MOF-associated complexes ensure stem cell identity and Xist repression

Tomasz Chelmicki, Friederike Dündar, Matthew Turley, Tasneem Khanam, Tugce Aktas, Fidel Ramírez, Anne-Valerie Gendrel, Patrick R Wright, Pavankumar Videm, Rolf Backofen, Edith Heard, Thomas Manke, Asifa Akhtar

DOI: http://dx.doi.org/10.7554/eLife.02024

Cite as: eLife 2014;10.7554/eLife.02024

Received: 6 December 2013
Accepted: 11 May 2014
Published: 19 May 2014

MOF-associated complexes ensure stem cell identity and *Xist* repression

Tomasz Chelmicki[1,2]*, Friederike Dündar[1,2,3]*, Matthew James Turley[1,2]^, Tasneem Khanam[1]^, Tugce Aktas[1]^, Fidel Ramírez[3], Anne-Valerie Gendrel[4], Patrick Rudolf Wright[5], Pavankumar Videm[5], Rolf Backofen[5,6,7,8], Edith Heard[4], Thomas Manke[3] and Asifa Akhtar[1]#

[1] Max Planck Institute of Immunobiology and Epigenetics, Department of Chromatin Regulation, 79108 Freiburg im Breisgau, Germany

[2] University of Freiburg, Faculty of Biology, 79104 Freiburg im Breisgau, Germany

[3] Max Planck Institute of Immunobiology and Epigenetics, Bioinformatics Department, 79108 Freiburg im Breisgau, Germany

[4] Mammalian Developmental Epigenetics Group, Institut Curie, 75248 Paris Cedex 05, France

[5] Bioinformatics Group, Department of Computer Science, University of Freiburg, D-79110 Freiburg, Germany

[6] BIOSS Center for Biological Signalling Studies, University of Freiburg, D-79104 Freiburg, Germany

[7] Center for Biological Systems Analysis, University of Freiburg, D-79104 Freiburg, Germany

[8] Center for Non-Coding RNA in Technology and Health, University of Copenhagen, DK-1870 Frederiksberg C, Denmark

* Joint first co-authors

^ These authors contributed equally to this work.

Corresponding author

akhtar@ie-freiburg.mpg.de
Phone: +49 (0)7615108565
Fax: +49 (0)7615108566

Summary

Histone acetyl transferases (HATs) play distinct roles in many cellular processes and are frequently misregulated in cancers. Here, we study the regulatory potential of MYST1-(MOF)-containing MSL and NSL complexes in mouse embryonic stem cells (ESCs) and neuronal progenitors. We find that both complexes influence transcription by targeting promoters as well as TSS-distal enhancers. In contrast to flies, the MSL complex is not exclusively enriched on the X chromosome yet it is crucial for mammalian X chromosome regulation as it specifically regulates *Tsix*, the major repressor of Xist lncRNA. MSL depletion leads to decreased *Tsix* expression, reduced REX1 recruitment, and consequently, enhanced accumulation of *Xist* and variable numbers of inactivated X chromosomes during early differentiation. The NSL complex provides additional, *Tsix*-independent repression of *Xist* by maintaining pluripotency. MSL and NSL complexes therefore act synergistically by using distinct pathways to ensure a fail-safe mechanism for the repression of X inactivation in ESCs.

Introduction

Histone acetyl transferases (HATs) are among the key architects of the cellular epigenetic landscape as the acetylation of histones is unanimously associated with transcriptionally active domains. Many HATs also have the ability to acetylate non-histone proteins extending their influence to diverse cellular pathways inside and outside of the nucleus (reviewed in (Sapountzi and Cote, 2011). Based on their catalytic domains, the HATs are classified into two major families, GCN5 N-acetyl transferases (GNATs) and MYST HATs (named after the founding members MOZ, Ybf2/Sas3, Sas2, Tip60), that encompass diverse sets of protein complexes. The individual complex members enhance and modulate the enzymes' activities, guiding the versatile HATs towards specific functions. GCN5, for example, is part of SAGA, ATAC, and SLIK complexes that are associated with distinct histone tail modifications and differential gene regulation (reviewed in (Lee and Workman, 2007, Nagy et al., 2010). In contrast, one of the well-known members of the MYST family, MOF (also known as: KAT8, MYST1), is rather substrate-specific for lysine 16 of histone H4 (H4K16) (Akhtar and Becker, 2000) and its interaction partners are thought to mainly alter the specificity and extent of MOF's H4K16 acetylation (H4K16ac). As part of the male-specific lethal (Koolen et al.) complex (MSL1, MSL2, MSL3, MOF, MLE, roX1 and roX2 lncRNAs) in *Drosophila*

63 *melanogaster*, MOF is recruited to the single X chromosome of male flies. The subsequent
64 spreading of H4K16 acetylation results in transcriptional upregulation of the male X
65 chromosome, the major means of *D. melanogaster* dosage compensation (reviewed in
66 (Conrad and Akhtar, 2011). In addition to the highly specialized MSL-associated role, MOF is
67 also involved in the more universal and sex-independent regulation of housekeeping genes
68 within the non-specific lethal (Denslow and Wade) complex (NSL1, NSL2, NSL3, MBD-R2,
69 MCRS2, MOF, WDS) (Mendjan et al., 2006, Raja et al., 2010, Lam et al., 2012, Feller et al.,
70 2012).

71 MOF and most of its interaction partners are conserved in mammals where MOF is also
72 responsible for the majority of H4K16 acetylation (Smith et al., 2005, Taipale et al., 2005).
73 MOF is essential for mammalian embryonic development and unlike the male-specific
74 lethality in *Drosophila*, deletion of *Mof* in mice is lethal for both sexes (Gupta et al., 2008,
75 Thomas et al., 2008). More specifically, mammalian MOF is critical for physiological nuclear
76 architecture (Thomas et al., 2008), DNA damage repair (Gupta et al., 2008), maintenance of
77 stem cell pluripotency (Li et al., 2012), differentiation of T cells (Gupta et al., 2013) and
78 survival of post-mitotic Purkinje cells (Kumar et al., 2011). Compared to MOF, mammalian
79 MSL and NSL complex members are poorly understood. Nevertheless, the individual complex
80 members appear to have important functions *in vivo* as mutations of the NSL complex
81 member KANSL1 cause the core phenotype of the 17q21.31 microdeletion syndrome (Zollino
82 et al., 2012, Koolen et al., 2012) and are common amongst patients with both Down
83 syndrome and myeloid leukaemia (Yoshida et al., 2013). Another NSL-associated protein,
84 PHF20 has been shown to associate with methylated Lys370 and Lys382 of p53 (Cui et al.,
85 2012) and to be required for somatic cell reprogramming (Zhao et al., 2013a). WDR5 was
86 shown to be an essential regulator of the core transcription network in embryonic stem cells
87 (Ang et al., 2011). The mammalian counterpart of *Drosophila* MSL2 was shown to have the
88 capacity to ubiquitylate p53 (Kruse and Gu, 2009) and lysine 34 of histone 2B (Wu et al.,
89 2011).

90 In the study presented here, we set out to dissect the mammalian MOF functions within the
91 MSL and NSL complexes using genome-wide chromatin immunoprecipitation and
92 transcriptome profiles and biochemical experiments for the core members of MSL and NSL
93 complexes in mouse embryonic stem cells (ESCs) and neuronal progenitor cells (NPCs). We

found that the MSL and NSL members possess concurrent as well as independent functions and that effects generally attributed to MOF are frequently accompanied by the NSL complex. The NSL complex abundantly binds to promoters of broadly expressed genes in ESCs and NPCs. These genes are predominantly downregulated upon depletion of either MOF or KANSL3. In contrast, the MSL complex shows more restricted binding in ESCs, which expands after differentiation, particularly at NPC-specific genes. In addition to promoter-proximal binding we discover several thousand binding sites of KANSL3 and MSL2 at promoter-distal loci with enhancer-specific epigenetic signatures. The majority of these distal regulatory sites are bound in ESCs, but not in differentiated cells, and genes that are predicted to be targeted by TSS-distal binding of MSL2 are frequently downregulated in sh*Msl2*-treated cells. The distinct, yet synergistic actions of both complexes become very apparent at the X inactivation center (XIC) that encodes numerous non-coding RNAs involved in the silencing of one of the two X chromosomes in differentiating female cells. We show that the MSL but not the NSL complex directly promotes expression of *Tsix*, the inverse transcript and the key murine repressor of *Xist* during early differentiation. Depletion of MSL proteins results in attenuation of *Tsix* transcription, enhanced *Xist* RNA accumulation and "chaotic" inactivation of variable numbers of X chromosomes during early differentiation. In addition to the very specific effect of MSL1/MSL2-depletion on the XIC genes, we show that MOF together with the NSL complex also influences *Xist* levels, but instead of affecting *Tsix*, MOF and KANSL3 depletion diminish key pluripotency factors involved in repressing *Xist*. Our study provides novel insights into the intricate interplay between MSL and NSL complexes in orchestrating gene expression. Furthermore we demonstrate how MSLs and NSLs ensure the active state of two X chromosomes in mouse embryonic stem cells via distinct mechanisms.

Results

MOF and its complexes show distinct chromatin binding dynamics during differentiation

To examine the behavior of MSL and NSL proteins in a cell type specific manner, we derived homogeneous populations of multipotent neuronal progenitor cells (NPCs) from mouse embryonic stem cells (ESCs) (Conti et al., 2005, Gendrel et al., 2014, Splinter et al., 2011). We followed the progress of the differentiation process by monitoring cell morphology (Figure

1A), as well as protein (Figure 1B) and transcript levels of ESC- and NPC-specific markers (Figure 1—figure supplement 1A-C). To gain a better understanding of how MOF-associated complexes behave throughout the differentiation process, in parallel to cell type-specific markers, we also monitored the RNA and protein levels of MOF, MSL (MSL1, MSL2) and NSL (KANSL1, KANSL3, MCRS1) complex members (Figure 1B and Figure 1—figure supplement 1A). Interestingly, MSL and NSL complex members showed distinct RNA and protein dynamics during the process of differentiation: KANSL1 and KANSL3 protein levels remained unchanged, whereas MSL1, MSL2 and MOF became more abundant in NPCs accompanied by increased H4K16 acetylation (H4K16ac) (Figure 1B). These results were confirmed using another ES cell-line and its NPC derivative (Figure 1—figure supplement 1D). The specificities of the antibodies were confirmed by co-immunoprecipitation assays (Figure 1—figure supplement 2A-C) as well as shRNA-mediated knockdowns followed by western blot analyses (Figures 3G and 6C).

