

Glaser, S., J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K. R. Tufteland, R. Aasland, K. Anastassiadis, S. L. Ang, and A. F. Stewart. 2006. 'Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development', *Development*, 133: 1423-32.

Gomez-Garcia, P. A., S. Portillo-Ledesma, M. V. Neguembor, M. Pesaresi, W. Oweis, T. Rohrlich, S. Wieser, E. Meshorer, T. Schlick, M. P. Cosma, and M. Lakadamyali. 2021. 'Mesoscale Modeling and Single-Nucleosome Tracking Reveal Remodeling of Clutch Folding and Dynamics in Stem Cell Differentiation', *Cell Rep*, 34: 108614.

Goveas, N., C. Waskow, K. Arndt, J. A. Heuberger, Q. Y. Zhang, D. A. Alexopoulou, A. Dahl, W. A. Birchmeier, K. Anastassiadis, A. F. A. Stewart, and A. Kranz. 2021. 'MLL1 is required for maintenance of intestinal stem cells', *Plos Genetics*, 17.

Grimm, J. B., B. P. English, H. Choi, A. K. Muthusamy, B. P. Mehl, P. Dong, T. A. Brown, J. Lippincott-Schwartz, Z. Liu, T. Lionnet, and L. D. Lavis. 2016. 'Bright photoactivatable fluorophores for single-molecule imaging', *Nat Methods*, 13: 985-88.

Grinat, J., J. Heuberger, R. O. Vidal, N. Goveas, F. Kosel, A. Berenguer-Llergo, A. Kranz, A. Wulf-Goldenberg, D. Behrens, B. Melcher, S. Sauer, M. Vieth, E. Batlle, A. F. Stewart, and W. Birchmeier. 2020. 'The epigenetic regulator Mll1 is required for Wnt-driven intestinal tumorigenesis and cancer stemness', *Nat Commun*, 11: 6422.

Grinat, J., F. Kosel, N. Goveas, A. Kranz, D. Alexopoulou, K. Rajewsky, M. Sigal, A. F. Stewart, and J. Heuberger. 2022. 'Epigenetic modifier balances Mapk and Wnt signalling in differentiation of goblet and Paneth cells', *Life Sci Alliance*, 5.

Grosswendt, S., H. Kretzmer, Z. D. Smith, A. S. Kumar, S. Hetzel, L. Wittler, S. Klages, B. Timmermann, S. Mukherji, and A. Meissner. 2020. 'Epigenetic regulator function through mouse gastrulation', *Nature*.

Gu, B., T. Swigut, A. Spencley, M. R. Bauer, M. Chung, T. Meyer, and J. Wysocka. 2018. 'Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements', *Science*, 359: 1050-55.

Hackett, J. A., and M. A. Surani. 2014. 'Regulatory principles of pluripotency: from the ground state up', *Cell Stem Cell*, 15: 416-30.

Haghverdi, L., A. T. L. Lun, M. D. Morgan, and J. C. Marioni. 2018. 'Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors', *Nat Biotechnol*, 36: 421-27.

Hammal, F., P. de Langen, A. Bergon, F. Lopez, and B. Ballester. 2022. 'ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments', *Nucleic Acids Res*, 50: D316-D25.

Hansen, J. C., K. Maeshima, and M. J. Hendzel. 2021. 'The solid and liquid states of chromatin', *Epigenetics Chromatin*, 14: 50.

Hao, Y., S. Hao, E. Andersen-Nissen, W. M. Mauck, 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, and R. Satija. 2021. 'Integrated analysis of multimodal single-cell data', *Cell*, 184: 3573-87 e29.

Hayashi, K., H. Ohta, K. Kurimoto, S. Aramaki, and M. Saitou. 2011. 'Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells', *Cell*, 146: 519-32.

Hofemeister, H., G. Ciotta, J. Fu, P. M. Seibert, A. Schulz, M. Maresca, M. Sarov, K. Anastassiadis, and A. F. Stewart. 2011. 'Recombineering, transfection, Western, IP and ChIP methods for protein tagging via gene targeting or BAC transgenesis', *Methods*, 53: 437-52.

Hsieh, T. H., A. Weiner, B. Lajoie, J. Dekker, N. Friedman, and O. J. Rando. 2015. 'Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C', *Cell*, 162: 108-19.

Hsieh, T. S., C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, X. Darzacq, and R. Tjian. 2022. 'Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1', *Nat Genet*, 54: 1919-32.

Hsieh, T. S., C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, and X. Darzacq. 2020. 'Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding', *Mol Cell*, 78: 539-53 e8.

Hu, D., A. S. Garruss, X. Gao, M. A. Morgan, M. Cook, E. R. Smith, and A. Shilatifard. 2013. 'The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells', *Nat Struct Mol Biol*, 20: 1093-7.

Hubner, M. R., M. A. Eckersley-Maslin, and D. L. Spector. 2013. 'Chromatin organization and transcriptional regulation', *Curr Opin Genet Dev*, 23: 89-95.

Iida, S., S. Shinkai, Y. Itoh, S. Tamura, M. T. Kanemaki, S. Onami, and K. Maeshima. 2022. 'Single-nucleosome imaging reveals steady-state motion of interphase chromatin in living human cells', *Sci Adv*, 8: eabn5626.

Jaenisch, R., and A. Bird. 2003. 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals', *Nat Genet*, 33 Suppl: 245-54.

Joshi, O., S. Y. Wang, T. Kuznetsova, Y. Atlasi, T. Peng, P. J. Fabre, E. Habibi, J. Shaik, S. Saeed, L. Handoko, T. Richmond, M. Spivakov, D. Burgess, and H. G. Stunnenberg. 2015. 'Dynamic Reorganization of Extremely Long-Range Promoter-Promoter Interactions between Two States of Pluripotency', *Cell Stem Cell*, 17: 748-57.

Kalkan, T., N. Olova, M. Roode, C. Mulas, H. J. Lee, I. Nett, H. Marks, R. Walker, H. G. Stunnenberg, K. S. Lilley, J. Nichols, W. Reik, P. Bertone, and A. Smith. 2017. 'Tracking the embryonic stem cell transition from ground state pluripotency', *Development*, 144: 1221-34.

Kalkan, T., and A. Smith. 2014. 'Mapping the route from naive pluripotency to lineage specification', *Philos Trans R Soc Lond B Biol Sci*, 369.

Kassis, J. A., J. A. Kennison, and J. W. Tamkun. 2017. 'Polycomb and Trithorax Group Genes in Drosophila', *Genetics*, 206: 1699-725.

Kinoshita, M., M. Barber, W. Mansfield, Y. Cui, D. Spindlow, G. G. Stirparo, S. Dietmann, J. Nichols, and A. Smith. 2021. 'Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency', *Cell Stem Cell*, 28: 453-71 e8.

Klymenko, T., and J. Muller. 2004. 'The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins', *EMBO Rep*, 5: 373-7.

Krietenstein, N., S. Abraham, S. V. Venev, N. Abdennur, J. Gibcus, T. S. Hsieh, K. M. Parsi, L. Yang, R. Maehr, L. A. Mirny, J. Dekker, and O. J. Rando. 2020. 'Ultrastructural Details of Mammalian Chromosome Architecture', *Mol Cell*, 78: 554-65 e7.

Kruse, K., C. B. Hug, and J. M. Vaquerizas. 2020. 'FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data', *Genome Biol*, 21: 303.

Kubo, N., P. B. Chen, R. Hu, Z. Ye, H. Sasaki, and B. Ren. 2024. 'H3K4me1 facilitates promoter-enhancer interactions and gene activation during embryonic stem cell differentiation', *Mol Cell*, 84: 1742-52 e5.

Lackner, A., R. Sehlke, M. Garmhausen, G. Giuseppe Stirparo, M. Huth, F. Titz-Teixeira, P. van der Lelij, J. Ramesmayer, H. F. Thomas, M. Ralser, L. Santini, E. Galimberti, M. Sarov, A. F. Stewart, A. Smith, A. Beyer, and M. Leeb. 2021. 'Cooperative genetic networks drive embryonic stem cell transition from naive to formative pluripotency', *EMBO J*: e105776.

Lackner, Andreas, Robert Sehlke, Marius Garmhausen, Giuliano Giuseppe Stirparo, Michelle Huth, Fabian Titz-Teixeira, Petra van der Lelij, Julia Ramesmayer, Henry Fabian Thomas, Meryem Ralser, Laura Santini, Mihail Sarov, A. Francis Stewart, Austin Smith,

Andreas Beyer, and Martin Leeb. 2020. 'Cooperative molecular networks drive a mammalian cell state transition', *bioRxiv*: 2020.03.23.000109.

Ladopoulos, V., H. Hofemeister, M. Hoogenkamp, A. D. Riggs, A. F. Stewart, and C. Bonifer. 2013. 'The histone methyltransferase KMT2B is required for RNA polymerase II association and protection from DNA methylation at the MagohB CpG island promoter', *Mol Cell Biol*, 33: 1383-93.

Lange, M., V. Bergen, M. Klein, M. Setty, B. Reuter, M. Bakhti, H. Lickert, M. Ansari, J. Schniering, H. B. Schiller, D. Pe'er, and F. J. Theis. 2022. 'CellRank for directed single-cell fate mapping', *Nat Methods*, 19: 159-70.

Langmead, B., and S. L. Salzberg. 2012. 'Fast gapped-read alignment with Bowtie 2', *Nat Methods*, 9: 357-9.

Lee, S., C. R. Bakker, C. Vitzthum, B. H. Alver, and P. J. Park. 2022. 'Pairs and Pairix: a file format and a tool for efficient storage and retrieval for Hi-C read pairs', *Bioinformatics*, 38: 1729-31.

Leeb, M., S. Dietmann, M. Paramor, H. Niwa, and A. Smith. 2014. 'Genetic exploration of the exit from self-renewal using haploid embryonic stem cells', *Cell Stem Cell*, 14: 385-93.

Lew, M., et al. 2013. 'Easy-DHPSF open-source software for three-dimensional localization of single molecules with precision beyond the optical diffraction limit.', *Protocol Exchange*.

Li, H., and R. Durbin. 2009. 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25: 1754-60.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.

Li, L., H. Liu, P. Dong, D. Li, W. R. Legant, J. B. Grimm, L. D. Lavis, E. Betzig, R. Tjian, and Z. Liu. 2016. 'Real-time imaging of Huntingtin aggregates diverting target search and gene transcription', *Elife*, 5.

Liao, Y., G. K. Smyth, and W. Shi. 2014. 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30: 923-30.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science*, 326: 289-93.

Logie, C., and A. F. Stewart. 1995. 'Ligand-regulated site-specific recombination', *Proc Natl Acad Sci U S A*, 92: 5940-4.

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

Lubitz, S., S. Glaser, J. Schaft, A. F. Stewart, and K. Anastassiadis. 2007. 'Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2', *Mol Biol Cell*, 18: 2356-66.

Ma, H., L. C. Tu, Y. C. Chung, A. Naseri, D. Grunwald, S. Zhang, and T. Pederson. 2019. 'Cell cycle- and genomic distance-dependent dynamics of a discrete chromosomal region', *J Cell Biol*, 218: 1467-77.

Marks, H., T. Kalkan, R. Menafra, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A. F. Stewart, A. Smith, and H. G. Stunnenberg. 2012. 'The transcriptional and epigenomic foundations of ground state pluripotency', *Cell*, 149: 590-604.

Mas, G., E. Blanco, C. Ballare, M. Sanso, Y. G. Spill, D. Hu, Y. Aoi, F. Le Dily, A. Shilatifard, M. A. Marti-Renom, and L. Di Croce. 2018. 'Promoter bivalency favors an open chromatin architecture in embryonic stem cells', *Nat Genet*, 50: 1452-62.

Miron, E., R. Oldenkamp, J. M. Brown, D. M. S. Pinto, C. S. Xu, A. R. Faria, H. A. Shaban, J. D. P. Rhodes, C. Innocent, S. de Ornellas, H. F. Hess, V. Buckle, and L. Schermelleh. 2020. 'Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin', *Sci Adv*, 6.

Mishra, B. P., K. M. Zaffuto, E. L. Artinger, T. Org, H. K. Mikkola, C. Cheng, M. Djabali, and P. Ernst. 2014. 'The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis', *Cell Rep*, 7: 1239-47.

Mohammed, H., I. Hernando-Herraez, A. Savino, A. Scialdone, I. Macaulay, C. Mulas, T. Chandra, T. Voet, W. Dean, J. Nichols, J. C. Marioni, and W. Reik. 2017. 'Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation', *Cell Rep*, 20: 1215-28.

Nagashima, R., K. Hibino, S. S. Ashwin, M. Babokhov, S. Fujishiro, R. Imai, T. Nozaki, S. Tamura, T. Tani, H. Kimura, M. Shribak, M. T. Kanemaki, M. Sasai, and K. Maeshima. 2019. 'Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II', *J Cell Biol*, 218: 1511-30.

Neagu, A., E. van Genderen, I. Escudero, L. Verwegen, D. Kurek, J. Lehmann, J. Stel, R. A. M. Dirks, G. van Mierlo, A. Maas, C. Eleveld, Y. Ge, A. T. den Dekker, R. W. W. Brouwer, IJcken W. F. J. van, M. Modic, M. Drukker, J. H. Jansen, N. C. Rivron, E. B. Baart, H. Marks, and D. Ten Berge. 2020. 'In vitro capture and characterization of embryonic rosette-stage pluripotency between naive and primed states', *Nat Cell Biol*, 22: 534-45.

Nichols, J., and A. Smith. 2009. 'Naive and primed pluripotent states', *Cell Stem Cell*, 4: 487-92.

Nora, E. P., A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, and B. G. Bruneau. 2017. 'Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization', *Cell*, 169: 930-44 e22.

Nozaki, T., R. Imai, M. Tanbo, R. Nagashima, S. Tamura, T. Tani, Y. Joti, M. Tomita, K. Hibino, M. T. Kanemaki, K. S. Wendt, Y. Okada, T. Nagai, and K. Maeshima. 2017. 'Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging', *Mol Cell*, 67: 282-93 e7.

Open2C, Nezar Abdennur, Geoffrey Fudenberg, Ilya M. Flyamer, Aleksandra A. Galitsyna, Anton Goloborodko, Maxim Imakaev, and Sergey V. Venev. 2023. 'Pairtools: from sequencing data to chromosome contacts', *bioRxiv*: 2023.02.13.528389.

Platania, A., C. Erb, M. Barbieri, B. Molcrette, E. Grandgirard, M. A. C. de Kort, W. Pomp, K. Meaburn, T. Taylor, V. M. Shchuka, S. Kocanova, M. Nazarova, G. M. Oliveira, J. A. Mitchell, E. Soutoglou, T. L. Lenstra, N. Molina, A. Papantonis, K. Bystricky, and T. Sexton. 2024. 'Transcription processes compete with loop extrusion to homogenize promoter and enhancer dynamics', *Sci Adv*, 10: eadq0987.

Quinlan, A. R., and I. M. Hall. 2010. 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26: 841-2.

Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. 2014. 'A 3D map of the

human genome at kilobase resolution reveals principles of chromatin looping', *Cell*, 159: 1665-80.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. 'NIH Image to ImageJ: 25 years of image analysis', *Nat Methods*, 9: 671-5.

Schuettengruber, B., H. M. Bourbon, L. Di Croce, and G. Cavalli. 2017. 'Genome Regulation by Polycomb and Trithorax: 70 Years and Counting', *Cell*, 171: 34-57.

Seeber, A., M. H. Hauer, and S. M. Gasser. 2018. 'Chromosome Dynamics in Response to DNA Damage', *Annu Rev Genet*, 52: 295-319.

Shahsavari, Arash, Andi Munteanu, and Irina Mohorianu. 2022. 'ClustAssess: tools for assessing the robustness of single-cell clustering', *bioRxiv*: 2022.01.31.478592.

Silva, J., and A. Smith. 2008. 'Capturing pluripotency', *Cell*, 132: 532-6.

Singh, A. M., Y. Sun, L. Li, W. Zhang, T. Wu, S. Zhao, Z. Qin, and S. Dalton. 2015. 'Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency', *Stem Cell Reports*, 5: 323-36.

Smith, A. 2017. 'Formative pluripotency: the executive phase in a developmental continuum', *Development*, 144: 365-73.

Som, A., C. Harder, B. Greber, M. Siatkowski, Y. Paudel, G. Warsow, C. Cap, H. Scholer, and G. Fuellen. 2010. 'The PluriNetWork: an electronic representation of the network underlying pluripotency in mouse, and its applications', *PLoS One*, 5: e15165.

Steindel, M., I. Orsine de Almeida, S. Strawbridge, V. Chernova, D. Holcman, A. Ponjavic, and S. Basu. 2022. 'Studying the Dynamics of Chromatin-Binding Proteins in Mammalian Cells Using Single-Molecule Localization Microscopy', *Methods Mol Biol*, 2476: 209-47.

Stevens, T. J., D. Lando, S. Basu, L. P. Atkinson, Y. Cao, S. F. Lee, M. Leeb, K. J. Wohlfahrt, W. Boucher, A. O'Shaughnessy-Kirwan, J. Cramard, A. J. Faure, M. Ralser, E. Blanco, L. Morey, M. Sanso, M. G. S. Palayret, B. Lehner, L. Di Croce, A. Wutz, B. Hendrich, D. Klenerman, and E. D. Laue. 2017. '3D structures of individual mammalian genomes studied by single-cell Hi-C', *Nature*, 544: 59-64.

Stirling, D. R., M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, and A. Goodman. 2021. 'CellProfiler 4: improvements in speed, utility and usability', *BMC Bioinformatics*, 22: 433.

Strawbridge, Stanley E, Guy B Blanchard, Austin Smith, Hillel Kugler, and Graziano Martello. 2020. 'Embryonic stem cells commit to differentiation by symmetric divisions following a variable lag period', *bioRxiv*: 2020.06.17.157578.

Talbert, P. B., and S. Henikoff. 2021. 'The Yin and Yang of Histone Marks in Transcription', *Annu Rev Genomics Hum Genet*, 22: 147-70.

Tolhuis, B., R. J. Palstra, E. Splinter, F. Grosveld, and W. de Laat. 2002. 'Looping and interaction between hypersensitive sites in the active beta-globin locus', *Mol Cell*, 10: 1453-65.

Tosolini, M., V. Brochard, P. Adenot, M. Chebrout, G. Grillo, V. Navia, N. Beaujean, C. Francastel, A. Bonnet-Garnier, and A. Jouneau. 2018. 'Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells', *Sci Rep*, 8: 5776.

Van den Berge, K., H. Roux de Bezieux, K. Street, W. Saelens, R. Cannoodt, Y. Saeys, S. Dudoit, and L. Clement. 2020. 'Trajectory-based differential expression analysis for single-cell sequencing data', *Nat Commun*, 11: 1201.

Wang, W., C. Lin, D. Lu, Z. Ning, T. Cox, D. Melvin, X. Wang, A. Bradley, and P. Liu. 2008. 'Chromosomal transposition of PiggyBac in mouse embryonic stem cells', *Proc Natl Acad Sci U S A*, 105: 9290-5.

Wolf, F. A., P. Angerer, and F. J. Theis. 2018. 'SCANPY: large-scale single-cell gene expression data analysis', *Genome Biol*, 19: 15.

Xie, G., J. E. Lee, A. D. Senft, Y. K. Park, Y. Jang, S. Chakraborty, J. J. Thompson, K. McKernan, C. Liu, T. S. Macfarlan, P. P. Rocha, W. Peng, and K. Ge. 2023. 'MLL3/MLL4 methyltransferase activities control early embryonic development and embryonic stem cell differentiation in a lineage-selective manner', *Nat Genet*, 55: 693-705.

Xu, W., Q. Zhong, D. Lin, Y. Zuo, J. Dai, G. Li, and G. Cao. 2021. 'CoolBox: a flexible toolkit for visual analysis of genomics data', *BMC Bioinformatics*, 22: 489.

Yang, P., S. J. Humphrey, S. Cinghu, R. Pathania, A. J. Oldfield, D. Kumar, D. Perera, J. Y. H. Yang, D. E. James, M. Mann, and R. Jothi. 2019. 'Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency', *Cell Syst*, 8: 427-45 e10.

Yang, S. H., M. Andrabi, R. Biss, S. Murtuza Baker, M. Iqbal, and A. D. Sharrocks. 2019. 'ZIC3 Controls the Transition from Naive to Primed Pluripotency', *Cell Rep*, 27: 3215-27 e6.

Yang, S. H., T. Kalkan, C. Morissroe, H. Marks, H. Stunnenberg, A. Smith, and A. D. Sharrocks. 2014. 'Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency', *Cell Rep*, 7: 1968-81.

Yang, T., F. Zhang, G. G. Yardimci, F. Song, R. C. Hardison, W. S. Noble, F. Yue, and Q. Li. 2017. 'HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient', *Genome Res*, 27: 1939-49.

Ying, Q. L., J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith. 2008. 'The ground state of embryonic stem cell self-renewal', *Nature*, 453: 519-23.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. 2008. 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biol*, 9: R137.

Zhao, Z., G. Tavoosidana, M. Sjolinder, A. Gondor, P. Mariano, S. Wang, C. Kanduri, M. Lezcano, K. S. Sandhu, U. Singh, V. Pant, V. Tiwari, S. Kurukuti, and R. Ohlsson. 2006. 'Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions', *Nat Genet*, 38: 1341-7.

Zheng, G. X., J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, and J. H. Bielas. 2017. 'Massively parallel digital transcriptional profiling of single cells', *Nat Commun*, 8: 14049.

Zidovska, A. 2020. 'Chromatin: Liquid or Solid?', *Cell*, 183: 1737-39.

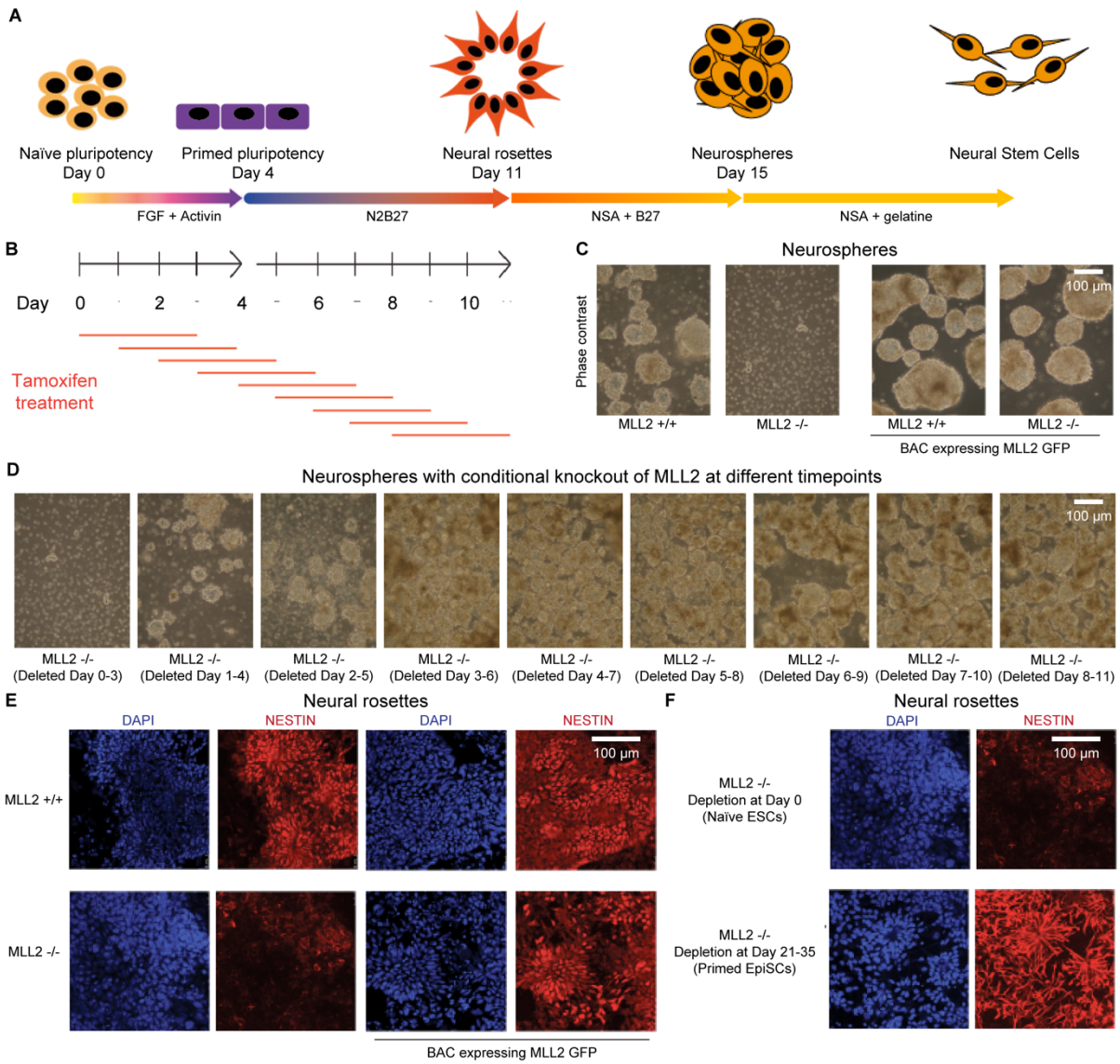


Figure 1. MLL2, which is required for neuroectoderm differentiation, exerts its effect during a priming window. **(A)** Schematic showing protocol for priming of naïve pluripotent cells (PCs) and subsequent neuroectoderm differentiation. **(B)** Time course strategy for tamoxifen induced conditional knockout (cKO) of *Mll2*. **(C)** Phase contrast images at the neurosphere stage of control (MLL2 +/+) and *Mll2* cKO (MLL2-/-) cells show that *Mll2* cKO cells fail to generate neural stem cells or neurospheres. Expression of exogenous GFP-tagged MLL2 from an integrated bacterial artificial chromosome (BAC) rescues failure to form neural stem cells and neurospheres in cells lacking endogenous *Mll2* (n=3). **(D)** Phase contrast images at the neurosphere stage from cells subjected to tamoxifen-induced *Mll2* cKO at different days during neuroectoderm differentiation as indicated above and in **Figure S1D**. **(E)** Fluorescence images of DAPI stain (blue) and NESTIN expression (red) at the neural rosette stage of control (MLL2 +/+) and *Mll2* cKO (MLL2-/-) cells show that *Mll2*cKO cells fail to generate neural rosettes while ectopic expression of exogenous MLL2 rescues this failure. **(F)** MLL2 is required prior to priming for successful ES cell differentiation to neural rosettes.

Fluorescence images of DAPI stain (blue) and NESTIN expression (red) at the neural rosette stage of control (MLL2 +/+), *Mll2*cKO cells where tamoxifen-induced MLL2 depletion is carried out at Day 0 (MLL2-/- naïve ESCs) or after 21-35 days of FGF-Activin addition (MLL2-/- primed ESCs).

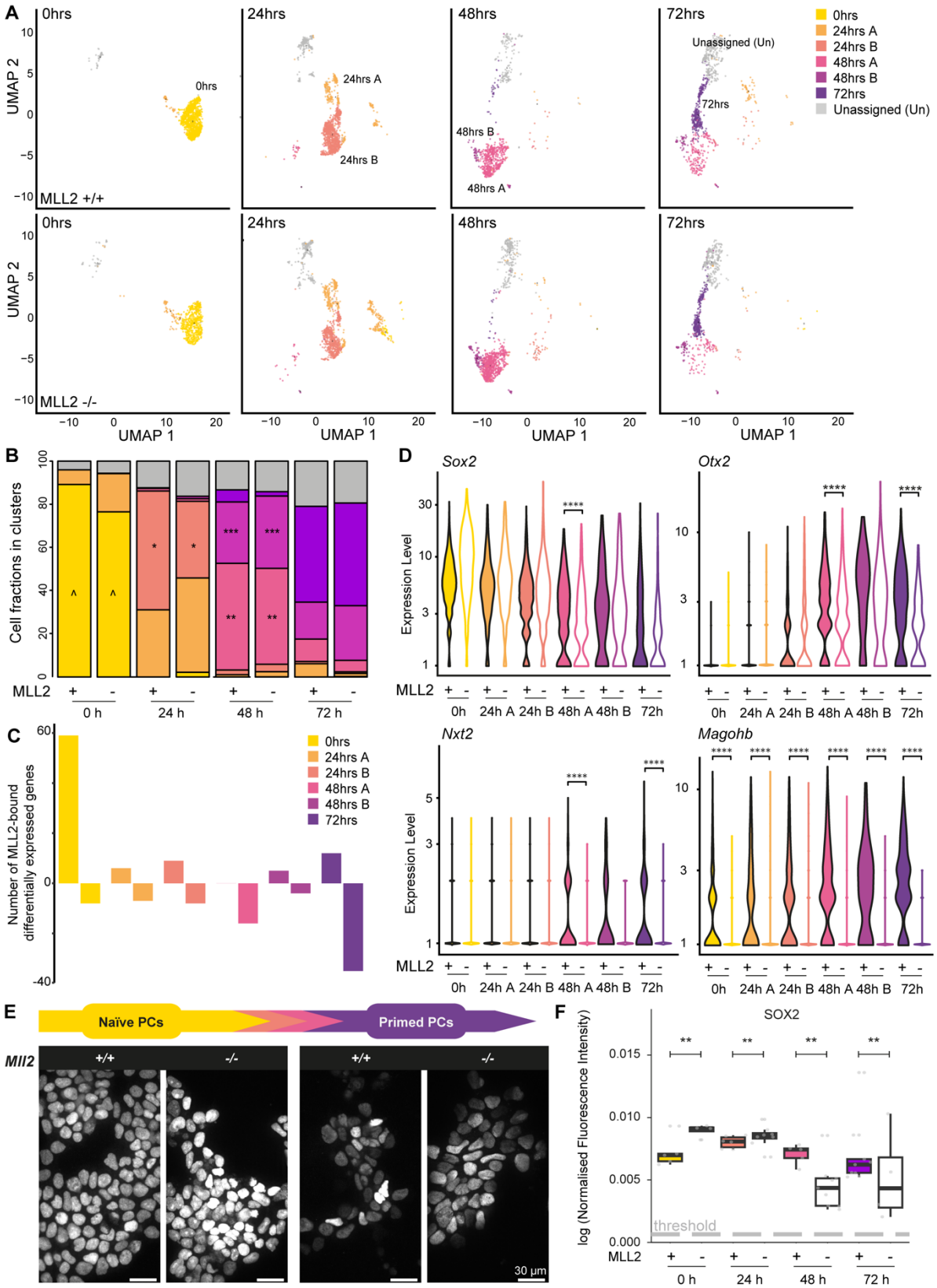


Figure 2. MLL2 has minor effect on cell types generated during priming but increases expression level of key neuroectoderm-specifying genes in primed pluripotent cells.

(A) Uniform manifold approximation and projection (UMAP) plot showing wild-type and *Mll2* cKO cells during transition from naïve to primed pluripotency with cells collected from each time point and sample shown on separate plots. Cells are coloured differently if they have been assigned to different clusters (see **Methods**). This shows that wild-type and *Mll2* cKO cells both contribute to all assigned clusters with one/two clusters predominating each timepoint. Number of cells: 1215/638 (0 hr +/- MLL2), 428/532 (24 hr A +/- MLL2), 998/671 (24 hr B +/- MLL2), 759/914 (48 hr A +/- MLL2), 133/223 (48 hr B +/- MLL2), 348/343 (72 hr +/- MLL2) and 716/819 (Unassigned +/- MLL2). **(B)** Fraction of cells in wild-type and *Mll2* cKO cells show initial differences in pluripotent cell types that resolve into similar primed pluripotent cell types (^p = $2.48*10^{-6}$, *p = 0.004, **p = 0.043, ***p = 0.011, Fisher's exact test). Number of cells: 1308/700 (0 hr +/- MLL2), 1486/1335 (24 hr +/- MLL2), 817/1271 (48 hr +/- MLL2), 986/834 (72 hr +/- MLL2). **(C)** Bar plot of the number of up-/down-regulated differentially expressed genes (DEGs) when comparing wild-type and *Mll2* cKO cells for each cell cluster identified in **Figure 2** ($p < 0.001$, $\log_2FC > 0.5$ or $\log_2FC < -0.5$). Colours depict clusters identified in **Figure 2**. Genes are filtered to ensure they have MLL2 bound within 1 kb of the gene transcription start site. **(D)** MLL2 up-regulates *Sox2* and *Otx2* transcription during priming. Violin plots generated from wild-type and *Mll2* cKO cells show the transcript level in each cluster for the differentially expressed genes *Sox2*, *Otx2*, *Nxt2* and *Magohb*. Number of cells: 1215/638 (0 hr +/- MLL2), 428/532 (24 hr A +/- MLL2), 998/671 (24 hr B +/- MLL2), 759/914 (48 hr A +/- MLL2), 133/223 (48 hr B +/- MLL2), 348/343 (72 hr +/- MLL2). [Wilcoxon rank sum tests with Bonferroni correction for *Sox2*, $p = 10^{-4}$ (48 hr A +/- MLL2); for *Otx2*, $p = 10^{-11}$ (48 hr A +/- MLL2), $p = 10^{-9}$ (72 hr +/- MLL2); for *Nxt2*, $p = 10^{-28}$ (48 hr A +/- MLL2), $p = 10^{-8}$ (72 hr +/- MLL2); for *Magohb*, $p = 10^{-52}$ (0 hr +/- MLL2), $p = 10^{-64}$ (24 hr A +/- MLL2), $p = 10^{-37}$ (24 hr B +/- MLL2), $p = 10^{-146}$ (48 hr A +/- MLL2), $p = 10^{-26}$ (48 hr B +/- MLL2), $p = 10^{-68}$ (72 hr +/- MLL2)]. **(E)** MLL2 up-regulates SOX2 protein levels in primed pluripotent cells. Representative immunofluorescence images of SOX2 at 0 and 72 hours of priming in *Mll2* cKO and control cells. **(F)** Boxplot of SOX2 expression at 0, 24, 48 and 72 hours of priming coloured by timepoint. Mean fluorescence per cell is shown as grey dots for each field of view and background threshold as a grey dotted line. Number of images (~10-100 cells/image): 6/4 (0 hr +/- MLL2), 6/10 (24 hr +/- MLL2), 5/10 (48 hr +/- MLL2), 11/4 (72 hr +/- MLL2). [unpaired two-sided t-test, p = 0.007 (0 hr +/- MLL2), p = 0.008 (24 hr +/- MLL2), p = 0.007 (48 hr +/- MLL2), p = 0.007 (72 hr +/- MLL2).]

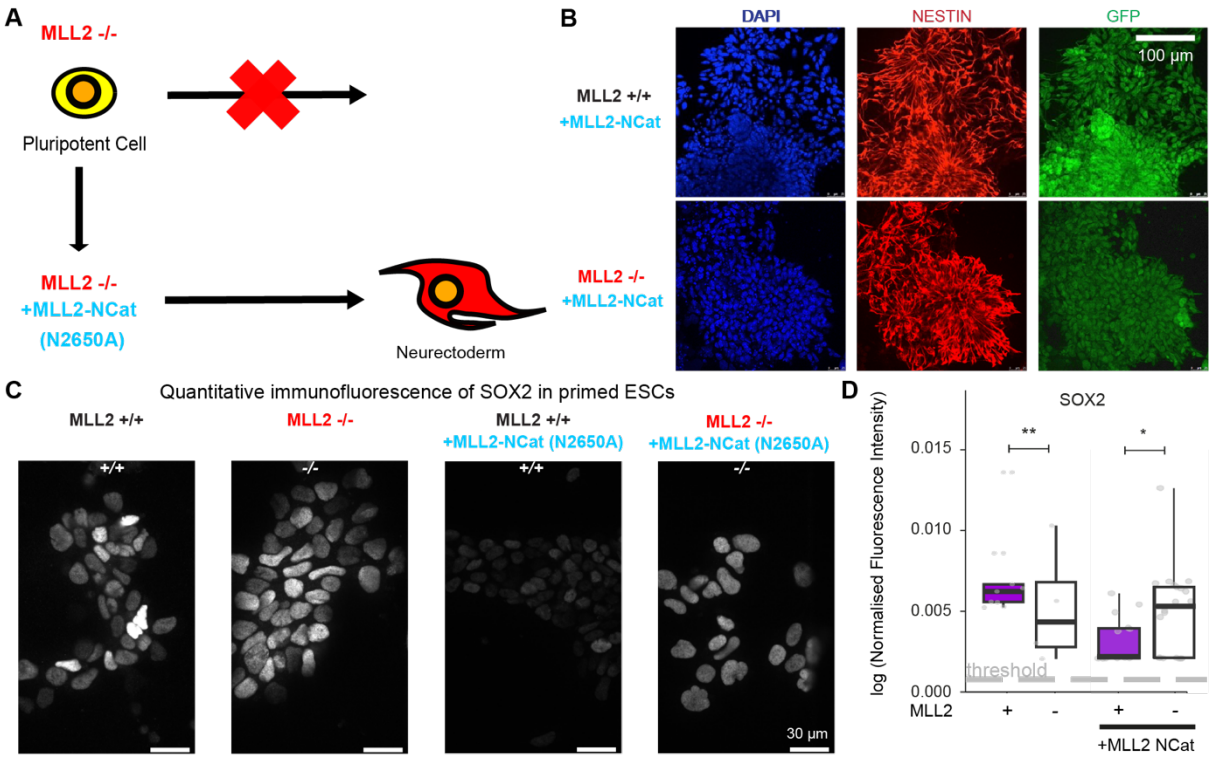


Figure 3. Introduction of catalytically-dead MLL2 (MLL2 NCat) rescues neuroectoderm differentiation and transcriptional changes during priming. (A) Schematic showing how introduction of a catalytically-dead MLL2 (MLL2 NCat) to MLL2 cKO cells rescues differentiation. **(B)** Fluorescence images of DAPI stain (blue), NESTIN expression (red) and expression of the GFP-tagged MLL2 NCat (green) at the neural rosette stage show that ectopic expression of exogenous MLL2 rescues failure of MLL2 cKO cells to generate neural rosettes (see **Figure 1**). (top) MLL2 +/+ cells ectopically expressing MLL2 NCat. (bottom) MLL2 -/- cells ectopically expressing MLL2 NCat. **(C)** SOX2 protein levels no longer decrease in MLL2 cKO primed pluripotent cells upon ectopic expression of exogenous MLL2. Representative immunofluorescence images of SOX2 in primed ESCs fors MLL2 cKO and control cells with/without ectopic expression of MLL2 NCat. **(D)** Boxplot of SOX2 expression in primed ESCs for MLL2 cKO and control cells with/without ectopic expression of MLL2 NCat. Mean fluorescence per cell is shown shown as grey dots for each field of view and background threshold as a grey dotted line. Number of images (10-100 cells/image): 11/4 (72 hr +/- MLL2), 11/21 (72 hr +/- MLL2 +MLL2NCat-GFP). [unpaired two-sided t-test, **p = 0.007 (72 hr +/- MLL2), *p = 0.03 (72 hr +/- MLL2 +MLL2NCat-GFP).]

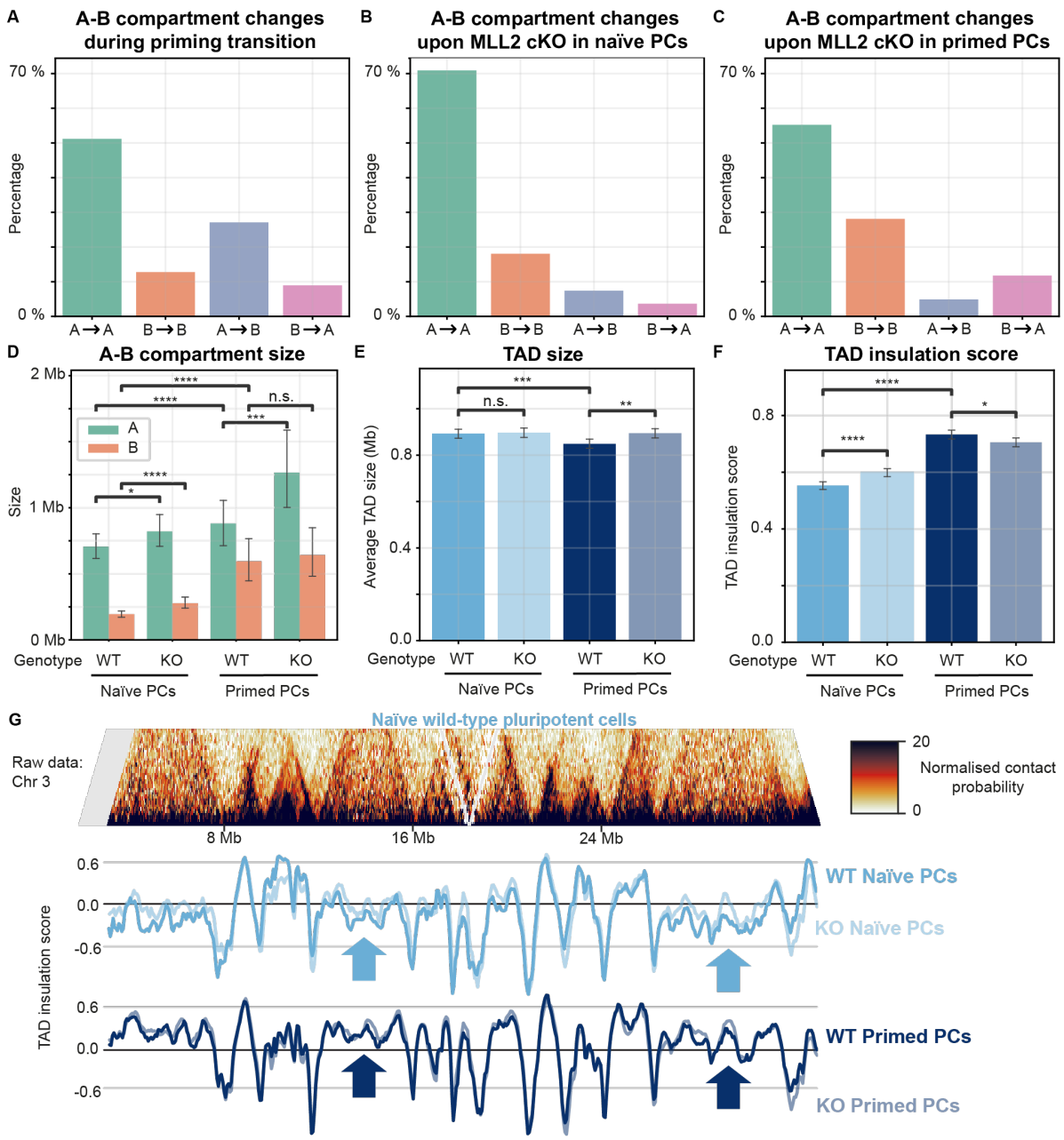


Figure 4. MLL2 has small effect on compartmentalisation of the genome into A-B compartments and TADs (topologically associating domain). (A) Percentage of 40 kb genomic bins that transition between A and B compartments when comparing naïve and primed wild-type PCs. **(B)** Percentage of 40 kb genomic bins that transition between A and B compartments when comparing wild-type (WT) and MLL2 cKO (KO) naïve PCs. **(C)** Percentage of 40 kb genomic bins that transition between A and B compartments when comparing WT and KO primed PCs. **(D)** Barplot showing mean A/B compartment size in WT and KO conditions for naïve and primed PCs, with error bars represent 95 % confidence interval. [n = 3031/3010, 2487/2466, 1854/1833 and 1435/1414 A/B compartments identified for WT naïve, KO naïve, WT primed and KO primed PCs respectively; Mann-Whitney-

Wilcoxon two-sided test with Benjamini-Hochberg correction, $p = 10^{-12}/10^{-8}$ for WT naïve vs primed A/B compartment size; $p = 0.04/10^{-12}$ for WT vs KO naïve PC A/B compartment size; $p = 10^{-4}/0.38$ for WT vs KO primed PC A/B compartment size] **(E)** Barplot of mean TAD size, with error bars represent 95 % confidence interval, shows that KO increases TAD size in primed but not in naïve PCs. [n = 2831, 2820, 2980 and 2828 TADs identified for WT naïve, KO naïve, WT primed and KO primed PCs respectively; Mann-Whitney-Wilcoxon two-sided test with Benjamini-Hochberg correction, $p = 10^{-4}$ for WT naïve vs primed PCs; p = 0.7 for WT vs KO naïve PCs; $p = 10^{-3}$ for WT vs KO primed PCs] **(F)** Barplot of mean TAD insulation score, with error bars represent 95 % confidence interval, shows stabilisation of TADs when comparing naïve to primed PC scores, with more subtle changes observed between WT and KO samples. [n = 2831, 2820, 2980 and 2828 TADs identified for WT naïve, KO naïve, WT primed and KO primed PCs respectively; Mann-Whitney-Wilcoxon two-sided test with Benjamini-Hochberg correction, $p = 10^{-76}$ for WT naïve vs primed PCs; $p = 10^{-8}$ for WT vs KO naïve PCs; p = 0.02 for WT vs KO primed PCs] **(G)** (Above) Normalised Micro-C contact map for chromosome 3 showing density of contacts allows visual identification of topologically associating domains (TADs). (Below) TAD insulation scores calculated for naïve (light blue) and primed PCs (dark blue) of priming for WT (darker) and KO (lighter) conditions showing stabilisation of TADs when comparing naïve to primed PC scores, as indicated by arrows, with more subtle changes observed between WT and KO samples.