To assess the distinct behaviors of the complexes in more detail, we generated genome-wide chromatin binding profiles for MSL1, MSL2 (MSL complex), KANSL3, MCRS1 (NSL complex) and MOF (MSL and NSL). ChIP-seq experiments in ESCs and NPCs yielded large numbers of high-quality DNA sequence reads and excellent agreements between the biological replicates (Figure 2—figure supplement 1A, Supplementary file 1A). Using MACS for peak calling (Zhang et al., 2008) and additional stringent filtering (see Methods and Materials), we scored between 1,500 and 15,000 regions of significant enrichments for the different proteins (Supplementary file 1B).

To uncover patterns of co-occurrence and independent binding, we used unsupervised clustering on the input-normalized signals. This unbiased approach allowed us to determine five main groups of binding distinguished by different combinations of the proteins and cell-type specific dynamics. As shown in Figure 2, three large clusters of binding sites encompassed regions where at least 1 of the investigated proteins was present both in ESCs as well as NPCs (clusters A, B and C). The binding sites of clusters A and B predominantly overlapped with annotated transcription start sites (TSS) in contrast to the regions that were bound exclusively in ESCs, which tended to contain inter- and intragenic regions (clusters D and E, Figure 2). The width of the enrichments did not differ profoundly between the groups (cluster E: 836 bp median width, cluster A: 1,782 bp median width). We found surprisingly

few regions where MOF associated primarily with MSL complex members. Instead, approximately 80 % of all MOF peaks displayed strong KANSL3 and MCRS1 signals (cluster B, see Figure 2 and Figure 2—figure supplement 1B), suggesting a predominant role of the NSL complex among MOF-associated complexes and a more specific role for the MSL complex at subsets of promoters and numerous intergenic and intronic regions. As the different clusters showed distinct enrichment patterns and diverse genomic localization, we set out to analyze the individual groups of binding in more detail.

The MSL and NSL complexes co-occur on active promoters of constitutively expressed genes in ESCs and NPCs

We first focused on the characterization of target promoters as the majority of MOF-binding was found around the TSS (mostly clusters A and B in Figure 2, Figure 3A). We identified 8,947 TSSs overlapping with ChIP-seq peaks of KANSL3 and/or MCRS1 in ESCs that encompassed virtually all MOF- and MSL-bound TSSs (Figure 3B). This pattern did not change substantially in NPCs where TSSs overlapping with MOF peaks almost always (99 %) showed significant enrichments of KANSL3 and in 35 % of the cases additionally contained a peak of MSL2 (Figure 3B, middle panel). Genes that were TSS-bound in ESCs tended to be bound in NPCs as well (Figure 3B, middle panel and Figure 3—figure supplement 1A). We next generated RNA-seq data for ESCs and NPCs, determined genes that were expressed in both cell types (FPKM > 4) and found that all ChIPed proteins preferably bound to the promoters of active genes (Figure 3C). Interestingly, in ESCs, genes whose TSSs were bound by members of both complexes showed higher median expression values than genes bound by only one complex (Figure 4—figure supplement 1B). In contrast to the differing expression values, analysis of gene ontology (GO) using DAVID (Huang da et al., 2009) revealed basic housekeeping functions for both gene groups, regardless of whether they were bound by the NSL complex only or by both MOF-complexes together (Figure—figure supplement 1C). Consistently, the promoters of all target gene groups were enriched for motifs associated with broad, non-cell-type-specific expression such as ELK1, YY1, CREB and E2F (Xie et al., 2005, Farre et al., 2007) and showed profound enrichments of CpG islands Figure 3—figure supplement 1D) which is indicative of housekeeping genes (Landolin et al., 2010). Interestingly, when we analyzed the subset of genes that gained binding of either KANSL3 or MSL2 in NPCs, we found strong enrichments of GO terms related to embryonic development for KANSL3 targets and cell migration and neuronal development for MSL2 targets.

The TSS-binding of the mouse NSL complex resembles that of the NSL complex in *D. melanogster*

MOF has traditionally been associated with a widespread enrichment along male X-linked genes in flies that is dependent on the MSL proteins (Figure 3D and Figure 3—figure supplement 2A). In our mammalian profiles, despite the presence of the MSLs, we could neither detect X-specific enrichments of MOF, nor broad domains of binding along gene bodies. Furthermore, promoter-distal binding sites consisted of narrow peaks and no evidence of spreading from intronic or intergenic regions was observed (Figures 2, 3A and 3D).

We then examined whether there was a correlation between NSL complex binding in *D. melanogaster* and mouse cells. Indeed, we found that mouse genes that were homologous to NSL complex targets in *D. melanogaster* had a high probability of being bound by the murine NSL complex as well (Pearson's Chi squared test of independence between NSL binding in the fly and the mouse, p-value < 2.2e-16). We additionally observed that mouse genes expressed in ESCs and NPCs, whose fly homologues were NSL targets, showed stronger signals for H3K4me3, MOF, KANSL3 and MCRS1 (but not for MSL1 or MSL2) than the mouse homologues of non-NSL-bound *D. melanogaster* genes (Figure 3—figure supplement 2B; lists of NSL-bound and –non-bound fly genes were from Lam et al., 2012). These findings support the notion that the function in housekeeping gene regulation by the *D. melanogaster* NSL complex is evolutionary conserved.

Depletion of MSL and NSL complex members results in genome-wide downregulation of TSS-target genes

To dissect the biological consequences of the gene targeting by the different MSL and NSL proteins in ESCs, we systematically depleted core members of both complexes (MOF, KANSL3, MSL1, MSL2) (Figure 3—figure supplement 3A). Interestingly, MOF- or KANSL3-depleted cells showed more severe proliferation defects than MSL1- and MSL2-depleted cells (Figure 3—figure supplement 3B). We subsequently performed RNA-seq experiments from shRNA-treated cells and determined their differential expression against the scrambled control to dissect transcriptional outcomes of the depletions at a global level. We found a striking overlap between the differential expression of MSL1 and MSL2 knockdowns and a higher resemblance of MOF-dependent differential expression to that of KANSL3-depletion

(Figure 3E). When we specifically focused on genes that we had identified as TSS-bound in our ChIP-seq samples, we found that their transcripts tended to be downregulated in all four knockdowns in comparison to untargeted genes which showed higher fractions of upregulation. These effects were independent of the wild-type expression status of the gene or the chromosome (Figure 3F and Figure 3—figure supplement 3C).

TSS-binding of MSL1 and KANSL3 does not require MOF

Turning to the assessment of protein levels in shRNA-treated cells, we detected markedly reduced bulk H4K16 acetylation in MSL1- and MOF-depleted cells and only slight reduction upon KANSL3-depletion. This is consistent with previous reports that indicate MSL1 as the major enhancer of MOF's H4K16 acetylation (Kadlec et al., 2011) and demonstrate relaxed substrate specificity for the NSL complex (Zhao et al., 2013b). In addition, we found that MSL1-depletion affected the levels of MSL2 but not of NSL complex members while the depletion of KANSL3 moderately decreased protein levels for both complexes (Figure 3E). ChIP-qPCR assays in MOF-depleted cells revealed that MSL1 and KANSL3 do not require the presence of MOF to bind to gene promoters, which is in agreement with previous observations in *D. melanogaster* (Hallacli et al., 2012) (Figure 3—figure supplement 4).

In summary, our TSS-focused analysis shows that the localized binding of the NSL complex to the promoters of housekeeping genes appears to be a conserved feature between the mammalian and *Drosophila* systems. Unlike in the fly, we do not detect an MSL- and X-chromosome-specific binding mode of MOF in the mouse cell lines. Instead, both complexes narrowly bind to TSSs where their co-occurrence is associated with significantly higher median expression values than those solely bound by the NSL complex. Moreover, we found that MOF is dispensable for the TSS recruitment of its interaction partner and that depletions of the individual proteins predominantly result in the downregulation of TSS-bound genes, further supporting the fact that the promoter-binding of the MSL as well as the NSL complex is associated with active transcription.

MSL and NSL complex members individually bind to active enhancers in ESCs

In addition to promoter-proximal binding, where both the MSL and NSL complex tend to (co-)occur constitutively in ESCs and NPCs, we identified a large proportion of binding sites where the proteins were present in a dynamic fashion, i.e. their binding was observed only in ESCs but not in NPCs (Figure 2, clusters D and E). In contrast to the binding mode

represented by clusters A and B (Figure 2), here MSL2, MCRS1 and KANSL3 were predominantly enriched within introns and intergenic regions that underwent significant CpG methylation upon differentiation (e.g. from median 50 % CpG methylation in ESCs to more than 80 % in NPCs for cluster D; bisulfite sequencing data from Stadler et al., 2011). As shown in Figure 4A, CpG methylation in NPCs was particularly pronounced around the center of the regions with significant ChIP enrichments in ESCs, indicating a correlation between the loss of ChIP-seq signal for MOF, MSL1, MSL2, KANSL3 and MCRS1 and DNA methylation upon differentiation. In addition, the regions of cluster D and, to a lesser extent the MSL1-rich cluster E (Figure 2), showed highly localized enrichments of DNase hypersensitivity sites (DNase HS), RNA Polymerase II (Pol II), p300, methylation of histone 3 on lysine 4 (H3K4me1), and acetylation of histone 3 on lysine 27 (H3K27ac) in ESCs (Figure 4A), which are characteristic features of enhancer regions. We thus examined whether MOF and its interaction partners were enriched on known enhancer regions, using lists of typical and super enhancers defined by binding sites of the pluripotency factors SOX2, NANOG, and OCT4 (Whyte et al., 2013) as well as sets of active and poised enhancers based on histone mark signatures (Creyghton et al., 2010).

Interestingly, MSL2, KANSL3 and MCRS1, but not MOF and MSL1, showed profound enrichments for active and poised ESC enhancers (Figure 4—figure supplement 1A) as well as along the regions of super enhancers that have been described as being particularly important for maintenance of cell identity (Whyte et al., 2013).

The signals of MSL2 and KANSL3 were specific for ESC enhancers and wide-spread along super enhancer regions (Figure 4B and 4C). We noted that enhancers overlapping with MSL2 ChIP-seq peaks tended to show lower KANSL3 enrichments and vice versa, implying that MSL2 and KANSL3 preferred different enhancer regions (heatmaps in Figure 4B, Figure 4—figure supplement 1B). MOF was not enriched at super enhancers and generally, its binding to TSS-distal sites was much less pronounced than to gene promoters (Figure 4 – figure supplement 1A, 1C; Figure 2). Like for TSS-specific binding, MOF was not alone (87 % of TSS-distal MOF peak regions overlapped with either KANSL3 or MSL2). Since a recent report showed H4K16 acetylation to be present at p300- and H3K27-acetylation-independent enhancer regions (Taylor et al., 2013), we analyzed the moderate TSS-distal enrichments of MOF in more detail and observed a slight preference for TSS-distal regions that were not

overlapping with previously published ESC enhancer regions (Figure 4—figure supplement 2A). In fact, we detected the strongest MOF signals in regions with rather low enrichments of known enhancer marks (see DNase HS, p300, H3K4me1, H3K27ac in Figure 4—figure supplement 2B and 2C), which suggested a preferred binding of MOF outside canonical ESC regulatory regions.

In addition to ESC-specific binding of MSL2 and KANSL3 to predicted enhancers, we also identified a very distinct set of TSS-distal binding sites by MSL2 to introns and intergenic regions without enhancer-associated marks (cluster C in Figure 2). Approximately 81 % of these cluster C regions had solitary MSL2 enrichments without significant signals of any of the other ChIPed proteins. Interestingly, these MSL2 binding sites increased in number and binding strength upon differentiation to NPCs (829 solitary MSL2 peaks in ESCs compared to 3,635 in NPCs). In contrast to the previously described binding sites that were characterized by the prevalence of open, active chromatin (Figures 3 and 4), here MSL2 was excluded from hypo-methylated DNA regions (Figure 4—figure supplement 3A; note the different behavior of KANSL3). When we searched the unique MSL2 binding sites for DNA motifs, we obtained a $(CAGA)_n$ motif (Figure 4—figure supplement 3B) that was previously described as a binding site for SMAD3, a transcription factor that translates the TGF-beta receptor response into gene expression regulation (Zawel et al., 1998). When we subsequently scanned all the binding sites for the presence of the published, original SMAD3 motif, we found a strikingly specific signal for the center of the solitary MSL2 ChIP-seq peaks (Figure 4—figure supplement 3C).

We conclude that MOF, MSL2 and KANSL3 specifically recognize ESC enhancers. In contrast to MSL-MOF-NSL co-occurrence at housekeeping gene promoters, we found evidence for differential and independent binding of the individual proteins to gene bodies and intergenic regions suggesting the potential for distinct tissue-specific regulatory functions of MSL2 and KANSL3. This data reveals a newly evolved function of MSL2 and KANSL3 in mammals, which has not been observed in flies.

Genes associated with TSS-distal binding sites of MSL1 and MSL2 are frequently downregulated in cells lacking MSL1 or MSL2

To study the functional implications of the binding of MSL2 and KANSL3 to putative ESC enhancers, we first tested five different regions located near genes related to pluripotency

and self-renewal (Young, 2011, Hu et al., 2009). Using luciferase reporter constructs, we found strong transcriptional enhancement for all tested regions in ESCs, but not in NPCs or 3T3 cells which correlated with the presence of MSL2 and/or KANSL3 and MCRS1 in ESCs only (Figure 4D and Figure 4—figure supplement 4A).

We then used our RNA-seq data sets from MSL1-, MSL2-, MOF-, and KANSL3-depleted cells to assess the effects on the transcription of those genes that were not bound at promoters, but had been predicted by GREAT (McLean et al., 2010) to be regulated by TSS-distal binding sites of the respective protein. As shown in Figure 4E, we again found similar effects for KANSL3- and MOF-depleted cells compared to MSL1- and MSL2-depleted cells with the latter group showing genome-wide downregulation of predicted target genes. In fact, the numbers of TSS-distal targets of MSL1 or MSL2 that were significantly and negatively affected in the respective shRNA-treatments were markedly larger than for genes where MSL1 or MSL2 bound to the promoter (compare Figure 3F with Figure 4—figure supplement 4B). Moreover, in MSL2-, but not KANSL3-depleted cells, the effects on TSS-distally targeted genes were slightly stronger than for TSS-targets (Figure 4—figure supplement 4C).

Depletions of MOF and KANSL3, but not of MSL complex members affect key pluripotency factors

While TSS-binding predominantly occurred at housekeeping genes, we noticed that the majority of enhancer regions associated with key pluripotency factors (e.g. SOX2, ESRRB, MYC, REX1, TBX3, NANOG) were strongly enriched for MSL2 as well as KANSL3. We thus assessed the effects of the protein depletions on pluripotency factors in ESCs and found strongly reduced levels of NANOG, REX1 and ESRRB in MOF- or KANSL3-depleted cells. Surprisingly, the pluripotency factors remained almost unaffected in cells depleted for MSL1 or MSL2 (Figure 4F). These contrasting results were mirrored by decreased levels of alkaline phosphatase (AP) in MOF- and KANSL3-, but not in MSL1- or MSL2-depleted cells (Figure 4—figure supplement 4D).

These findings indicate that despite their frequent effects on TSS-distally targeted genes, MSL1 and MSL2 might not show dominant effects at genes that are bound by KANSL3 as well. Therefore, we specifically searched for regions without KANSL3 binding to identify putative MSL-specific functions.

The MSL complex binds multiple loci within the X inactivation center

As described previously, we identified only a small subset of regions in the mouse genome where MSL complex members were enriched exclusively (see cluster E in Figure 2). Strikingly, several of these binding sites fall into a region known as the X inactivation center (XIC). The XIC is the X-chromosomal region necessary and sufficient to control the inactivation of one of the two X chromosomes in females (reviewed in Pollex and Heard, 2012).

The XIC site with the strongest concomitant enrichments of MSL1, MSL2 and MOF was the major promoter (P2) of *Tsix* and its intronic minisatellite – *DXPas34* (Figure 5A and 5B). *DXPas34* is a well-characterized tandem repeat that serves as a binding platform for multiple transcription factors and contains bidirectional enhancing properties essential for the expression of *Tsix*, the antisense transcript of *Xist* (Navarro et al., 2010, Gontan et al., 2012, Donohoe et al., 2007, Debrand et al., 1999, Cohen et al., 2007). In rodents, *Tsix* antisense transcription across the *Xist* promoter is required for regulating the levels of *Xist* accumulation. In turn, *DXPas34* deletion impairs the recruitment of Pol II and TFIIB to the major promoter of *Tsix* causing its downregulation (Vigneau et al., 2006).

In addition to the *DXPas34* binding site, we detected MSL peaks on the promoters, gene bodies and intronic regions of other key XIC regulators including the genes of the long non-coding (lnc) RNAs *Xist* and *Jpx*. Additionally, we observed peaks upstream of the *Tsx* gene and both at the TSS and downstream of the *Rnf12* gene (Figure 5A). Products of all of these genes were shown to play important roles in orchestrating the process of X inactivation (Tian et al., 2010, Sun et al., 2013, Stavropoulos et al., 2001, Shin et al., 2010, Gontan et al., 2012, Chureau et al., 2011, Anguera et al., 2011).

The XIC binding of MSL-MOF was specific to ESCs, as almost all enrichments were abolished upon differentiation, except for some loci upstream of *Xist* where traces of binding could still be detected in NPCs (for example *Ftx* and *Jpx* TSS, Figure 5—figure supplement 1A).

We next confirmed the high ChIP-seq enrichments of MSL1, MSL2 and MOF and assessed H4K16 acetylation on the major promoter of *Tsix* and along *DXPas34* with ChIP-qPCR assays covering the entire region in male and female ESCs likewise (Figure 5C and Figure 5—figure supplement 1B). Interestingly, the recruitment of MOF was almost completely abolished in both MSL1- and MSL2-depleted cells whereas the depletion of MOF had no effect on MSL1

and MSL2 binding to the *Tsix* major promoter and *DXPas34* (Figure 5C). H4K16 acetylation ChIP signals were severely reduced in both MOF- as well as MSL2-depleted cells. These results are in agreement with our global observations (Figure 3G, Figure 3—figure supplement 4C) and indicate that MSL1 and MSL2 are together necessary and sufficient for the recruitment of MOF and for the deposition of H4K16 acetylation at *DXPas34*.

MSL1 and MSL2 are important for *Tsix* expression

To directly assess the functional outcome of MOF-, MSL1- and MSL2-depletions, we studied the expression of *Tsix* and *Xist* in shRNA-treated ESCs. Unexpectedly, only MSL1- and MSL2-, but not MOF-depletion led to pronounced downregulation of *Tsix* both in male and female ESCs (Figure 6A; note that in our RNA-seq data set for MSL2-depleted cells, *Tsix* was among the 5 most strongly downregulated genes). Downregulation of *Tsix* was accompanied by moderately elevated *Xist* RNA levels in MSL1- and MSL2-depleted ESCs whereas depletion of MOF yielded the most pronounced (8-15 fold) upregulation of *Xist* without affecting *Tsix*.

To determine the effects on *Tsix* in individual cells, we next performed RNA-FISH with probes against *DXPas34* and *Huwe1* in female ESCs (*Huwe1* was used to mark X chromosomes, for probe references see Methods and Materials). The RNA-FISH confirmed the qPCR results as we observed global reduction and in many cases elimination of *DXPas34* signals in MSL1- and MSL2-, but not in MOF-depleted cells (Figure 6B and Figure 6—figure supplement 1A-C).