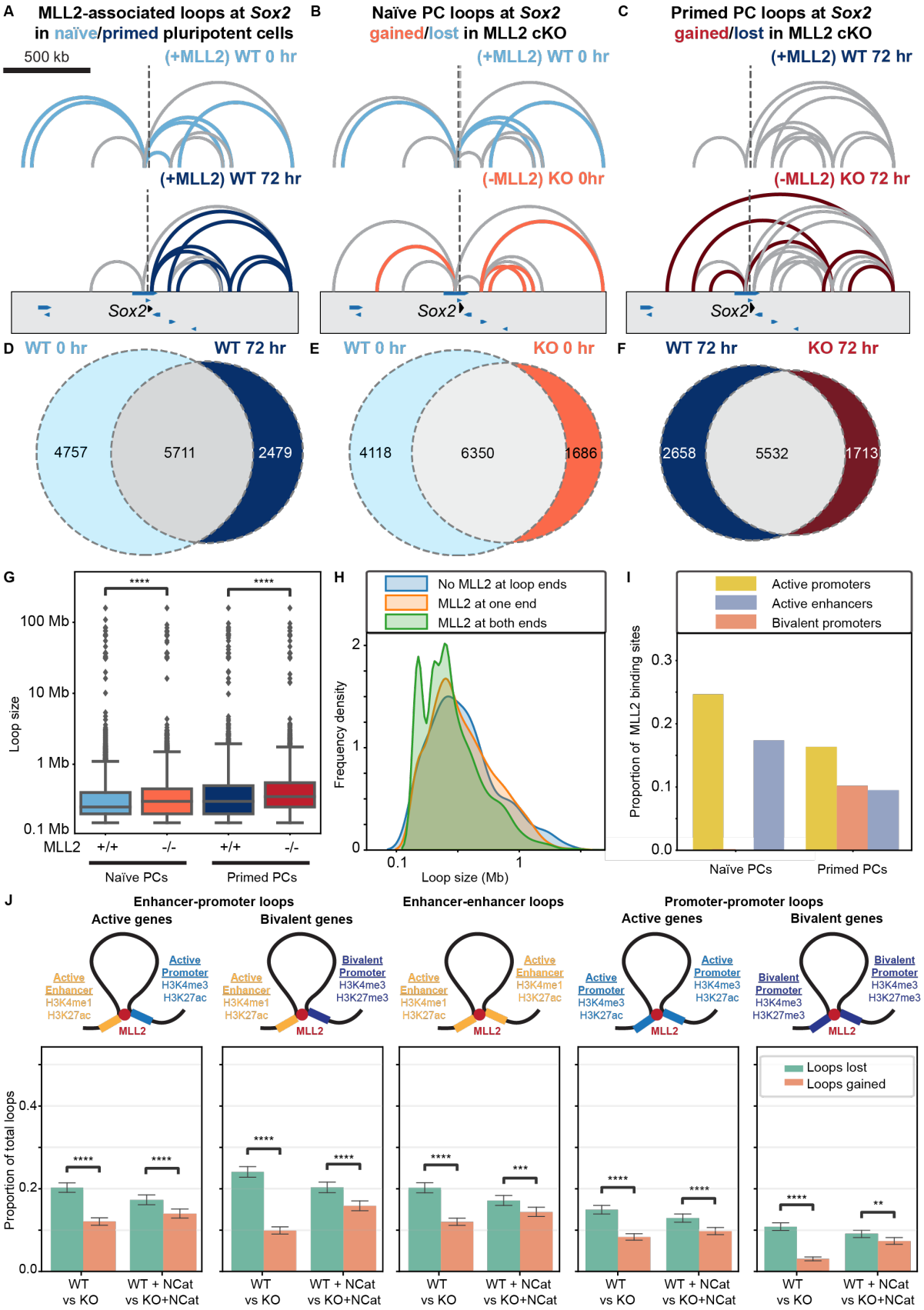

A
MLL2-associated loops at Sox2 in naïve/primed pluripotent cells
500 kb
(+MLL2) WT 0 hr
(+MLL2) WT 72 hr
Sox2
B
Naïve PC loops at Sox2 gained/lost in MLL2 cKO
(+MLL2) WT 0 hr
(-MLL2) KO 0hr
Sox2
C
Primed PC loops at Sox2 gained/lost in MLL2 cKO
(+MLL2) WT 72 hr
(-MLL2) KO 72 hr
Sox2
D
WT 0 hr
WT 72 hr
4757
5711
2479
E
WT 0 hr
KO 0 hr
4118
6350
1686
F
WT 72 hr
KO 72 hr
2658
5532
1713
G
Loop size
100 Mb
10 Mb
1 Mb
0.1 Mb
****
****
MLL2
+/+
-/-
+/+
-/-
Naïve PCs
Primed PCs
H
No MLL2 at loop ends
MLL2 at one end
MLL2 at both ends
Frequency density
Loop size (Mb)
I
Active promoters
Active enhancers
Bivalent promoters
Proportion of MLL2 binding sites
Naïve PCs
Primed PCs
J
Enhancer-promoter loops
Active genes
Bivalent genes
Enhancer-enhancer loops
Promoter-promoter loops
Active genes
Bivalent genes
Active Enhancer H3K4me1 H3K27ac
Active Promoter H3K4me3 H3K27ac
Bivalent Promoter H3K4me3 H3K27me3
MLL2
Loops lost
Loops gained
Proportion of total loops
WT vs KO
WT + NCat vs KO+NCat

Figure 5. MLL2 influences 3D chromatin looping but does not require its catalytic activity to form enhancer-promoter loops at bivalent genes in primed pluripotent cells. **(A)** High-confidence MLL2-associated loops at *Sox2* gene in naïve (above) and primed PCs (below) for wild-type (WT) cells. Grey loops are common to both conditions, light blue to naïve PCs and dark blue to primed PCs ($p < 0.02$). Grey dotted line represents transcription start site of gene locus. MLL2-associated loops are defined by MLL2 binding sites present at both sides of the loop. **(B)** High-confidence MLL2-associated loops at *Sox2* gene in naïve PCs for wild-type (above, WT) and MLL2 cKO cells (below, KO). Grey loops are common to both conditions, light blue to WT naïve PCs and light red to KO naïve PCs ($p < 0.02$). **(C)** High-confidence MLL2-associated loops at *Sox2* gene in primed PCs for WT (above) and KO cells (below). Grey loops are common to both conditions, dark blue to WT primed PCs and dark red to KO primed PCs ($p < 0.02$). **(D)** Euler diagram showing overlap of MLL2-associated loops detected in naïve PCs (0 hr) vs in primed PCs (72 hr). **(E)** Euler diagram showing overlap of MLL2-associated loops detected in WT vs KO cells for naïve PCs (0 hr). **(F)** Euler diagram showing overlap of MLL2-associated loops detected in WT vs KO cells for primed PCs (72 hr). **(G)** Box plot showing loop size distributions in WT vs KO cells for naïve and primed PCs. [Mann-Whitney-Wilcoxon two-sided test, $p = 10^{-24}$ for WT vs KO naïve ESCs; $p = 10^{-48}$ for WT vs KO primed ESCs] **(H)** Histogram showing distribution of loop sizes in WT cells for loops without MLL2, with MLL2 bound at one of the loop anchors or with MLL2 bound at both of the loop anchors [two-sided t-test, $p = 0.9$ for no MLL2 vs for MLL2 on one side; $p = 10^{-39}$ for MLL2 on one side vs for MLL2 on both sides; $p = 10^{-8}$ for no MLL2 vs for MLL2 on both sides] **(I)** Proportion of MLL2 binding sites that overlap with active enhancers, active promoters and bivalent promoters in WT naïve and primed PCs. Enhancers are defined by overlapping H3K4me1 and H3K27ac peaks, active promoters by overlapping H3K4me3 and H3K27ac peaks and bivalent promoters by overlapping H3K4me3 and H3K27me3 peaks. Peaks identified from wild-type cells [1]. **(J)** (Row 1) Schematic showing how categories of loops are defined as enhancer-promoter loops at active or bivalent genes, enhancer-enhancer loops and promoter-promoter loops at active or bivalent genes. (Row 2) Bar plots for categories in Row 1 showing proportion of loops gained (green) or lost (orange) when comparing WT and KO primed PCs but also when comparing WT and KO primed PCs that express the N2650A catalytically-dead mutant form of MLL2 (WT+NCat; KO+NCat). [Fisher's exact test with false discovery rate (FDR) multiple test correction, p-values: $10^{-26}/10^{-5}$, $10^{-74}/10^{-7}$, $10^{-27}/10^{-4}$, $10^{-23}/10^{-6}$ and $10^{-50}/10^{-3}$ (E-P, E-BP, E-E, P-P and BP-BP loops in WT vs KO/ WT+NCat vs KO+NCat)].

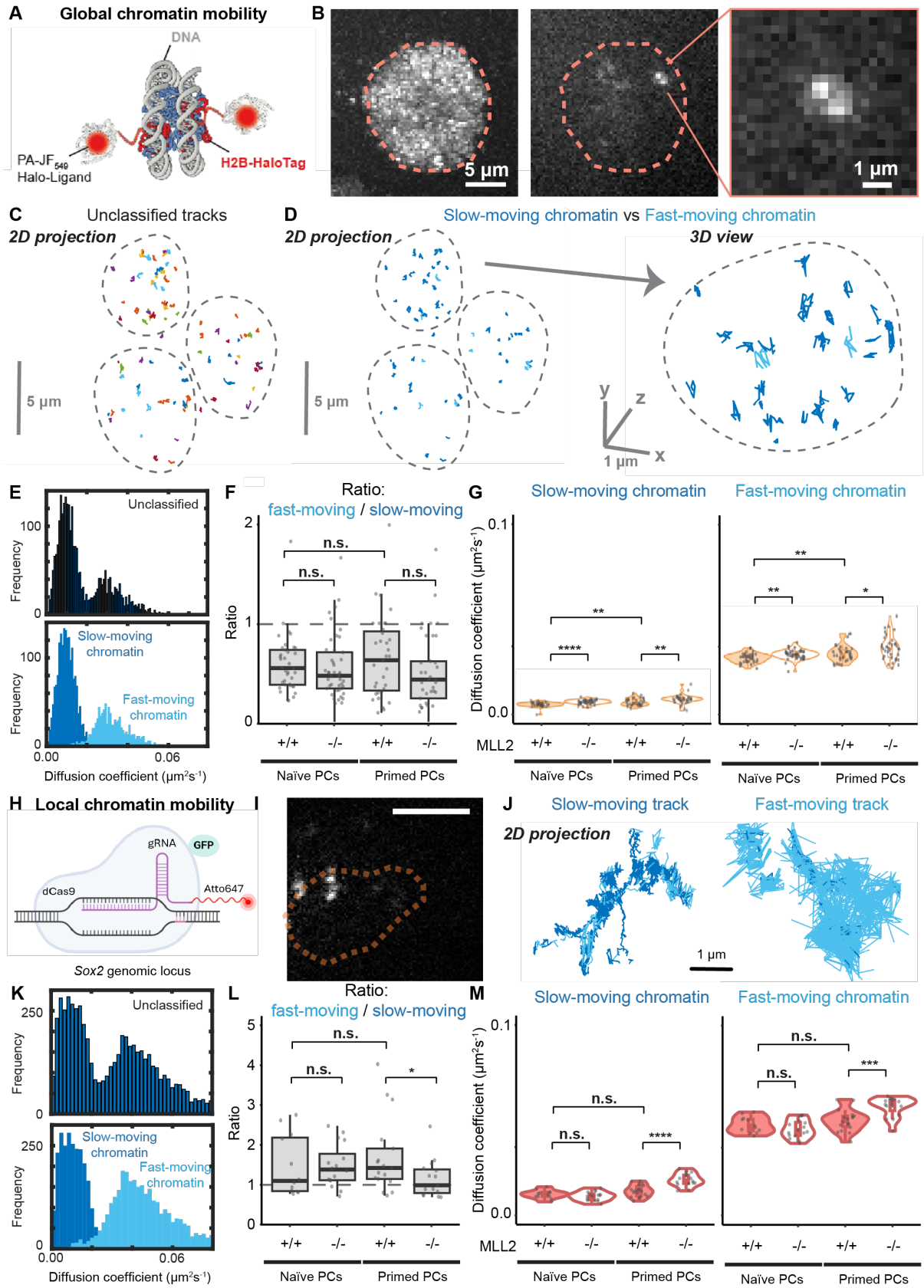


Figure 6. MLL2 restrains chromatin mobility genome-wide and at *Mll2*-dependent gene *Sox2* during priming of pluripotent cells. (A) Schematic showing PA-JF$_{549}$-HaloTag-tagged histone H2B. **(B)** (Left) Sum projection of single-molecule localisation microscopy (SMLM)