We next wanted to understand the mechanistic differences between the *Tsix*-specific and the *Tsix*-independent effects on *Xist* levels that we found for depletions of MSL1/MSL2 and MOF, respectively. As pluripotency factors are additional regulators of *Xist* (Navarro et al., 2008, Nesterova et al., 2011), we assessed the consequences of the different knockdowns on the *Xist*-related pluripotency network in female ESCs. Like for MOF- and KANSL3-depletions in male ESCs (see Figure 4F), the depletion of MOF (but not of MSL1 or MSL2) in female ESCs resulted in a significant decrease of transcript as well as protein levels of pluripotency factors that had previously been associated with *Xist* repression (eg. NANOG and REX1; see Figure 6C and Figure 6—figure supplement 1D).

Taken together, we detect direct binding of MSL complex members to several loci within the X inactivation center including the *Tsix/Xist* locus. Depletion of MSL1 or MSL2, but not MOF led to severe downregulation of *Tsix* expression while depletion of MOF, MSL1 or MSL2

resulted in elevated *Xist* levels. These results indicate a direct regulatory function of MSL1 and MSL2 on the *DXPas34* locus and an indirect NSL-associated MOF effect on *Xist* expression through the pluripotency network.

Depletion of MSL1 and MSL2 leads to impaired recruitment of REX1 and YY1 to regulatory regions of *Tsix*

As loss of MSL1 and MSL2 did not affect the core pluripotency network we set out to explore what might be the impact of MSL depletion on XIC genes (other than *Tsix* and *Xist*) as well as transcription factors involved in their regulation. As shown in Figure 6—figure supplement 1E, we observed mild effects on the expression of XIC-encoded genes involved in the regulation of X inactivation. Only depletion of MSL2 led to significant downregulation of *Ftx* and *Jpx* genes whose promoters were bound by MSL1 and/or MSL2 (see Figure 5A). On the other hand, depletion of MOF led to moderate upregulation of *Linx* lncRNA, which acts synergistically with *Tsix* (Nora et al., 2012).

Neither the depletion of MSL1 and MSL2, nor the depletion of MOF significantly influenced protein levels of RNF12, YY1 or CTCF which are known regulators of the XIC (Figure 6D) (Shin et al., 2010, Jonkers et al., 2009, Donohoe et al., 2007, Donohoe et al., 2009, Jeon and Lee, 2011). Since REX1 and YY1 bind and regulate the *Tsix* locus (Donohoe et al., 2007, Gontan et al., 2012), we subsequently tested whether MSL depletion would affect the recruitment of these factors to the *Tsix* major promoter and *DXPas34*. Indeed, the depletion of MSL2 led to significant reduction of REX1 ChIP signals across the *DXPas34* locus whereas the effect on YY1-targeting was less pronounced and restricted to the *Tsix* major promoter (P2) (Figure 6E).

Knockdown of *Msl1* and *Msl2* results in enhanced accumulation of *Xist* and X-chromosomal coating in differentiating female ESCs

We next assessed the consequence of MSL-dependent reduction of *Tsix* levels and concomitant upregulation of *Xist* at a cellular level using RNA-FISH for *Xist* upon depletion of individual MSL complex members (for probe reference see Methods and Materials). Interestingly, we observed accumulating *Xist* lncRNA and X-chromosomal coating in a small fraction of MSL1- and MSL2-depleted female ESCs (but not MOF-depleted cells; 4-5 % of the cell population in sh*Msl1* and sh*Msl2* with comparison to 0.5 % in scrambled control, see Figure 7A and Figure 7—figure supplement 1A-C). These findings suggest that the MSL1-and

MSL2-dependent downregulation of *Tsix* is sufficient to cause occasional accumulation of *Xist* lncRNA in undifferentiated female ESCs. The different outcomes following MOF and MSL1/MSL2 depletion on *Xist* confirmed the notion that MOF and MSL1/MSL2 influence the XIC via different mechanisms.

Previous studies have shown that the effects of *Tsix* depletion on *Xist* accumulation and X inactivation become fully apparent after induction of differentiation (Clerc and Avner, 1998, Debrand et al., 1999, Lee and Lu, 1999, Luikenhuis et al., 2001, Ohhata et al., 2006, Sun et al., 2006). We therefore depleted MSL1, MSL2 and MOF and induced differentiation for 3 days by withdrawing LIF and placing the ESCs in N2B27 media. Consistent with our previous results, the induction of differentiation resulted in a stronger elevation of *Xist* RNA levels in MSL1- and MSL2-depleted cells in comparison to the scrambled control (Figure 7B). As *Tsix* expression was not affected in MOF-depleted ESCs and *Xist* levels were already high before induction of differentiation, *Xist* upregulation between day 2 and 3 of differentiation was similar to the scrambled control.

To monitor the effect on the X chromosome more closely, we next performed *Xist* RNA-FISH in MSL1-, MSL2- and MOF-depleted cells after 3 days of differentiation. All three knockdowns resulted in enhanced *Xist* accumulation and X-chromosomal coating (63.1 %, 61.1 % and 50.6 % of all counted cells in sh*Msl1-,* sh*Msl2*- and sh*Mof*-treated ESCs, respectively, in comparison to scrambled control with 39.1 % of counted cells; see Figure 7C, 7D and Figure 7—figure supplement 2A and 2B). Interestingly, we observed that MSL1- and MSL2-depleted differentiating cells contained numerous cells with two inactive X chromosomes. The fraction of cells where both X chromosomes underwent XCI was approximately 10 fold higher in *Msl1* and *Msl2* knockdown compared to the scrambled control (Figure 7E). These results are in agreement with previously published data from homozygous *Tsix* mutants that exhibit irregular, “chaotic” choice for X inactivation (Lee, 2005).

Taken together, our data establishes MSL1 and MSL2 among the key regulators of *Tsix* transcription as the depletion of MSL proteins results in severe downregulation of *Tsix* transcription and enhanced accumulation of *Xist* during early differentiation.

Discussion

We present a thorough characterization of the histone acetyltransferase MOF and its two known complexes in mouse embryonic stem cells (ESCs) and neuronal progenitor cells (NPCs). We determined 5 basic modes of co-occurrence that revealed cell-type-specific as well as constitutive functions of the different proteins and support the notion that the NSL complex has general, housekeeping functions whereas the MSL complex predominantly performs more specialized tasks. We show that MOF and its associated proteins are involved in gene expression regulation via different means: first, they all target the promoters of housekeeping genes in a cell-type-independent manner and second, members of both complexes occupy different sets of ESC-specific enhancers that are essential for the maintenance of stem cell identity. We demonstrate the distinct and novel functions carried out by the MOF-associated complex members by revealing that both complexes contribute to the repression of X inactivation in ESCs via different means: While we establish the MSL complex as a direct regulator of *Tsix*, MOF and the NSL complex play an important role in the maintenance of pluripotency factors (Figure 8).

Global effects of MOF are correlated with the NSL complex

Our study sheds light onto the interplay between MOF and its complexes in mammals. Despite the fact that the depletion of KANSL3 does not strongly reduce global H4K16 acetylation levels, we observed strikingly similar protein and transcriptome changes in KANSL3- or MOF-depleted cells (Figure 3E-G). On the other hand, MSL1- and MSL2-depletion caused marked decreases of H4K16 acetylation (Figure 3G). This is consistent with previous reports that established MSL proteins as the main enhancers of MOF's H4K16 acetylation activity, while the NSL complex was shown to possess broader substrate specificity and can crosstalk with histone methylases (Kadlec et al., 2011, Cai et al., 2010, Zhao et al., 2013b). Unexpectedly, we observed remarkably different phenotypic changes in MSL1- or MSL2-depleted cells compared to MOF- and KANSL3-depleted cells (Figure 3E, Figure 3—figure supplement 3A). A striking example was the strong reduction of key pluripotency factors in KANSL3- and MOF-depleted cells that remain unaffected in MSL1- and MSL2-knockdowns (Figure 3G, Figure 4F). These results support the recent finding that MOF is vital for the maintenance of pluripotency (Li et al., 2012), but we furthermore show that this is an NSL- and not MSL-related function of MOF independent of H4K16 acetylation deposition.

Taken together, our data shows that while MOF is the major acetyltransferase for lysine 16 of histone 4 (Taipale et al., 2005), MSL-dependent H4K16 acetylation is one of several means through which MOF exerts its crucial biological functions. This notion was further supported by the finding that MOF predominantly binds to promoters of broadly expressed genes as part of the NSL complex and subsequently supports their transcription (Figure 3A-F). MSL1 and MSL2, on the other hand, bound to a relatively small subset of broadly expressed MOF-NSL-targeted genes that were significantly stronger expressed than those where MOF was exclusively present with NSL complex members (Figure 3B, Figure 3—figure supplement 1B). The additive effects of the complexes on gene expression were intriguing, and whether they influence each other's activity or exert their functions separately should be studied in the future. We propose that the MSL complex fine-tunes MOF's activity and ensures precise regulation of more specific targets – after all, their presence is essential for the recruitment of MOF to NSL-independent targets (Figure 5B). Our model is surprisingly similar to the picture that is emerging from *Drosophila* research where the NSL complex regulates housekeeping genes (Lam et al., 2012, Feller et al., 2012) while the MSL complex fulfills a highly specialized role on the male X chromosome (reviewed in (Conrad and Akhtar, 2011).

MSL2 and KANSL3 can contribute to transcription via enhancer binding

In addition to insights about MOF-related functions of MSL and NSL complexes, we show for the first time additional binding of MSL and NSL proteins to TSS-distal regions with enhancer characteristics. On a global scale, MOF did not yield strong enrichments for canonical enhancers; however, both MSL2 as well as KANSL3 showed robust signals for TSS-distal regions in ESCs, but not in NPCs which reflected the transcriptional activity of these regions (Figure 2, Figure 4A). This apparent MOF-independent binding of the individual proteins (that tended to prefer different sets of enhancers; Figure 4B) suggests that KANSL3 and MSL2 stimulate transcription even in the absence of the histone acetyltransferase. Both proteins are in principle capable of supporting transcription: the *Drosophila* homologue of KANSL3 can directly activate transcription *in vitro* (Raja et al., 2010) and human MSL2 acts as an E3 ubiquitin ligase at lysine 34 of H2B (H2BK34ub) (Wu et al., 2011) which has been suggested to promote methylation of H3K4, and thus gene expression (Wu et al., 2011). Indeed, we observed several hundred genes that had been predicted to be regulated by TSS-distal binding sites of MSL2 or KANSL3 to be downregulated in the respective knockdowns with particularly high frequencies in MSL2-depleted cells (Figure 4E). It is important to note that

the subset of ESC enhancers for key pluripotency factors (e.g. *Klf4*, *Sox2*) were bound concomitantly by KANSL3 as well as MSL2 and only the depletion of KANSL3, but not of MSL1 or MSL2 diminished protein and transcript levels of these key ESC molecules (see above). It is possible that KANSL3 could rescue loss of MSL2 at certain loci, but the exact mechanisms through which KANSL3 affects transcription via enhancer-binding need to be studied further. Furthermore, the pluripotency network and/or Mediator-related functions at super enhancers may be sufficient and dominant over MSL2 to maintain the expression of the pluripotency factors in the absence of MSL2, but may well be dependent on the function of KANSL3 at these regions.