video shows nuclear localisation of histone H2B within a nucleus indicated by an orange dotted line. (Middle) Representative SMLM image showing single PA-JF_{549}-HaloTag-tagged histone H2B molecules in a live pluripotent cell (PC) imaged using a double helix point spread function (DHPSF) microscope. Each molecule appears as two dots where the midpoint encodes the x,y position and where the angle between the dots encodes the z position. The nucleus is indicated by an orange dotted line based on the sum projection (Right) Inset showing zoomed in image of a single histone H2B molecule. **(C)** Tracks pre-classification. **(D)** (Left) Tracks classified as slow- (dark blue) and fast-moving chromatin (light blue). (Right) Inset showing zoomed in image of a few classified trajectories. **(E)** Histogram showing apparent diffusion coefficient D_{app} of histone H2B trajectories (above) pre-classification and (below) post-classification into slow- or fast-moving chromatin. **(F)** Box-and-violin plot showing ratio of H2B trajectories classified as fast-moving/slow-moving for naïve and primed PCs in WT/KO conditions. Dots represent ratios calculated from independent videos (n>36 videos, 3-5 cells per video) [unpaired two-sided t-test, p-values: 0.8 (WT naïve vs primed PCs), 0.3 (WT vs KO in naïve PCs), 0.09 (WT vs KO in primed PCs)] **(G)** Box-and-violin plot showing apparent diffusion coefficient D_{app} of (left) slow- and (right) fast-moving chromatin for naïve and primed PCs in WT/KO conditions. Dots represent mean values calculated from independent videos (n>36 videos, 3-5 cells per video) [unpaired two-sided t-test, p-values: $**10^{-3}$ (slow-moving WT naïve vs primed PCs), $****10^{-6}$ (slow-moving WT vs KO in naïve PCs), $**10^{-3}$ (slow-moving WT vs KO in primed PCs), $**10^{-3}$ (fast-moving WT naïve vs primed PCs), $**10^{-3}$ (fast-moving WT vs KO in naïve PCs), *0.02 (fast-moving WT vs KO in primed PCs)] **(H)** Schematic showing dCas9-GFP targeted to genomic repeat near the Sox2 gene locus using a specific Atto647-tagged gRNA. **(I)** Representative SMLM image of three labelled Sox2 gene loci in a live PC imaged using a DHPSF microscope. The nucleus is indicated by an orange dotted line based on the bright field image. **(J)** Representative trajectories of Sox2 gene loci that are either mostly slow-moving or mostly fast-moving. **(K)** Histogram showing apparent diffusion coefficient D_{app} of *Sox2* gene sub-trajectories (above) pre-classification and (below) post-classification into slow- or fast-moving chromatin. **(L)** Box-and-violin plot showing ratio of *Sox2* gene sub-trajectories classified as fast-moving/slow-moving gene for naïve and primed PCs in WT/KO conditions. Dots represent ratios calculated from independent videos (n>12 videos, 3-5 cells per video) [unpaired two-sided t-test, p-values: 0.3 (WT naïve vs primed PCs), 1.0 (WT vs KO in naïve PCs), *0.03 (WT vs KO in primed PCs)] **(M)** Box-and-violin plot showing apparent diffusion coefficient D_{app} of (left) slow- and (right) fast-moving *Sox2* gene for naïve and primed PCs in WT/KO conditions. Dots represent mean values calculated from independent videos (n>12 videos, 3-5 cells per video) [unpaired two-sided t-test, p-values: 0.08 (slow-moving WT naïve vs primed PCs), 0.2 (slow-moving WT vs KO in naïve PCs),

****10^{-5} (slow-moving WT vs KO in primed PCs), 0.2 (fast-moving WT naïve vs primed PCs), 0.2 (fast-moving WT vs KO in naïve PCs), ***10^{-4} (fast-moving WT vs KO in primed PCs)].

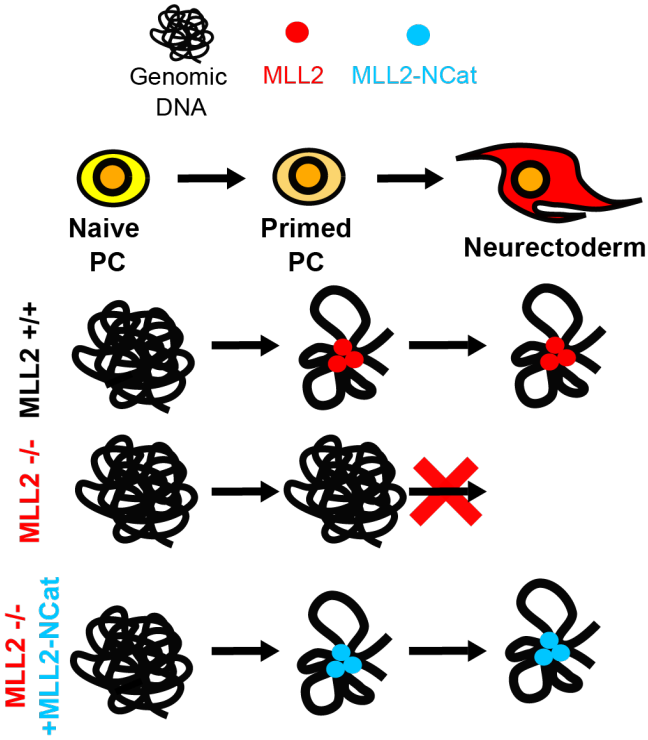


Figure 7. Model: MLL2 tethers enhancer-promoter loops during the priming of pluripotent cells to allow cells to differentiate into neuroectoderm. Schematic showing how in wild-type cells MLL2 increases tethering of enhancer-promoter loops at bivalent genes during priming of pluripotent cells to allow neuroectoderm differentiation. MLL2 cKO disrupts enhancer-promoter loops and increases formation of repressive loops during priming of pluripotent cells, blocking neuroectoderm differentiation. Introduction of a catalytically-dead MLL2 restores enhancer-promoter loops at bivalent genes and restores differentiation.

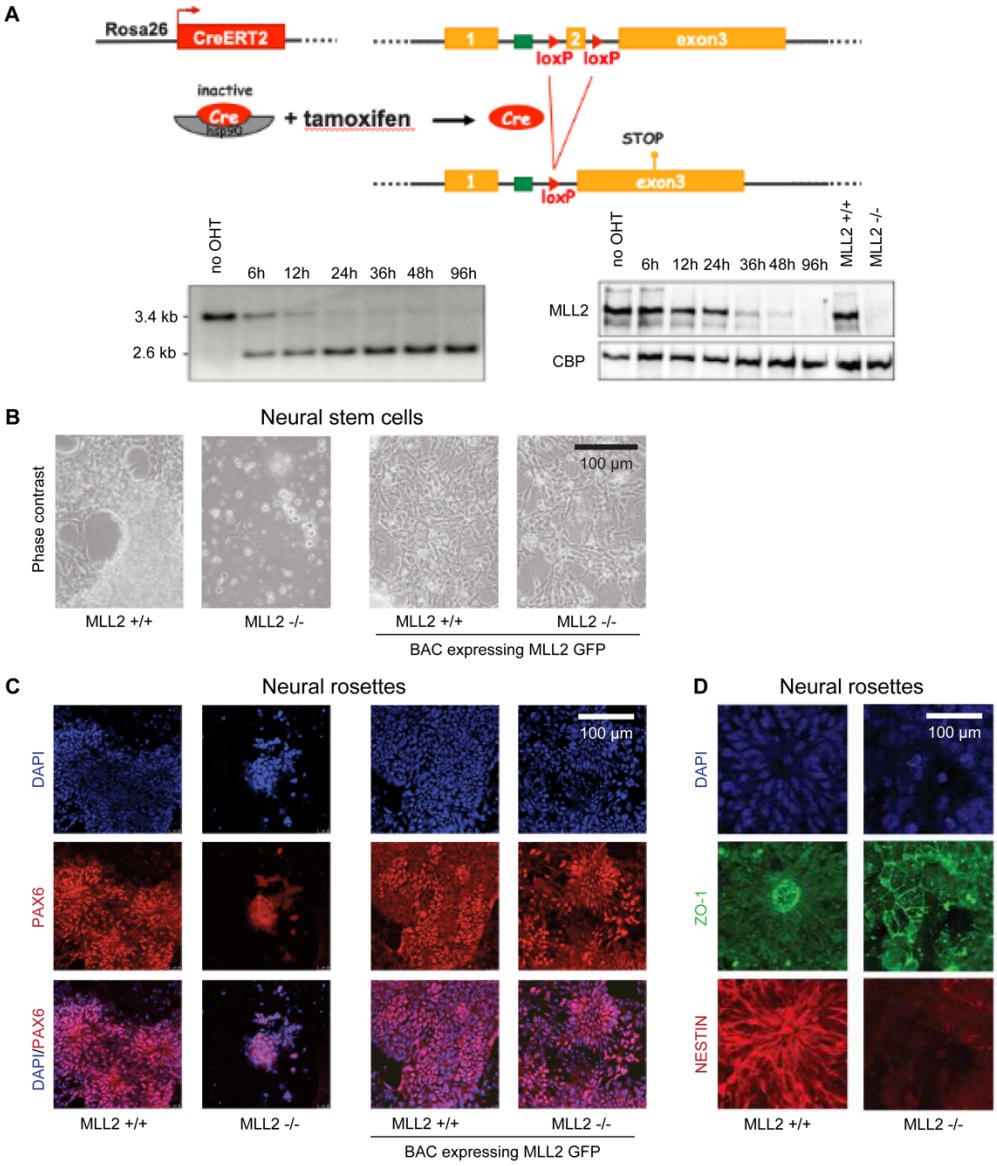


Fig. S1. MLL2 is required for priming neuroectoderm specification *in vitro*. (A) (Top) Schematic representation of the conditional knockout strategy. MLL2 cKO cell line has loxP sites either side of *Mll2* exon 2 and expresses CreERT2 from a Rosa26 promoter to allow tamoxifen-induced removal of *Mll2* exon 2. (Bottom left) Genomic PCR confirms removal of *Mll2* exon 2 within 24 hr. (Bottom right) Western blot confirms depletion of MLL2 protein within 48 hr. **(B)** Phase contrast images at the neural stem cell stage of control (MLL2 +/+) and MLL2 cKO (MLL2-/-) cells show that MLL2 cKO cells fail to generate neural stem cells. **(C)** DAPI stain (blue), PAX6 expression (red) and merge at the neural rosette stage in control (MLL2 +/+), tamoxifen-induced MLL2-deficient (MLL2 -/- cells) and MLL2 -deficient expressing exogenous GFP-tagged MLL2 from an integrated bacterial artificial chromosome (BAC). **(D)** DAPI stain (blue), ZO-1 expression (red) and NESTIN expression (red) at the neural rosette stage in control (MLL2 +/+), tamoxifen-induced MLL2-deficient (MLL2 -/- cells).

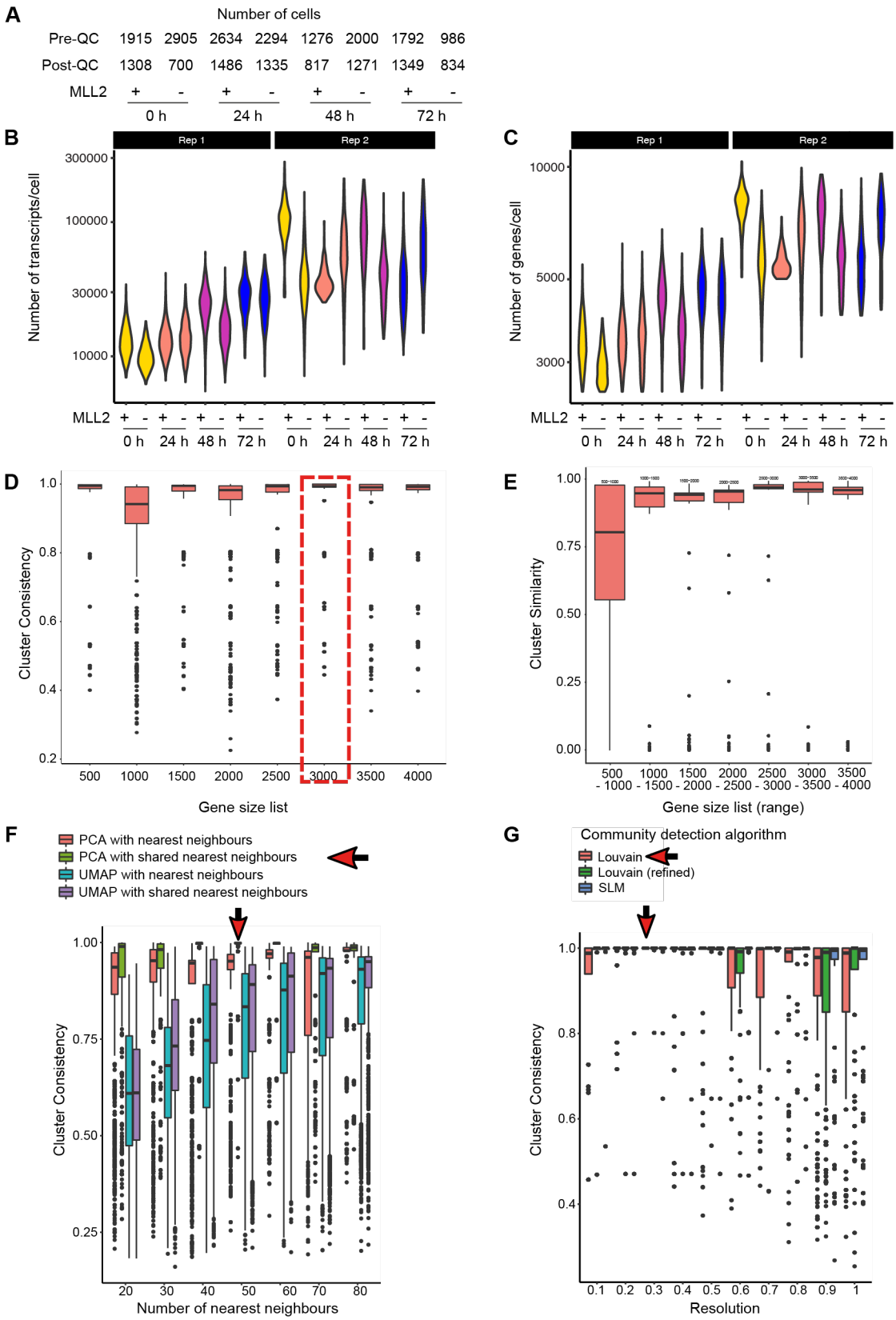


Fig. S2 Quality control of 10x scRNAseq datasets and cluster analysis. (A) Table containing number of cells per sample. **(B)** Violin plot showing number of transcripts/cell post filtering [see **(A)** for sample size]. **(C)** Violin plot of number of genes per cell from 10x scRNAseq post-filtering [see **(A)** for sample size]. **(D-E)** The effect of gene list size (of the

most variable genes), used in the PCA dimensionality reduction step, on **(D)** cluster consistency and **(E)** similarity (n = 100 iterations of randomly sampled genes). **(F)** The effect of different base embeddings (PCA vs UMAP), nearest neighbouring algorithm (standard or shared) and number of neighbours on clustering consistency. **(G)** The effect of different community detection algorithms on cluster consistency. **(D-G)** red box and arrow indicate the parameters selected for scRNAseq analysis. (n = 100 iterations of randomly sampled genes)

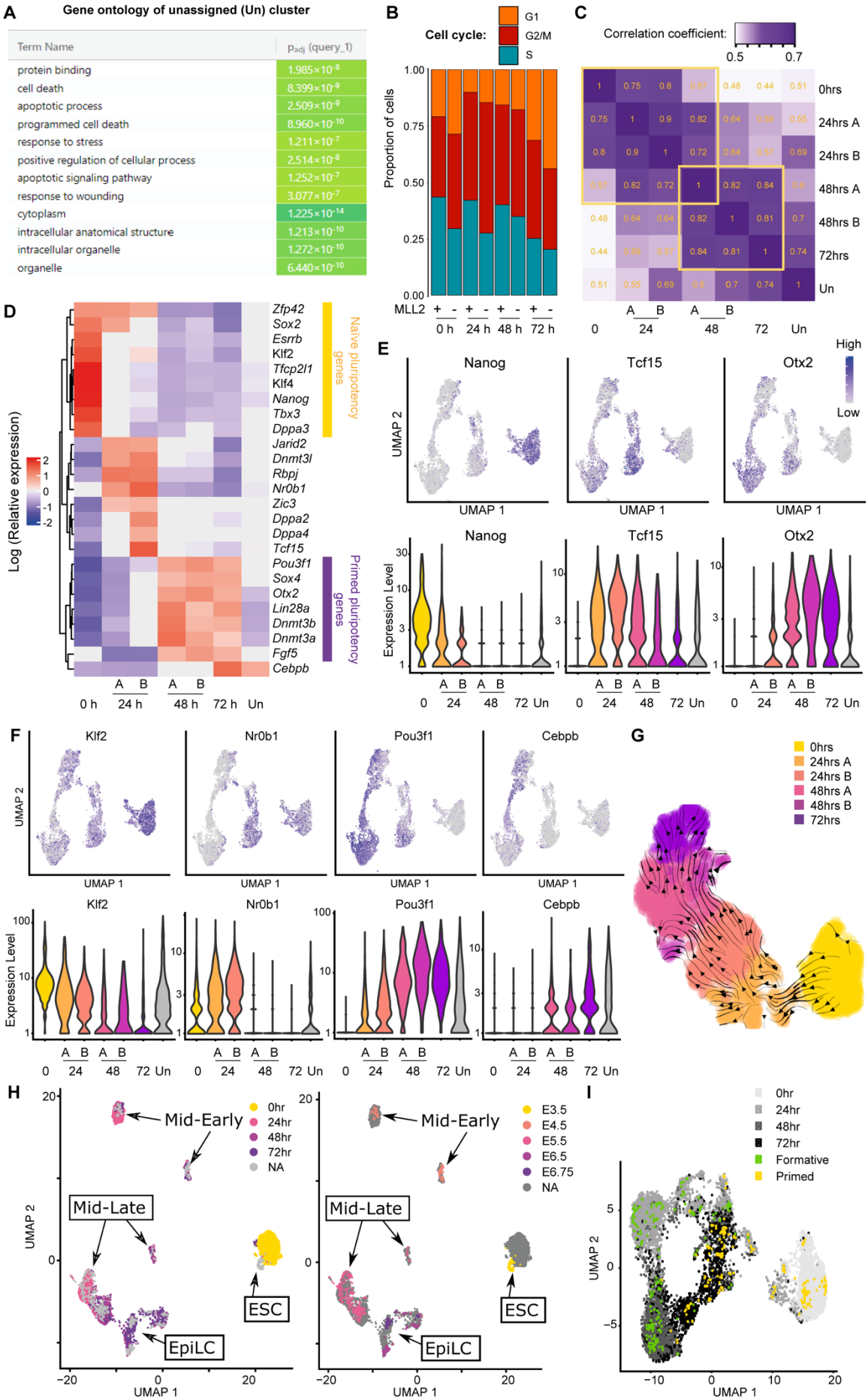

A
Gene ontology of unassigned (Un) cluster
Term Name
p_{adj} (query_1)
protein binding 1.985×10-8
cell death 8.399×10-9
apoptotic process 2.509×10-9
programmed cell death 8.960×10-10
response to stress 1.211×10-7
positive regulation of cellular process 2.514×10-8
apoptotic signaling pathway 1.252×10-7
response to wounding 3.077×10-7
cytoplasm 1.225×10-14
intracellular anatomical structure 1.213×10-10
intracellular organelle 1.272×10-10
organelle 6.440×10-10
B
Cell cycle:
G1
G2/M
S
Proportion of cells
MLL2
0 h
24 h
48 h
72 h
C
Correlation coefficient:
0.5
0.7
0hrs
24hrs A
24hrs B
48hrs A
48hrs B
72hrs
Un
D
Log (Relative expression)
Zfp42
Sox2
Esrrb
Klf2
Tfcp2l1
Klf4
Nanog
Tbx3
Dppa3
Jarid2
Dnmt3l
Rbpj
Nr0b1
Zic3
Dppa2
Dppa4
Tcf15
Pou3f1
Sox4
Otx2
Lin28a
Dnmt3b
Dnmt3a
Fgf5
Cebpb
Naïve pluripotency genes
Primed pluripotency genes
0 h
24 h
48 h
72 h
Un
E
Nanog
Tcf15
Otx2
High
Low
UMAP 1
UMAP 2
Expression Level
F
Klf2
Nr0b1
Pou3f1
Cebpb
G
0hrs
24hrs A
24hrs B
48hrs A
48hrs B
72hrs
H
Mid-Early
Mid-Late
EpiLC
ESC
0hr
24hr
48hr
72hr
NA
E3.5
E4.5
E5.5
E6.5
E6.75
I
0hr
24hr
48hr
72hr
Formative
Primed

Fig. S3. Cluster assignment and cluster-specific genes. (**A**) G profiler output for unassigned cluster shows a signature of apoptosis and programmed cell death. **(B)** Cell cycle analysis show significant changes between wild-type and *Mll2* cKO cells with more cells in G1 and fewer cells in S phase [see **(Fig. S2A)** for sample size]. **(C)** Spearman correlation of all genes assigned to the second highest expression quartile instead of only variable genes used for dimensionality reduction. **(D)** Heatmap of annotated gene set showing relative gene expression in assigned clusters. The gene set is based on previous publications of genes involved in pluripotent cell priming. **(E-F)** Single-cell expression of naïve, intermediate and primed genes from **Figure 2B** +/- *Mll2* for **(E)** well-known genes and **(F)** less well-known genes. Expression represented as (top) a UMAP plot coloured by transcript level (grey to blue for low to high expression) and (bottom) as violin plots of transcript level per assigned cluster. Number of cells: 1215/638 (0 hr +/- MLL2), 428/532 (24 hr A +/- MLL2), 998/671 (24 hr B +/- MLL2), 759/914 (48 hr A +/- MLL2), 133/223 (48 hr B +/- MLL2), 348/343 (72 hr +/- MLL2) and 716/819 (Unassigned). **(G)** UMAP plot regenerated from **Figure 2B** after removal of unassigned cluster and using parameters that better represent distance between cells rather than between clusters. Cells in UMAP arranged by pseudotime metric calculated from CellRank pipeline and with arrows indicating directionality. **(H)** UMAP plot combining our data (n = 8737 cells) with published mouse embryo *in vivo* datasets (n = 721 cells) [2]. **(I)** UMAP plot combining our data (n = 8737 cells) with published *in vitro* datasets of formative-like (n = 192 cells) and primed pluripotent cells (n = 192 cells) [3].

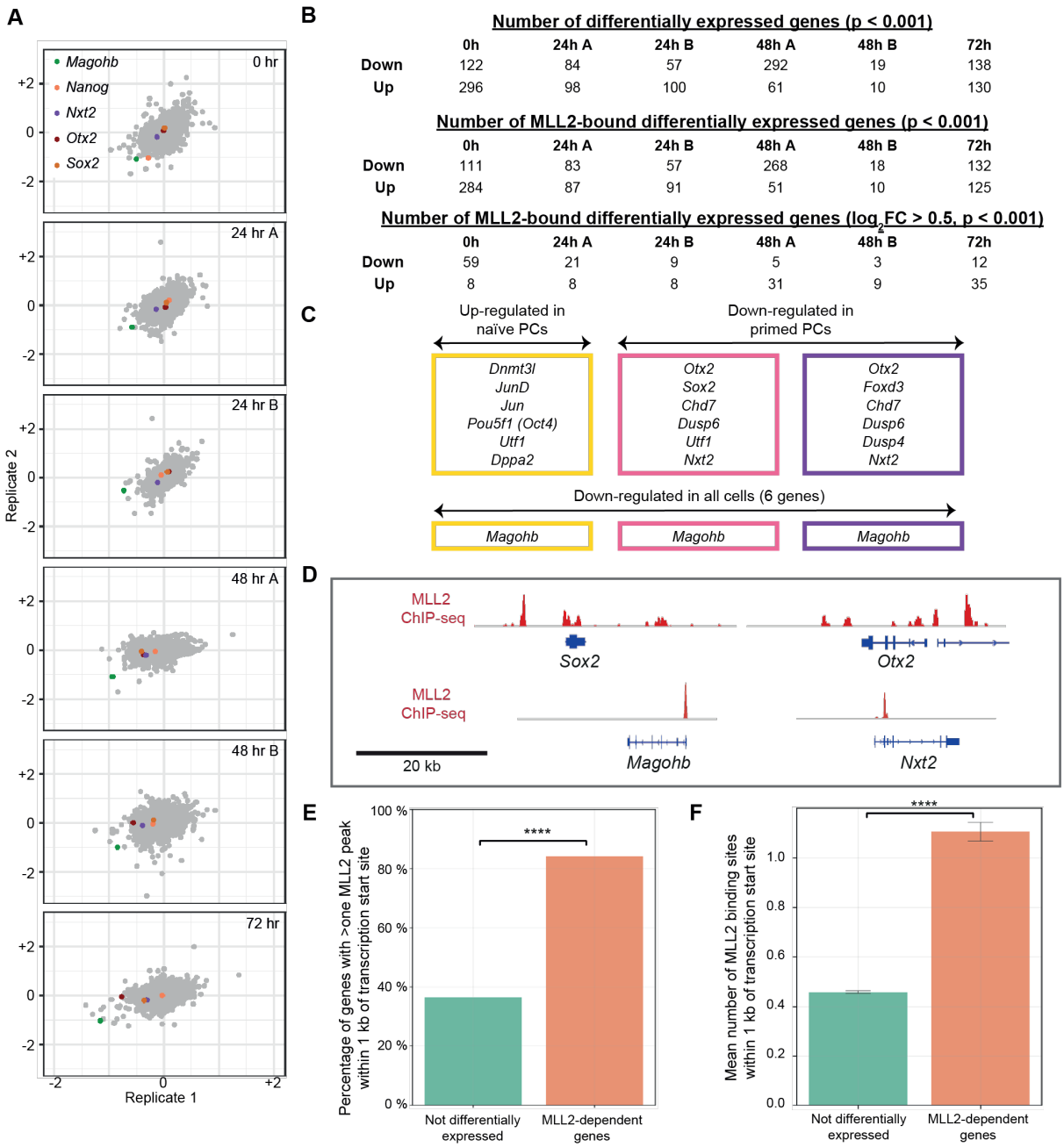


Figure S4. Identification of MLL2-dependent genes. (A) Comparison of scRNAseq replicates show reproducibility in $\log_2$FC for *Mll2*-dependent genes. **(B)** Table showing number of DEGs (top) and number of DEGs with MLL2 bound within 1 kb of the gene's transcription start site: $\log_2$FC > 0 (middle) or $\log_2$FC > 0.5 (bottom). **(C)** Table showing example DEGs up-regulated in naïve PCs and down-regulated in primed PCs. **(D)** UCSC genome browser tracks showing MLL2 ChIP-seq [4] for the MLL2-dependent genes *Sox2*, *Otx2*, *Magohb* and *Nxt2*. **(E)** Bar chart showing an increase in the proportion of genes bound by MLL2 at differentially expressed genes versus at those with no transcriptional change. [Chi-squared test: $p = 10^{-267}$] **(F)** Bar chart showing an increase in the mean number of MLL2 peaks at differentially expressed genes versus at those with no transcriptional change. [error bars represent 95 % confidence interval, Mann Whitney Test: $p = 10^{-273}$]