MSL1 and MSL2 repress X inactivation by regulating *Tsix* expression

When we specifically searched for regions where KANSL3 was not present together with MSL1 and MSL2, we found that the X inactivation center (XIC) showed numerous signals for the MSL complex (Figure 5). The XIC, a hot-spot of regulatory lncRNAs, is an X-chromosomal region that contains the main regulators of X chromosome inactivation (XCI). The proper function of XIC-located non-coding RNAs is influenced by the spatial organization of the XIC and governed by a sophisticated interplay of multiple transcription factors such as pluripotency factors (Deuve and Avner, 2011, Donohoe et al., 2007, Gontan et al., 2012, Navarro et al., 2010, Nora et al., 2012).

We found that depletion of MSL1 and MSL2 severely reduced *Tsix* expression in male as well as in female ESCs, moderately increased *Xist* levels (Figure 6A), but left pluripotency factors unaffected (Figure 6C). In contrast, MOF-depleted cells showed downregulation of pluripotency factors and much higher *Xist* levels. Previous studies demonstrated that in undifferentiated ESCs, where pluripotency factors are highly abundant, even severe downregulation of *Tsix*, or *Tsix*-deletion, have almost no effect on *Xist* transcription (Navarro et al., 2005, Morey et al., 2001, Nesterova et al., 2011). Thus, the pronounced *Xist* upregulation seen in MOF-depleted cells seems to be an indirect effect due to the downregulation of pluripotency factors, while the reduction of *Tsix* transcripts in MSL1- and MSL2-depleted cells, where pluripotency factors remain unaffected, has milder consequences on *Xist* levels.

Consequently, we could show that once ESCs are forced to initiate differentiation, the depletion of MOF has mild effects while MSL1- and MSL2-depleted cells, in which *Tsix*

expression is prematurely downregulated, indeed suffer from enhanced *Xist* accumulation accompanied by "chaotic" X inactivation (different numbers of inactivated X chromosomes within a population of cells; Figure 7B-E). This is consistent with the notion that the repressive potential of *Tsix* on *Xist* accumulation and the role of *Tsix* and the *DXPas34* locus in the process of counting and choice of XCI (Lee, 2005, Vigneau et al., 2006) becomes fully apparent during early stages of differentiation where additional repressive factors such as pluripotency factors are downregulated (reviewed in (Rougeulle and Avner, 2004).

Conclusion

We show that NSL and MSL complex members can function in concert to ensure proper regulation of gene expression, but our findings also strongly imply that members of both complexes have the capacity to act independently. In the case of the X inactivation center, we observe that the MOF-interacting proteins, despite engaging different regulatory means (MSL1, MSL2 through direct regulation of *Tsix*, and MOF-NSL through the pluripotency network) synergize to ensure the proper expression of the X chromosomes in undifferentiated ES cells (Figure 8). Our study sets the ground for future research to dissect the intricate interactions and specific functions of MOF and its associated major regulatory proteins in more detail.

Acknowledgements

We would like to thank T. Jenuwein from the MPI-IE, Freiburg for providing the WT26 ES cell line and U. Riehle from the Deep Sequencing Unit of the MPI-IE, Freiburg for sequencing all samples. We would also like to thank P. Kindle (MPI-IE, Freiburg, Imaging facility), S. Toscano and M. Shvedunova for help with imaging. TC is especially grateful to A. Chatterjee for providing support and insightful discussions. We thank all members of the labs for helpful discussions. This work was supported by DFG-SFB992 and BIOSSII awarded to AA and RB; DFG-SFB746 awarded to AA. AA is also a member of the EU-NoE-EpiGeneSys.

Methods and Materials

Cell culture

All cell culture was performed in a humidified incubator at 37°C and 5 % CO2. The feeder-dependent mouse female embryonic stem cell line (Zollino et al.) F1-21.6 was cultivated on mitomycin-C-inactivated or irradiated mouse embryonic fibroblasts (MEFs). The feeder-independent mouse male ES cell line WT26, a kind gift from the lab of Thomas Jenuwein, was cultivated on gelatin-coated dishes in ESC culture media KnockOut-DMEM (Gibco, Cat. No. 10829-018) supplemented with 1 % L-glutamine (Gibco, Cat. No. 25030-081), 1 % penicillin/streptavidin (Gibco, Cat. No. 15070-063), 1% non-essential amino acids (Gibco, Cat. No. 11140-050), 1 % sodium pyruvate, 1 % 2-mercaptoethanol. All ESC media contained 15 % FBS and 1000 U/ml (for feeder-dependentfree) or 2000 U/ml (for feeder-independent) of leukemia inhibitory factor (Anguera et al.).

Male and female neuronal progenitor cell (NPC) lines were derived from previously mentioned ES cell lines (see below). Mouse 3T3 cells (for luciferase assays) and human HEK293-FT cells (for lentiviral production) were cultivated in DMEM (high glucose, with glutamine, Gibco, Cat. No. 41965) supplemented with 10 % heat-inactivated serum (PAA, Cat. No. A15-101), 1 % L-glutamine, 1 % penicillin/streptavidin (Gibco, Cat. No. 25030-081 and 15070-063, respectively).

NPC differentiation

Mouse ESCs were differentiated into neuronal progenitor cells (NPC) as previously described (Splinter et al., 2011, Conti et al., 2005). In brief, $1x10^6$ ESCs (deprived of feeder cells) were plated on 0.1 % gelatin-coated dishes in N2B27 medium and cultured for 7 days with daily media changes. The cells were then dissociated from the plate using accutase (Sigma) and 3×10^6 cells were plated on a bacterial petri dish to induce formation of embryoid bodies in N2B27 medium supplemented with 10 ng/ml EGF and FGF2 (Peprotech). After 72 hours, embryoid bodies were transferred to 0.1 % gelatin-coated dishes to allow adhesion and expansion of NPCs from the embryoid bodies. NPC lines were maintained in N2B27 medium supplemented with EGF and FGF2 (10 ng/ml each), on 0.1 % gelatin-coated flasks. For FISH analysis, F1-21.6 ESCs were grown on gelatin-coated coverslips with a MEF-inactivated monolayer for 24 hours.

Western blot analysis

Immunoprecipitation and co-immunoprecipitation experiments were performed in IP buffer (25 mM HEPES pH 7.6, 150 mM KCl, 5 mM $MgCl_2$, 0.5 % Tween20, 0.2 mg/ml BSA, 1× complete protease inhibitors tablet) with 1 ml of nuclear extract (0.5 mg/ml). Extracts were incubated with 5 µg of the respective antibody or normal-rabbit/normal rat serum. For MSL1, 15 µl of antibody serum was used. Extracts were incubated for 2 hours, rotating at 4°C. Protein-A Sepharose beads (GE Healthcare), blocked with 1 mg/ml yeast tRNA and 1 mg/ ml BSA (NEB), were used for all ChIP and IP assays.

Immunoprecipitation assays (IP and ChIP)

For (co)immunoprecipitation (IP, co-IP) experiments, 1 ml of nuclear extract (0.5 mg/ml) was used. IPs were performed in IP buffer (25 mM HEPES pH 7.6, 150 mM KCl, 5 mM MgCl2, 0.5 % Tween20, 0.2 mg/ml BSA, 1× complete protease inhibitors tablet). Extracts were incubated with 5 µg of the respective antibody or normal-rabbit/normal rat serum. For MSL1 15 µl of antibody serum was used. Extracts were incubated for 2 hours, rotating at 4°C. Sepharose-A beads were used.

ChIP assays were performed as previously described (Pauli, 2010) with minor changes. Cells were fixed in 1 % molecular biology grade formaldehyde (Sigma) 9 minutes before being quenched with glycine (0.125 M final concentration). Cells were washed twice with ice-cold PBS and lysed on ice for 10 minutes with 10 ml of Farnham lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5 % NP-40 + Roche Protease Inhibitor Cocktail Tablet, filtered through 0.2 micron filter unit). Lysates were transferred to a Kontes dounce tissue grinder (K885300-0015, size B) and dounced 15 times in order to break the cells and keep nuclei mostly intact. Crude nuclear prep was transferred to 15 ml falcon tube and nuclei pelleted by centrifugation at 2,000 rpm at 4°C for 5 minutes. Nuclei were resuspended in RIPA lysis buffer (1×PBS, 1 % NP-40, 0.5 % sodium deoxycholate, 0.1 % SDS + Roche Protease Inhibitor Cocktail Tablet, filtered through 0.2 micron filter unit). The nuclear extract was subjected to chromatin shearing using the Diagenode Bioruptor Plus sonicator (at high setting for a total time of 25 minutes, 30 seconds ON, 30 seconds OFF). The sonicated mixture was centrifuged at 14,000 rpm at 4°C for 5 minutes and supernatant was collected. Chromatin was supplemented with 5 µg of primary antibody and incubated for 16 hours (antibodies used for ChIP are listed below). After incubation, 50 µl of 50 % slurry bead solution was added for

another incubation period (2 hours), then beads were washed: 4 times for 15 minutes with RIPA lysis buffer, 2 times for 1 minute with LiCl IP wash buffer (250 mM LiCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 % NP-40, 0.5 % DOC, filtered through 0.2 micron filter unit), 2 times for 1 minute with TE buffer (1 mM Tris-HCl pH 8.0, 1 mM EDTA, filtered through 0.2 micron filter unit). Washed beads were resuspended in 100 µl of IP elution buffer and subjected to overnight reverse cross-linking (RNase and proteinase K digestions) followed by DNA purification (DNA was purified using Minelute PCR purification kit from Qiagen). For single IP assay 50 µl of bead solution was used. Purified ChIPed DNA was subjected to qPCR amplification (Applied Biosystems). Input was used for normalization control. For primer pairs see Supplementary file 3.