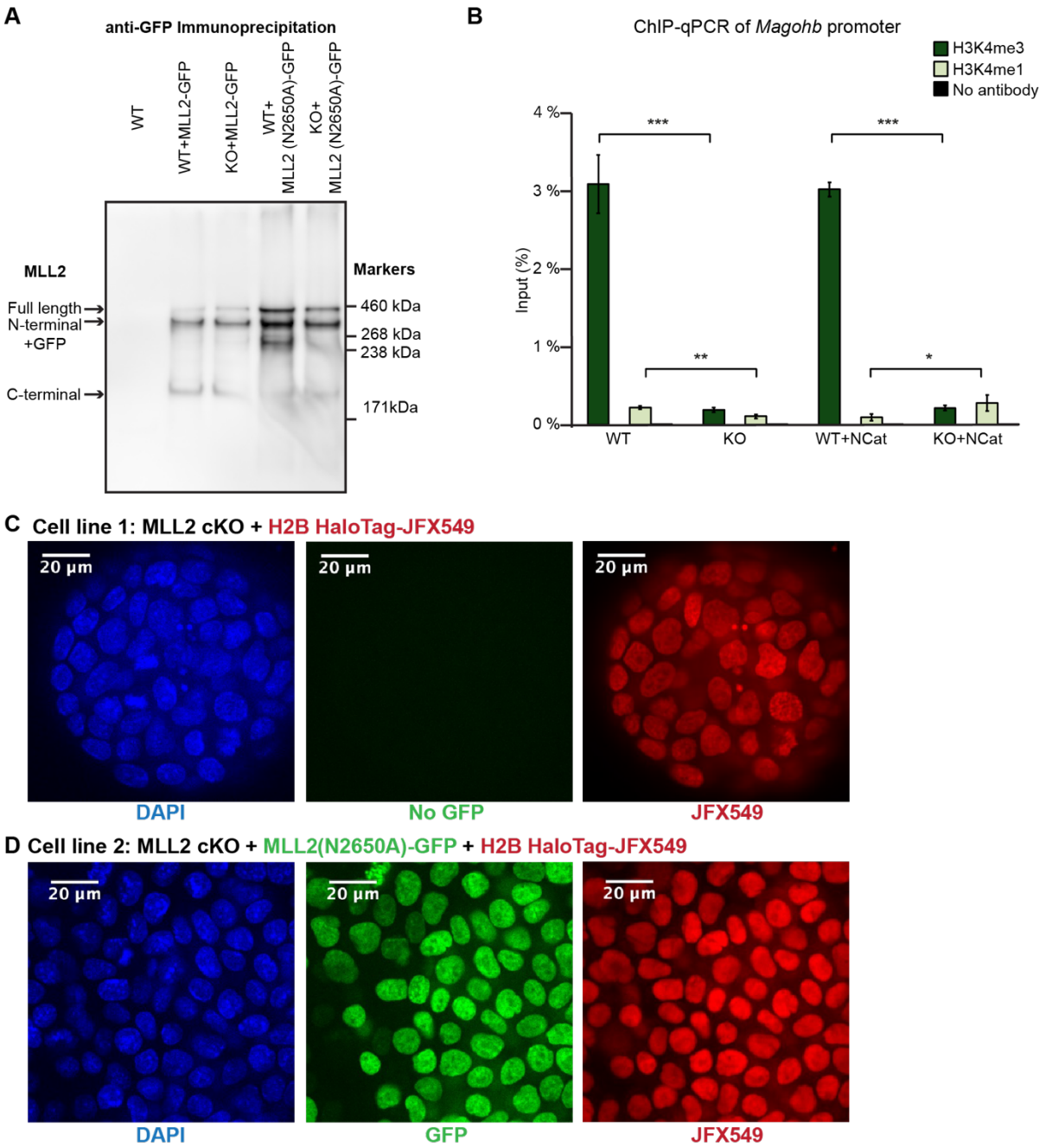


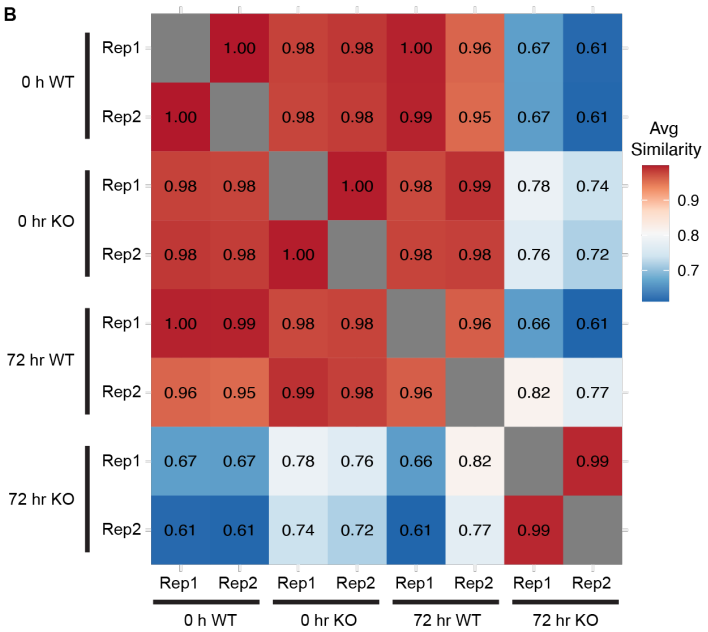
Figure S5. Generation of cell lines expressing catalytically-dead MLL2 and/or HaloTag-tagged histone H2B. (A) Immunoprecipitation using anti-GFP followed by SDS PAGE gel shows that GFP-tagged MLL2 proteins are expressed. **(B)** ChIP-qPCR at *Magohb* gene promoter for H3K4me3 and H3K4me1 [unpaired two-sided t-test, p-values: ***10^{-4} (H3K4me3 WT vs KO), **10^{-3} (H3K4me1 WT vs KO), ***10^{-4} (H3K4me3 WT + NCat vs KO + NCat), *0.04 (H3K4me1 WT + NCat vs KO + NCat)] **(C)** Representative images of MLL2 cKO cell line expressing HaloTag-tagged histone H2B, where histone H2B molecules were labelled using JFX_{549} dye (red) and imaged alongside DAPI (blue) to confirm nuclear localisation. **(D)** Representative images of MLL2 cKO cell line ectopically expressing a GFP-tagged catalytically-dead form of MLL2 and HaloTag-tagged histone H2B, where histone H2B

molecules were labelled using JFX_{549} dye (red) and imaged alongside DAPI (blue) and GFP-tagged MLL2 (green) to confirm expression of catalytically-dead form of MLL2 and nuclear localisation.

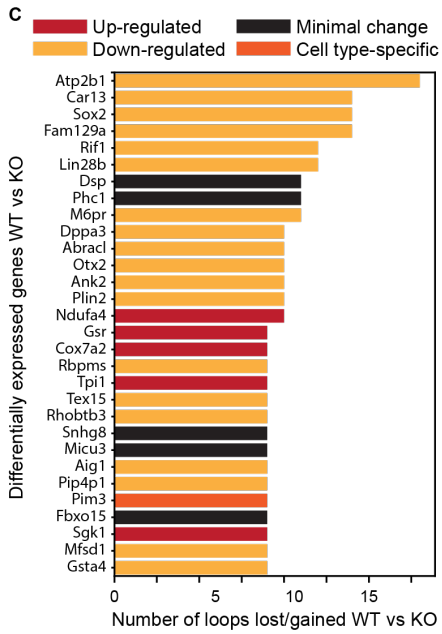
A

Sample	Replicate	Total Reads	Mapped Reads	Unique Reads	Total Unique Reads	Total cis Contacts >40 kb	A/B compartment number	Number of TADs	Number of loops
0h WT	1	607,912,217	460,358,376	170,728,864	277,041,238	52,141,265	3031/3010	2753	13,435
	2	375,949,065	288,543,274	106,312,374					
0h WT +N2650A	1	338,947,336	248,575,492	204,148,485	418,573,010	94,207,251	1758/1737	1737	13,159
	2	358,599,413	267,858,976	214,424,525					
0h KO	1	436,518,271	328,343,059	223,488,443	313,691,100	63,251,327	2487/2466	2547	11,654
	2	217,702,133	160,551,113	90,202,658					
0h KO +N2650A	1	277,365,909	204,817,093	169,416,895	336,299,030	83,977,034	2221/2200	2200	9667
	2	298,012,751	213,628,048	166,882,135					
72h WT	1	309,369,243	237,526,777	178,014,027	452,080,924	61,967,130	1854/1833	2834	13,380
	2	441,260,987	330,181,230	274,066,897					
72h WT +N2650A	1	316,021,981	236,515,098	187,507,100	391,826,635	76,830,470	2120/1833	2099	13,962
	2	348,244,578	264,855,898	204,319,535					
72h KO	1	323,615,588	241,263,984	199,328,071	419,563,124	84,751,843	1435/1414	2683	13,999
	2	371,760,813	270,964,123	220,235,053					
72h KO +N2650A	1	356,224,911	268,672,957	211,741,010	540,866,526	103,033,859	1864/1843	1843	14,008
	2	581,998,413	434,217,388	329,125,516					

B



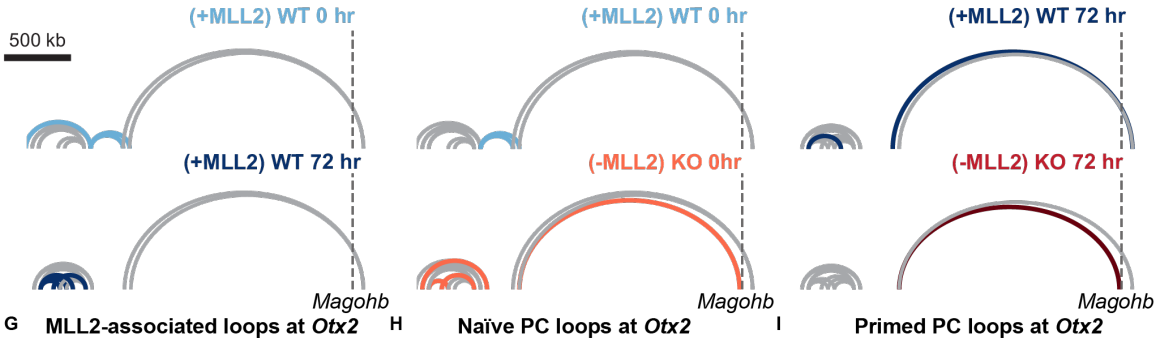
C



D MLL2-associated loops at *Magohb* in naïve/primed pluripotent cells

E Naïve PC loops at *Magohb* gained/lost in MLL2 cKO

F Primed PC loops at *Magohb* gained/lost in MLL2 cKO



G MLL2-associated loops at *Otx2* in naïve/primed pluripotent cells

H Naïve PC loops at *Otx2* gained/lost in MLL2 cKO

I Primed PC loops at *Otx2* gained/lost in MLL2 cKO

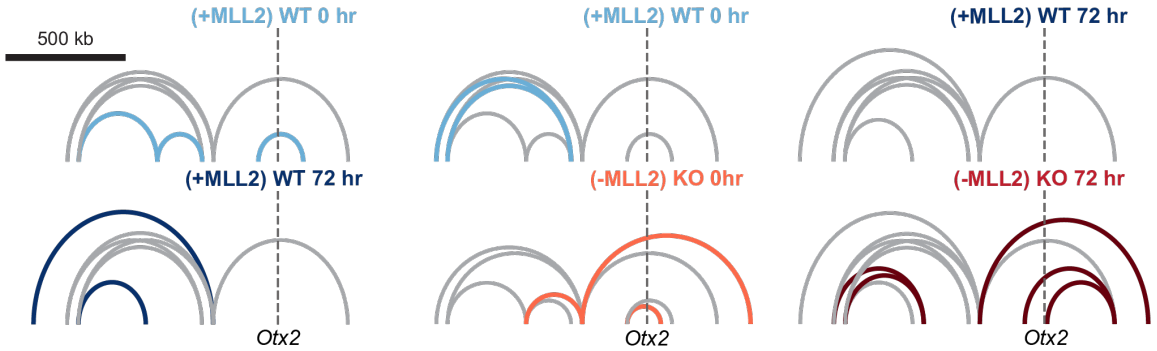


Figure S6. Quality control and example of high-confidence loops from Micro-C analysis. **(A)** Table showing number of reads, 3D looping contacts, compartments, TADs and significant loops detected by HICCUPS per sample ($p < 0.02$). **(B)** HICRep shows consistency between replicates when comparing contacts at a range of bin sizes (0.25, 0.5, 1 and 2.5 Mb). **(C)** Differentially expressed genes ranked by number of loops lost/gained when comparing WT vs KO conditions. **(D, G)** High-confidence MLL2-associated loops detected by HICCUPS ($p < 0.02$) at **(D)** *Magohb* and **(G)** *Otx2* genes in naïve (above) and primed PCs (below) for wild-type (WT) cells. Grey loops are common to both conditions, light blue to naïve PCs and dark blue to primed PCs ($p < 0.02$). Grey dotted line represents transcription start site of gene locus. **(E,H)** High-confidence MLL2-associated loops at **(E)** *Magohb* and **(H)** *Otx2* genes in naïve PCs for wild-type (above, WT) and MLL2 cKO cells (below, KO). Grey loops are common to both conditions, light blue to wild-type naïve PCs and light red to MLL2 cKO naïve PCs ($p < 0.02$). Grey dotted line represents transcription start site of gene locus. **(E, H)** High-confidence MLL2-associated loops at **(F)** *Magohb* and **(I)** *Otx2* genes in primed PCs for wild-type (above, WT) and MLL2 cKO cells (below, KO). Grey loops are common to both conditions, dark blue to wild-type primed PCs and dark red to MLL2 cKO primed PCs ($p < 0.02$). Grey dotted line represents position of gene locus.

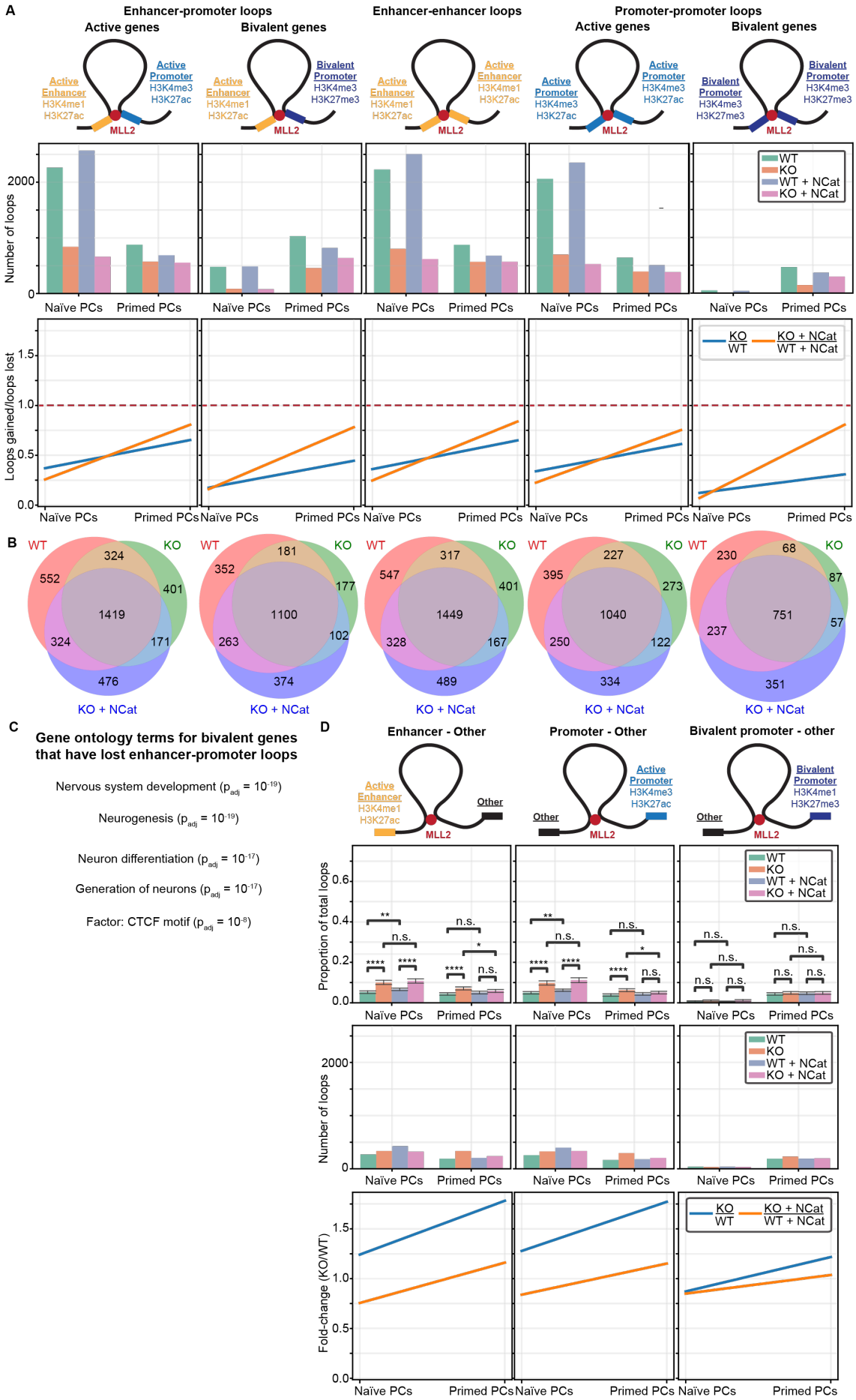

A
Enhancer-promoter loops
Active genes
Bivalent genes
Enhancer-enhancer loops
Promoter-promoter loops
Active genes
Bivalent genes
Active Enhancer H3K4me1 H3K27ac
Active Promoter H3K4me3 H3K27ac
Bivalent Promoter H3K4me3 H3K27me3
MLL2
Number of loops
WT
KO
WT + NCat
KO + NCat
Naïve PCs
Primed PCs
Loops gained/loops lost
KO / WT
KO + NCat / WT + NCat
B
WT
KO
KO + NCat
552
324
401
1419
324
171
476
352
181
177
1100
263
102
374
547
317
401
1449
328
167
489
395
227
273
1040
250
122
334
230
68
87
751
237
57
351
C
Gene ontology terms for bivalent genes that have lost enhancer-promoter loops
Nervous system development ($p_{adj} = 10^{-19}$)
Neurogenesis ($p_{adj} = 10^{-19}$)
Neuron differentiation ($p_{adj} = 10^{-17}$)
Generation of neurons ($p_{adj} = 10^{-17}$)
Factor: CTCF motif ($p_{adj} = 10^{-8}$)
D
Enhancer - Other
Promoter - Other
Bivalent promoter - other
Active Enhancer H3K4me1 H3K27ac
Other
Active Promoter H3K4me3 H3K27ac
Bivalent Promoter H3K4me1 H3K27me3
MLL2
Proportion of total loops
Number of loops
Fold-change (KO/WT)
n.s.
**
*
****

Figure S7. Enhancer-promoter analysis shows that MLL2 increases specific and decreases non-specific looping at enhancers and promoters. (A) (Row 1) Schematic showing how categories of loops are defined as enhancer-promoter loops at active genes; enhancer-promoter loops at bivalent genes; enhancer-enhancer loops; promoter-promoter loops at active genes; or promoter-promoter loops at bivalent genes. (Row 2) Barplot showing number of loops loops gained (green) or lost (orange) when comparing wild-type (WT) and MLL2 cKO (KO) conditions for both naïve and primed pluripotent cells (PCs) or the proportion of loops gained (blue) or lost (pink) when comparing WT and KO conditions for naïve and primed PCs that express the N2650A catalytic mutant form of MLL2 (WT + NCat; KO + NCat). (Row 3) Line graph showing fold-change of loops gained over loops lost when comparing KO to WT cells (orange) and when comparing (KO + NCat) to (WT + NCat) (blue). **(B)** Euler diagram showing overlap between WT, KO and (KO + NCat) conditions shows that loops recovered in (KO + NCat) conditions are not always the same as in WT cells. **(C)** GO analysis shows that bivalent genes with loop changes are associated with nervous system development. **(D)** (Row 1) Schematic showing how categories of non-specific loops are defined as enhancer-other loops and promoter-other loops at active or bivalent genes. (Row 2) Barplot showing proportion of non-specific loops gained (green) and lost (orange) in naïve and primed PCs when comparing WT and KO conditions [Fisher's exact test with false discovery rate (FDR) multiple test correction, p-values: $10^{-15}/10^{-8}$ (E-other loops in naïve/primed PCs), $10^{-16}/10^{-7}$ (P-other loops in naïve/primed PCs, 0.78/0.78 (BP-other loops in naïve/primed PCs)]. Barplot also shows loops gained (blue) and lost (pink) when comparing (WT + NCat) and (KO + NCat) conditions. [Fisher's exact test with false discovery rate (FDR) multiple test correction, p-values: $10^{-11}/0.15$ (E-other loops in naïve/primed PCs), $10^{-15}/0.19$ (P-other loops in naïve/primed PCs, 0.09/0.86 (BP-other loops in naïve/primed PCs)] (Row 3) Barplot showing number of non-specific loops gained (green) and lost (orange) in naïve and primed PCs when comparing WT and KO conditions or loops gained (blue) and lost (pink) when comparing (WT + NCat) and (KO + NCat) conditions. (Row 4) Line graph showing fold-change of loops gained over loops lost when comparing KO to WT cells (orange) and when comparing (KO + NCat) to (WT + NCat) (blue).

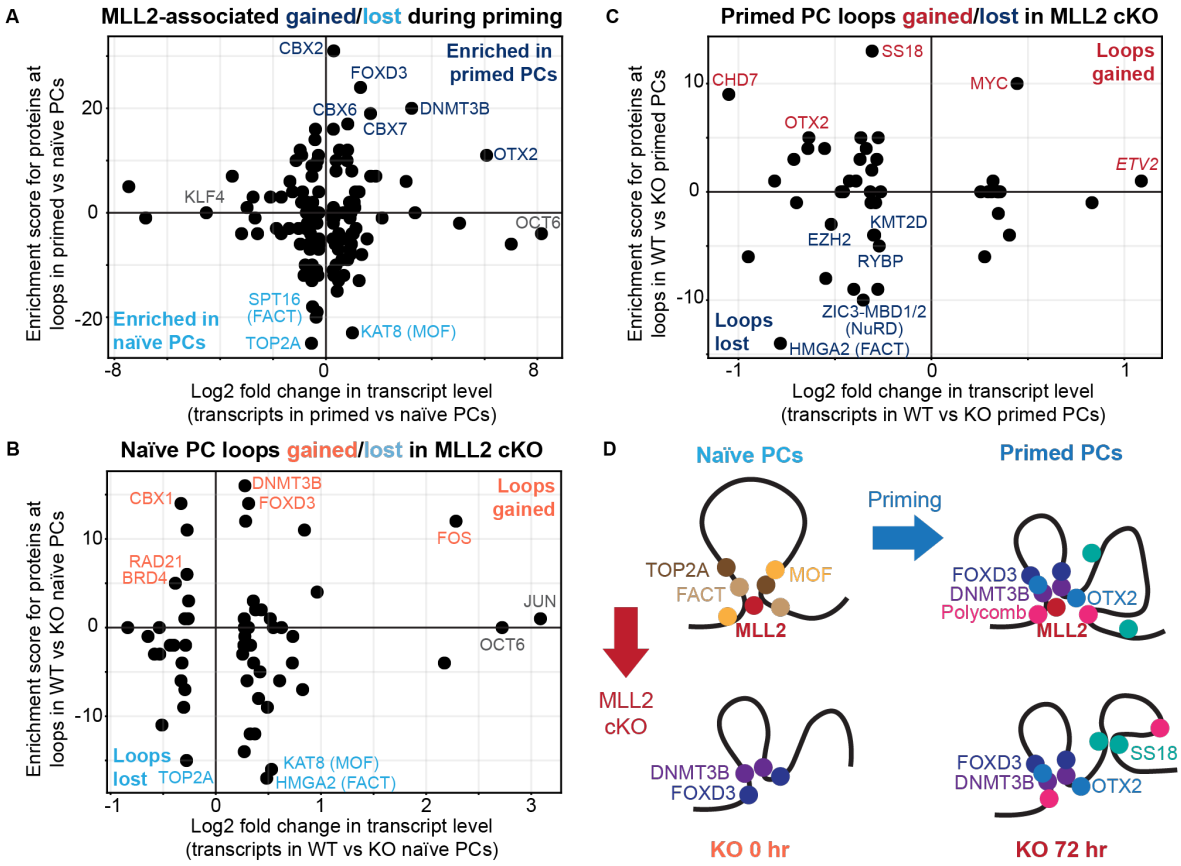


Figure S8. MLL2 enhances 3D chromatin loops linked to gene activation and inhibits loops linked to DNA methylation and heterochromatin. (A) Correlation plot showing how changes in protein enrichment at MLL2-associated loops relate to change in expression of enriched proteins when comparing naïve (bottom left) and primed PCs (top right). Proteins enriched in naïve and primed PCs are coloured light blue and dark blue respectively. **(B)** Correlation plot showing how changes in protein enrichment at MLL2-associated loops relate to change in expression of enriched proteins when comparing wild-type (bottom left) and MLL2 cKO (top right) naïve PCs (0 hr). Proteins enriched in wild-type and MLL2 cKO PCs are coloured light blue and light red respectively. **(C)** Correlation plot showing how changes in protein enrichment at MLL2-associated loops relate to change in expression of enriched proteins when comparing wild-type (bottom left) and MLL2 cKO (top right) primed PCs (72 hr). Proteins enriched in wild-type and MLL2 cKO PCs are coloured dark blue and dark red respectively. **(D)** Schematic showing changes in proteins at MLL2-associated loops during priming of pluripotent cells for wild-type and MLL2 cKO cells.

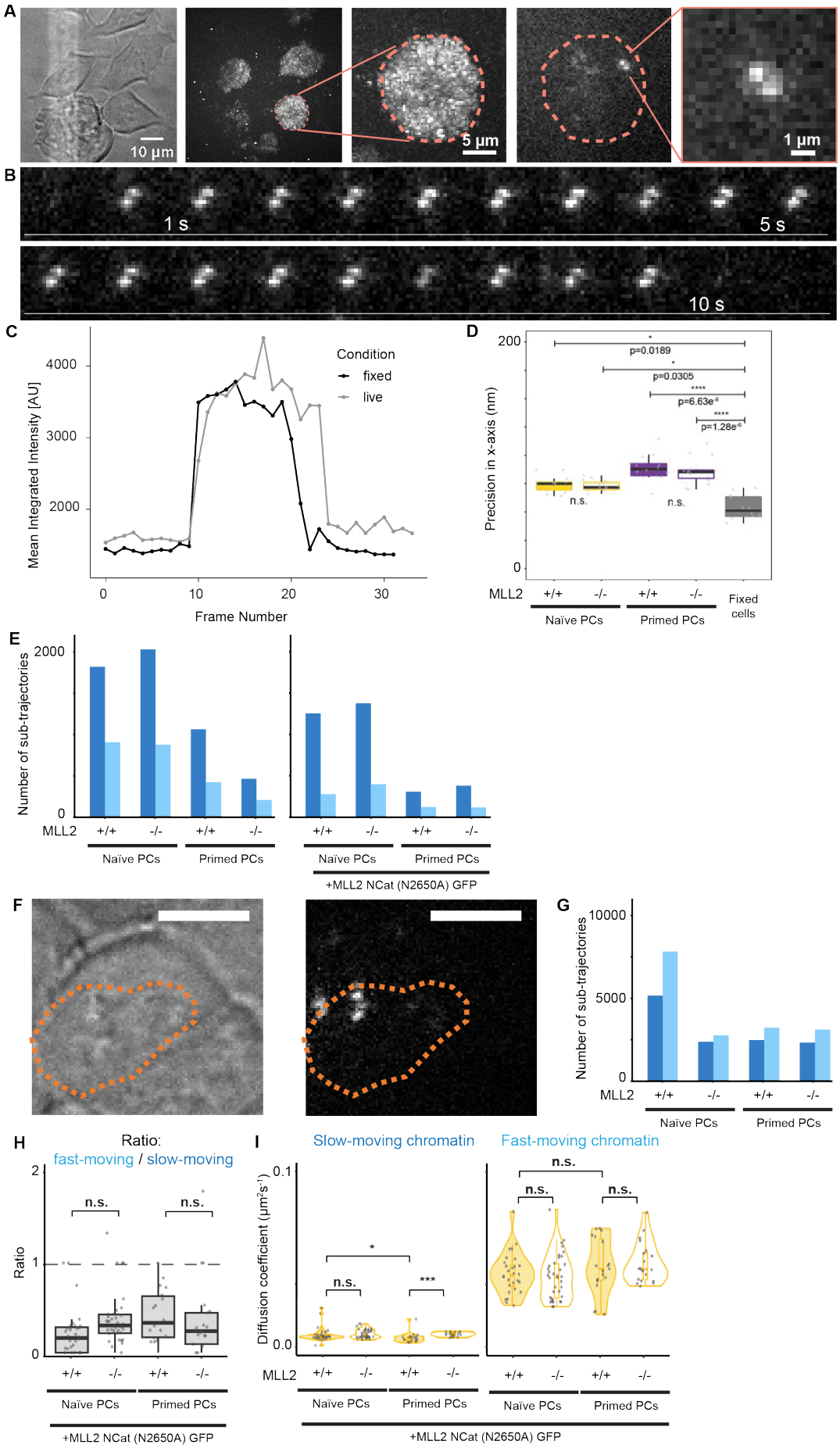

A
10 µm
5 µm
1 µm
B
1 s
5 s
10 s
C
Mean Integrated Intensity [AU]
Frame Number
Condition
fixed
live
D
Precision in x-axis (nm)
p=0.0189
p=0.0305
p=6.63e-6
p=1.28e-6
n.s.
MLL2
+/+
-/-
Naïve PCs
Primed PCs
Fixed cells
E
Number of sub-trajectories
+MLL2 NCat (N2650A) GFP
F
G
Number of sub-trajectories
H
Ratio:
fast-moving / slow-moving
Ratio
I
Slow-moving chromatin
Fast-moving chromatin
Diffusion coefficient (µm²s⁻¹)

Figure S9. Control experiments for SMLM. (A) (Left to right) Bright-field image of a field of view containing 3-4 cells. Next, we have a sum project of a single-molecule localisation microscopy (SMLM) video showing nuclear localisation of histone H2B within 3-4 nuclei. We then show a zoom of one of these nuclei with the nuclear periphery indicated by an orange dotted line. Then we show a representative SMLM image with a single PA-JF_{549}-HaloTag-tagged histone H2B molecule. We then show a zoom of the single molecule in the final panel. **(B)** Representative time-lapse montage showing single-step appearance and disappearance of PA-JF_{549}-HaloTag-tagged histone H2B molecule over 11 s (or 22 frames) taken from a live-cell SMLM video. **(C)** Intensity trace of histone H2B molecules labelled using PA-JF_{549} dye and tracked at 500 ms time resolution show single-step photobleaching. **(D)** Distance travelled by H2B between frames significantly higher in live versus fixed cells along the x, y and z directions [Kruskal-Wallis test and a post-hoc Dunn's test with Holm correction, p-values indicated on plots]. **(E)** Number of trajectories in H2B samples. **(F)** Representative bright-field and fluorescence image of cells labelled for imaging of *Sox2* gene using deactivated Cas9 and Atto647-labelled gRNAs. **(G)** Number of trajectories in *Sox2* gene tracking samples. **(H)** Proportion of H2B molecules observed in slow- and fast-moving chromatin states in WT/KO cells expressing N2650A catalytically-dead form of MLL2. Dots represent ratios calculated from independent videos (n>25 videos, 3-5 cells per video) [unpaired two-sided t-test, p-values: 0.14 (WT+NCat vs KO+NCat in naïve PCs), 1.0 (WT+NCat vs KO+NCat in primed PCs)] (**I**) Box-and-violin plot showing apparent diffusion coefficient D_{app} of (left) slow- and (right) fast-moving chromatin for naïve and primed PCs in WT/KO cells expressing N2650A catalytically-dead form of MLL2. [unpaired two-sided t-test, p-values: *0.02 (slow-moving WT+NCat naïve vs primed PCs), 0.2 (slow-moving WT+NCat vs KO+NCat in naïve PCs), ***10^{-4} (slow-moving WT+NCat vs KO+NCat in primed PCs), 1.0 (fast-moving WT+NCat naïve vs primed PCs), 1.0 (fast-moving WT+NCat vs KO+NCat in naïve PCs), 1.0 (fast-moving WT+NCat vs KO+NCat in primed PCs)]

References

1. Yang, S.H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A.D. (2014). Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. Cell Rep *7*, 1968-1981. 10.1016/j.celrep.2014.05.037.
2. Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep *20*, 1215-1228. 10.1016/j.celrep.2017.07.009.
3. Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J., and Smith, A. (2021). Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. Cell Stem Cell *28*, 453-471 e458. 10.1016/j.stem.2020.11.005.
4. Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H.G., and Stewart, A.F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Development *141*, 526-537. 10.1242/dev.102681.