Antibodies

For MSL1 antibody production, a GST-mMSL1 fusion protein (C-terminal, residues 254-616) was used to immunize rabbits; the final bleed was used in experiments. Antibody specificity was verified with IP and MSL1-specific RNAi followed by Western blot analysis and ChIP assay. We used several commercial antibodies: a-KANSL1 (Abnova PAB20355), a-KANSL3 (SIGMA, HPA035018), a-MCRS1 (Proteintech, 11362-1-AP), a-MOF (BETHYL, A3000992A), a-MSL2 (SIGMA HPA003413), a-NANOG (BETHYL, A300-397A), a-OCT3/4 (Santa Cruz, sc-5279), a-REX1 (Abcam, Ab28141), a-ESRRB (Perseus Proteomics, PP-H6705-00), a-KLF4 (Abcam, Ab72543), a-SOX2 (R&D Systems, AF2018), a-YY1 (BETHYL, A302-779A), a-RNF12/RILM (Proteintech, 16121-1-AP) a-GAPDH (BETHYL, A300-639A), a-NESTIN (Abcam, Ab93666), a-CTCF (Abcam, Ab70303), a-H3 (Abcam, Ab1791), a-H4 (Abcam, Ab10158), a-H4K16ac (Millipore, 07-329).

Luciferase assays

Enhancer candidate regions (see below) were cloned into the firefly luciferase plasmids (Promega, pGL4.23) and transfected into mouse ESCs and 3T3 fibroblasts using Lipofectamine-2000 reagent and into NPCs using LTX-PLUS reagent (Invitrogen). Transfections were performed according to the manufacturer's guidelines except for using a 1:6 DNA to Lipofectamine ratio. Cells were seeded one day prior to transfection to achieve 70-80 % confluency at the time of transfection. Next, cells were fed with antibiotics-free medium (ES medium with LIF for ESCs and OPTIMEM for NPCs and 3T3s) at least 30 minutes before transfection and the medium was changed back 6-8 hours after transfection (basal

neural medium with FGF and EGF for NPCs). 100 ng of firefly construct with the cloned candidate region was co-transfected with 1 ng of renilla luciferase construct (pRL-TK of Promega) per 96-well and harvested for luciferase assay after 24 hours. Cells were harvested for luciferase assay 24 hours after transfection. The Dual Luciferase Kit (Promega) was used according to the manufacturer's protocol but with reduced substrate volumes of LARII and Stop&Glo reagents (50 µL per well of a 96-well plate with 10 µL cell lysate). Luminescence was measured by using Mithras plate reader (Berthold).

The transfection efficiency was normalized by firefly counts divided by the renilla counts. The fold enhancement value was calculated by an additional normalization to minimal promoter alone activities in each experiment (the graphs represent at least three independent experiments that were performed in technical triplicates each with error bars representing standard error of the mean).

Luciferase assays

The following enhancer candidate regions were amplified from mouse genomic DNA by PCR and cloned into BamHI-SalI sites (downstream of luciferase gene) of firefly luciferase plasmid pGL4.23 (Promega):

Intron of *Esrrb* (chr12:87,842,537-87,843,719) with primers introducing BamHI and XhoI sites: ATAGGATCCGAAGTAATTGTCTATTGTATCAG (forward), TATCTCGAGAAGAAGAAAGACTGTGTTCAACTCC (reverse).

Upstream of *Lefty* (chr1: 182854617-182855516) with primers introducing BamHI and SalI sites: ATAGGATCCCTTGCGGGGGATATGAGGC (forward), TATGTCGACCTGGGCCTTTCTAAGGC (reverse).

Upstream of *Trim28* (*Kap1*) (chr18: 34309039-34310140) with primer introducing BamHI and SalI sites: ATAGGATCCGAGGACTATTTGAAGGATCTATT (forward), TATGTCGACCTCACTCCCCAACCTCCATTTC (reverse).

Upstream of *Apc* (chr18: 34309039-34310140) with primers introducing BamHI and SalI sites: ATAGGATCCCTGAGCAATGCTCTTCCTCACAAGC (forward), TATGTCGACTTATACTCCAAATAGAATTGTCTG (reverse).

Intron of *Tbx3* (chr5: 120129690- 120130617) with primers introducing BamHI and SalI sites:
ATAGGATCCATAAATAAATAAATAAATATCTGATTG (forward),
TATGTCGACCGCGAGTCTGGCGATGCCTTGTC (reverse).

RNA extraction followed by cDNA synthesis and quantitative real time PCR

cDNA was synthesized from 500 ng – 1 µg of total RNA (extracted from circa 1 million cells using Rneasy kit, Qiagen) with random hexamers using SuperScript-III First Strand Synthesis kit (Invitrogen). The qPCRs were carried in a total reaction volume of 25 µl containing 0.5 - 1 µL of cDNA, 0.4 µmol of forward and reverse primer mix and 50 % 2 × SYBR Green PCR Master Mix (Roche). Gene expression was normalized to multiple controls (*RplP0* or *Hprt*), using the 7500 software V2.0.4 for analysis (Applied Biosystems). For primer pairs used for expression profiling see Supplementary File 3C.

Lentiviral-based RNAi in ESCs

shRNA constructs were either obtained from Sigma in pLKO.1 or designed using Genescript and cloned (please see below for details). For cloning, forward and reverse complimentary DNA oligonucleotides (Eurofins MWG Operon) designed to produce AgeI (5') and EcoRI (3') overhangs were annealed at a final concentration of 2 µM in NEBuffer. The pLKO.1-puro plasmid was digested with AgeI and EcoRI, ligated to the annealed oligonucleotides, and transformed into HB101 competent cells (Promega). Plasmid DNA was purified using the QIAprep Spin Miniprep kit (QIAGEN), and the sequence was validated.

For production of lentiviral particles, 70 % confluent HEK293FT cells in a 10 cm tissue culture plate were co-transfected with 3.33 µg lentiviral construct, 2.5 µg psPAX2 packaging plasmid and 1 µg pMD2.G envelope plasmid using Lipofectamine-2000 reagent (Invitrogen). To transduce ESCs, either concentrated or diluted lentiviral particles were used. For concentrated lentivirus, transfections were scaled up and OPTIMEM (Invitrogen) added to the HEK293FT cells following transfection and the lentiviral supernatant collected at 48 and 72 hours post-infection. This was then concentrated using Amicon Ultra-15 centrifugal filter units (Millipore) and added to ESC media supplemented with LIF and 10 µg/mL polybrene (Millipore). For diluted lentivirus, ESC media without LIF was added to the HEK293FT cells and the lentiviral supernatant was collected after 48 hours, filtered through 0.22 µm filters (Whatmann), and added 1:1 with fresh ESC media supplemented with LIF and polybrene to the ESCs. ESCs were then subjected to selection with 1.0 µg/mL puromycin, passaged once,

and harvested on day 3, 4, 5 or 6 of knockdown depending on the experiment (the numbers of days are indicated in the corresponding results section).

The following shRNA sequences were used for the knockdowns:
CCGGCCTAAGCACTCTCCCATTAAACTCGAGTTTAATGGGAGAGTGCTTAGGTTTTTG (sh*Msl1*, SIGMA, TRCN0000241378),
CCGGCCCAGTCTCTTAGCCATAATGCTCGAGCATTATGGCTAAGAGACTGGGTTTTTG (sh*Msl2*, SIGMA, TRCN0000243429),
CCGGAAGGCCGAGAAGAATTCTATCTCGAGATAGAATTCTTCTCGGCCTTTTTTTG (sh*Mof*, GENSCRIPT designed),
CCGGCTCCAGTCCTCTTCGTCATTGCTCGAGCAATGACGAAGAGGACTGGAGTTTTTG (sh*Kansl3*, SIGMA, TRCN0000266995),
CCGGAAGTGGCGCCTTAGCAACAACCTCGAGGTTGTTGCTAAGGCGCACTTTTTTTG (sh*Mcrs1*, GENSCRIPT designed),
CCGGCAACAAGATGAAGAGCACCAACTCGAGACAATTCGGAAGAAATCTGAGCTTTTTG (Non-targeting control, SIGMA, SHC002).

Cell proliferation assay

Cells treated with respective shRNAs and scramble control were performed as described earlier in feeder-free W26 mouse ESCs. The cell count was monitored for 6 days post knockdown at 24-hour intervals. In brief, after 4 days of knockdown 6 sets of 0.4×10^4 cells per well were seeded in triplicates in a 12-well gelatinized plate. The cells were grown in ES cell culture medium supplemented with 2000 U/ml LIF and 1 µg/ml puromycin; the medium was changed every 24 hours. For counting, cells were trypsinized and counted using the Neubauer hemocytometer.

Alkaline phosphatase staining

Detection of alkaline phosphatase, a surface marker and indicator of undifferentiated ESCs, was performed using the following method: feeder-free W26 ESCs were transduced (4 days) with scramble or the shRNAs against the genes of interest. Cells were washed twice with PBS followed by fixation with 4 % PFA for 2-3 minutes. The cells were washed twice with PBS and stained for 20 minutes with staining solution (25mM Tris-Maleic acid buffer pH 9.0, 0.4 mg/ml α-Naphthyl Phosphate (Sigma, cat.no. N7255), 1 mg/ml Fast Red TR Salt (Sigma,

cat.no. F8764), 8 mM $MgCl_2$, 0.01 % Deoxycholate, 0.02 % NP40). The reaction was stopped by washing with water followed by two washes with 1×PBS.

RNA extraction for RNA-seq

Total RNA was extracted from WT26 ESCs and NPCs as biological triplicates using TRIzol® Reagent and treated with the TURBO Dnase kit (Ambion). For RNA-seq analysis, cDNA libraries were prepared using the Illumina TruSeq Stranded mRNA kit with 3 µg DNase-treated samples.

Feeder-free W26 ESCs were transduced with shRNAs specific for *Msl1*, *Msl2*, *Mof*, *Kansl3* and control shRNA as biological triplicates as described above. Briefly, following transduction for 24 hours, cells were washed with PBS thrice to remove the viral supernatant and subjected to puromycin selection (1.5 µg/ml) for 24 hours. In case of *Msl1/2*, *Mof*, control shRNA the cells were maintained in puromycin selection for 4 days and in case of *Kansl3,* the cells were maintained in puromycin-selection for 84 hours. An additional set of control shRNA was performed alongside with *Kansl3* for 84 hours. Total RNA from all the shRNA-treated cells was extracted using TRIzol® Reagent. Following total RNA isolation the samples were treated with DNase using the TURBO DNase kit (Ambion). The quality of the RNA was analysed using the Bioanalyzer and samples with RIN values between 9-10 were used for RNA-seq.

RNA-FISH

Xist and *Huwe1* probes were described previously (Chow et al., 2010). *Tsix* was detected with a *DXPas34* plasmid (Debrand et al., 1999). Approximately 1×10^5 of F1-21.6 ESCs were plated on gelatin-coated coverslips and incubated for 24 to 48 hours. After fixation and permeabilization, coverslips with cells were washed and stored in 70 % EtOH at -20°C. Then the coverslips were dehydrated in 80 %, 95 % and 100 % EtOH (5 min each) and briefly air-dried. FISH probes were labeled by nick translation (Abbott) with Spectrum Red-dUTP or Spectrum Green-dUTP following the manufacturer's instructions. Labeled probes were precipitated in the presence of salmon sperm (10 µg) and Cot-1 DNA (3 µg), denatured and competed with Cot-1 DNA for 45 min at 37°C. Cells were then directly hybridized with labeled probes at 37°C overnight. Next, coverslips were washed 3 times in 50 % formamide/2 × SSC followed by 3 washes in 2 × SSC at 42°C. Cells were stained with DAPI (0.2 mg/ml).

Immunofluorescence Staining (against NESTIN)

Approximately 1×10^5 of male W26 ESCs and NPCs were plated on gelatin-coated coverslips and incubated for 24 hours. The cells were washed twice with PBS and fixed with pre-warmed 4 % formaldehyde for 8 minutes at 37°C. Next, cells were washed thrice with and incubated in Permeabilization buffer (1×PBS, 0.2 % Triton X-100) for 5 minutes at room temperature. After permeabilization cells were incubated in Blocking buffer (1×PBS, 5 % BSA, 0.05% Triton-X100) for 30 minutes, stained for 1 hour with primary antibody (rabbit polyclonal a-NESTIN, 1:500). Next, cells were washed thrice with Wash buffer (1×PBS, 0.05 % Triton-X100) and incubated in 10% goat normal serum solution (Invitrogen) for 20 minutes. Secondary antibody (goat anti-rabbit Alexa Fluor-488, 1:1000) was added upon coverslips and incubated for 45 minutes.

Microscopy

We used a spinning disk confocal microscope (Observer 1/Zeiss) with Plan Apochromat 63x1.4-oil objective for magnification. 500 ms exposure time was used for all lasers. Sequential z-axis images were collected in 0.5 µm steps. ZEN Blue software was used for image analysis.

Sequencing

All samples were sequenced by the Deep Sequencing Unit (MPI-IE, Freiburg) using Illumina HiSeq2000. Library preparation was carried out following Illumina standard protocols for paired-end sequencing (50 bp reads). All raw reads can be found in the GEO database under the accession number GSE51746.

RNA-seq data processing

RNA-seq reads were mapped to Ensembl annotation NCBIM37/mm9 using TopHat2 (Kim et al., 2013) with the options mate-inner-dist, mate-std-dev and library-type (fr-firststrand). The distance between read mates (mate-inner-dist and mate-std-dev) were assessed individually for each sequenced library based on the output of the sequencer for average fragment size and CV value.

For FPKM value generation, cufflinks (version 2.1.1) was used for each transcript in each condition (3 replicates for ESC and NPC) with default parameters; CummeRbund was used for quality checks and data access (Trapnell et al., 2013). Based on the distribution of FPKM

values, active genes were defined as transcripts with mean FPKM ≥ 4 (average over the replicates).

Differential gene expression analysis

After mapping of the RNA-seq reads from the shRNA-treated samples (including scrambled control), the reads that mapped to the genome were counted using htseq-count (doi: 10.1101/002824) with the stranded option set to reverse. The annotations present in the *Mus musculus* gtf file from the ENSEMBL release 67 were used as reference for counting.

DESeq2 was used for differential expression analysis (Anders and Huber, 2010). In this analysis, all libraries from knockdown cells were compared in a pairwise manner with its corresponding scrambled shRNA samples. Within the DESeq2 workflow, the cooks-cutoff parameter was set to "FALSE" and the genes with an adjusted p-value <= 0.01 were defined as significantly affected.

ChIP-seq analysis

Read mapping and normalizations

After mapping of the paired-end reads to the mouse genome (mm9) using bowtie version 2 (Langmead and Salzberg, 2012), we filtered for duplicate reads, reads with mapping qualities smaller than 2 and ambiguously mapped reads using samtools (Li et al., 2009). We also removed reads mapping to the mitochondrial genome and “random” chromosomes as well as known major satellites and duplicated genome regions to avoid coverage biases.

For normalization procedures, several modules of the deepTools suite (https://deeptools.github.io) were used (Ramirez et al., 2014). To ensure a thoroughly fair comparison between all data sets, first, the GC bias of all mapped reads was determined and, if necessary, corrected so that input and ChIP samples had similar GC distributions of their reads (correctGCbias module). In addition, all aligned read files were corrected for sequencing depth using the signal extraction method proposed by (Diaz et al., 2012) and normalized to the cell-type-specific input (bamCompare module).

Peak calling and replicate handling

MACS (version 1.4) was used for peak calling on every sample individually as well as on the merged files of two replicates (Zhang et al., 2008). Only peaks present in both replicates

were considered, using the borders and summits defined by peak calling results for the merged replicates. In addition, peaks with $-10\log_{10}$(p-values) lower than 50 and false-discovery rate values greater than 0.1 % were excluded from down-stream analyses.

Annotation used for genome-wide analyses

We used the RefSeq gene list for genome version mm9/NCBI37. Unless specified otherwise, alternative transcription start sites were scored as individual TSS in the respective analyses. The list of genes with homologues in different species was downloaded from HomoloGene and subsequently filtered for pairs of mouse and fly genes that belong to the same clusters of homology ID (Yoshida et al.). CpG island information was downloaded from the UCSC Genome Browser (Wu et al., 2010), mean observed over expected CpG ratios were extracted for the TSSs +/- 0.5 kb using UCSC tools.

Clustering

For Figure 2, a matrix containing the normalized ChIP-seq signals for all peaks was generated as follows: first, the union of peaks was created using mergeBed from the bedtools suite (Quinlan and Hall, 2010); then each region was binned to 2 kb and the normalized ChIP values were extracted in 50 bp windows. The ChIP signal values were rank-transformed, converted into euclidean distances using the R function 'dist' and subsequently ordered according to their similarity by the 'hclust' function (using Ward's method). The resulting dendrogram was pruned to 2 to 10 clusters for which the individual ChIP signals for unscaled regions were extracted (Figure 2). Visual inspection revealed no striking differences of the binding patterns between the individual clusters for more than 5 clusters.

The 3 clusters displayed in the lower part of Figure 4—figure supplement 2B were obtained similarly: first, a matrix was generated that contained the normalized ChIP-seq values of MOF, p300, H3K4me1 and DNase hypersensitivity sites for all regions of cluster D that did not overlap with ESC enhancers. The regions were then scaled to 1,400 bp and mean values were computed for 50 bp bins using the computeMatrix module of deepTools (Ramirez et al., 2014). Further processing was done as described above; the resulting dendrogram was pruned to k = 3 and the enrichments of the different factors were computed and visualized for 10 kb regions using the heatmapper module of deepTools.

GO term analysis

For GO term analyses, we used two approaches: the web interface of DAVID (Huang da et al., 2009) and GREAT (McLean et al., 2010).

For DAVID, we determined genes overlapping with the peaks of the individual ChIP-seq samples (TSS region +/- 500 bp) and supplied the corresponding RefSeq-IDs. The background list contained the union of all TSSs bound by at least one ChIPed protein. We used the Functional Annotation Clustering tool, filtered with the option “high stringency” and manually grouped the returned clusters of gene functions with enrichment scores above 1.3 into even broader terms.

To assess the GO terms of genes that might be regulated by the TSS-distal binding sites of MOF, MSL1, MSL2, KANSL3 and MCRS1, we used GREAT (McLean et al., 2010) with the mouse genome as the background data set and default settings. We obtained the top-ranked biological processes of the genes suggested to be *cis*-regulated by the regions combined in cluster D (Figure 2).

Analysis of transcription factor binding sites

For the analysis of enriched transcription factor binding sites, we used the R package ChIPEnrich (http://sartorlab.ccmb.med.umich.edu/chip-enrich) and TRAP (Thomas-Chollier et al., 2011). The ChIPEnrich package takes peak regions as input and uses a logistic regression approach to test for gene set enrichments while normalizing for mappability and locus length. We supplied the regions belonging to the individual clusters of binding (A-E from Figure 2) and obtained the corresponding enriched transcription factors.

To plot the occurrences of the SMAD3 motif (V$SMAD3_Q6, TRANSFAC name M00701; Figure 4—figure supplement 3C), TRAP was used with the following command to generate a bedgraph file where the log likelihood of a SMAD3 motif occurrence is stored for the entire genome: ANNOTATEv3.04_source/Release/ANNOTATE_v3.04 -s mm9.fa --pssm /transfac.pssm -g 0.5 --ttype balanced -name M00701 -d | awk 'BEGIN{OFS=t}{print $1, $4+7, $4+8, $6}' > SMAD3.pssm.bedgraph

Heatmap visualizations and summary plots

Heatmaps displaying normalized read densities of ChIP-seq samples, % methylated CpGs and SMAD3 motif score (Figures 2, 3C, 4A, Figure 4—figure supplement 2B and 3) were generated with the computeMatrix and heatmapper modules of the deepTools package (Ramirez et al., 2014) with 'reference-point' mode. Heatmaps of fractions of overlapped regions as in Figure 3—figure supplement 1 and Figure 4B as well as $\log_2$ fold changes (knockdown/control) from RNA-seq experiments (Figure 3E) were generated with the function 'heatmap.2' from the R gplots package.

The values underlying the summary plots such as the meta-gene and meta-enhancer plots in Figures 3D, 4B, Figure 3—figure supplement 2B, Figure 4—figure supplement 1B, 2A and 2C were generated with the computeMatrix module of the deepTools package using either 'reference-point' or 'scale-regions' mode and were visualized with the R package ggplots2.

Working with genomic intervals

For general assessments of overlaps between bed-files and to extract scores for defined regions the bedTools suite (Quinlan and Hall, 2010) and UCSC tools (Kuhn et al., 2013) were used. The snapshots of the binding profiles were obtained with IGV browser (Thorvaldsdottir et al., 2013).

Target definitions

For each knockdown condition for which RNA-seq data had been generated (see above), significantly affected genes were used (adjusted p-value <= 0.01, see above for differential gene expression analysis). Then they were subdivided into TSS- (ChIP-seq peak overlap with TSS +/- 1 kb), TSS-distal- (ChIP-seq peaks not overlapping with TSS +/- 1 kb) and non-targets (neither TSS overlap nor part of TSS-distal list). A gene was classified as TSS-distally regulated when at least one of the following criteria was true:

1. TSS-distal peaks overlapped its published super or typical enhancer (Whyte et al., 2013)
2. TSS-distal peaks were predicted by GREAT (McLean et al., 2010) to regulate the respective gene
3. TSS-distal peaks overlapped with at least one intron

937 Genes were defined as MSL targets when peaks of MOF and MSL1|MSL2 were overlapping
938 at the TSS +/- 1 kb or TSS-distal peaks were predicted to regulate the same putative target
939 gene. NSL targets were defined the same way, but with co-occurrences of peaks from MOF
940 and KANSL3|MCRS1.

941

References

AKHTAR, A. & BECKER, P. B. 2000. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell,* 5**,** 367-75.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biol,* 11**,** R106.

ANG, Y. S., TSAI, S. Y., LEE, D. F., MONK, J., SU, J., RATNAKUMAR, K., DING, J., GE, Y., DARR, H., CHANG, B., WANG, J., RENDL, M., BERNSTEIN, E., SCHANIEL, C. & LEMISCHKA, I. R. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell,* 145**,** 183-97.

ANGUERA, M. C., MA, W., CLIFT, D., NAMEKAWA, S., KELLEHER, R. J., 3RD & LEE, J. T. 2011. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet,* 7**,** e1002248.

CAI, Y., JIN, J., SWANSON, S. K., COLE, M. D., CHOI, S. H., FLORENS, L., WASHBURN, M. P., CONAWAY, J. W. & CONAWAY, R. C. 2010. Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem,* 285**,** 4268-72.

CHOW, J. C., CIAUDO, C., FAZZARI, M. J., MISE, N., SERVANT, N., GLASS, J. L., ATTREED, M., AVNER, P., WUTZ, A., BARILLOT, E., GREALLY, J. M., VOINNET, O. & HEARD, E. 2010. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell,* 141**,** 956-69.

CHUREAU, C., CHANTALAT, S., ROMITO, A., GALVANI, A., DURET, L., AVNER, P. & ROUGEULLE, C. 2011. Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum Mol Genet,* 20**,** 705-18.

CLERC, P. & AVNER, P. 1998. Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation. *Nat Genet,* 19**,** 249-53.

COHEN, D. E., DAVIDOW, L. S., ERWIN, J. A., XU, N., WARSHAWSKY, D. & LEE, J. T. 2007. The DXPas34 repeat regulates random and imprinted X inactivation. *Dev Cell,* 12**,** 57-71.

CONRAD, T. & AKHTAR, A. 2011. Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet,* 13**,** 123-34.

CONTI, L., POLLARD, S. M., GORBA, T., REITANO, E., TOSELLI, M., BIELLA, G., SUN, Y., SANZONE, S., YING, Q. L., CATTANEO, E. & SMITH, A. 2005. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol,* 3**,** e283.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

CUI, G., PARK, S., BADEAUX, A. I., KIM, D., LEE, J., THOMPSON, J. R., YAN, F., KANEKO, S., YUAN, Z., BOTUYAN, M. V., BEDFORD, M. T., CHENG, J. Q. & MER, G. 2012. PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nat Struct Mol Biol,* 19**,** 916-24.

DEBRAND, E., CHUREAU, C., ARNAUD, D., AVNER, P. & HEARD, E. 1999. Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression. *Mol Cell Biol,* 19**,** 8513-25.

DENSLOW, S. A. & WADE, P. A. 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene,* 26**,** 5433-8.

DEUVE, J. L. & AVNER, P. 2011. The coupling of X-chromosome inactivation to pluripotency. *Annu Rev Cell Dev Biol,* 27**,** 611-29.

DIAZ, A., PARK, K., LIM, D. A. & SONG, J. S. 2012. Normalization, bias correction, and peak calling for ChIP-seq. *Stat Appl Genet Mol Biol,* 11**,** Article 9.

DONOHOE, M. E., SILVA, S. S., PINTER, S. F., XU, N. & LEE, J. T. 2009. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature,* 460**,** 128-32.

DONOHOE, M. E., ZHANG, L. F., XU, N., SHI, Y. & LEE, J. T. 2007. Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell,* 25**,** 43-56.

FARRE, D., BELLORA, N., MULARONI, L., MESSEGUER, X. & ALBA, M. M. 2007. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol,* 8**,** R140.

FELLER, C., PRESTEL, M., HARTMANN, H., STRAUB, T., SODING, J. & BECKER, P. B. 2012. The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. *Nucleic Acids Res,* 40**,** 1509-22.

GENDREL, A. V., ATTIA, M., CHEN, C. J., DIABANGOUAYA, P., SERVANT, N., BARILLOT, E. & HEARD, E. 2014. Developmental dynamics and disease potential of random monoallelic gene expression. *Dev Cell,* 28**,** 366-80.

GONTAN, C., ACHAME, E. M., DEMMERS, J., BARAKAT, T. S., RENTMEESTER, E., VAN, I. W., GROOTEGOED, J. A. & GRIBNAU, J. 2012. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature,* 485**,** 386-90.

GUPTA, A., GUERIN-PEYROU, T. G., SHARMA, G. G., PARK, C., AGARWAL, M., GANJU, R. K., PANDITA, S., CHOI, K., SUKUMAR, S., PANDITA, R. K., LUDWIG, T. & PANDITA, T. K. 2008. The mammalian ortholog of Drosophila MOF that acetylates histone H4 lysine 16 is essential for embryogenesis and oncogenesis. *Mol Cell Biol,* 28**,** 397-409.

GUPTA, A., HUNT, C. R., PANDITA, R. K., PAE, J., KOMAL, K., SINGH, M., SHAY, J. W., KUMAR, R., ARIIZUMI, K., HORIKOSHI, N., HITTELMAN, W. N., GUHA, C., LUDWIG, T. & PANDITA, T. K. 2013. T-cell-specific deletion of Mof blocks their differentiation and results in genomic instability in mice. *Mutagenesis,* 28**,** 263-70.

HALLACLI, E., LIPP, M., GEORGIEV, P., SPIELMAN, C., CUSACK, S., AKHTAR, A. & KADLEC, J. 2012. Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in Drosophila. *Mol Cell,* 48**,** 587-600.

HU, G., KIM, J., XU, Q., LENG, Y., ORKIN, S. H. & ELLEDGE, S. J. 2009. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev,* 23**,** 837-48.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

JEON, Y. & LEE, J. T. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell,* 146**,** 119-33.

JONKERS, I., BARAKAT, T. S., ACHAME, E. M., MONKHORST, K., KENTER, A., RENTMEESTER, E., GROSVELD, F., GROOTEGOED, J. A. & GRIBNAU, J. 2009. RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell,* 139**,** 999-1011.

KADLEC, J., HALLACLI, E., LIPP, M., HOLZ, H., SANCHEZ-WEATHERBY, J., CUSACK, S. & AKHTAR, A. 2011. Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol,* 18**,** 142-9.

KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol,* 14**,** R36.

KOOLEN, D. A., KRAMER, J. M., NEVELING, K., NILLESEN, W. M., MOORE-BARTON, H. L., ELMSLIE, F. V., TOUTAIN, A., AMIEL, J., MALAN, V., TSAI, A. C., CHEUNG, S. W., GILISSEN, C., VERWIEL, E. T., MARTENS, S., FEUTH, T., BONGERS, E. M., DE VRIES, P., SCHEFFER, H., VISSERS, L. E., DE BROUWER, A. P., BRUNNER, H. G., VELTMAN, J. A., SCHENCK, A., YNTEMA, H. G. & DE VRIES, B. B. 2012. Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat Genet,* 44**,** 639-41.

KRUSE, J. P. & GU, W. 2009. MSL2 promotes Mdm2-independent cytoplasmic localization of p53. *J Biol Chem,* 284**,** 3250-63.

KUHN, R. M., HAUSSLER, D. & KENT, W. J. 2013. The UCSC genome browser and associated tools. *Brief Bioinform,* 14**,** 144-61.

KUMAR, R., HUNT, C. R., GUPTA, A., NANNEPAGA, S., PANDITA, R. K., SHAY, J. W., BACHOO, R., LUDWIG, T., BURNS, D. K. & PANDITA, T. K. 2011. Purkinje cell-specific males absent on the first (mMof) gene deletion results in an ataxia-telangiectasia-like neurological phenotype and backward walking in mice. *Proc Natl Acad Sci U S A,* 108**,** 3636-41.

LAM, K. C., MÜHLPFORDT, F., VAQUERIZAS, J. M., RAJA, S. J., HOLZ, H., LUSCOMBE, N. M., MANKE, T. & AKHTAR, A. 2012. The NSL complex regulates housekeeping genes in Drosophila. *PLoS Genet,* 8**,** e1002736.

LANDOLIN, J. M., JOHNSON, D. S., TRINKLEIN, N. D., ALDRED, S. F., MEDINA, C., SHULHA, H., WENG, Z. & MYERS, R. M. 2010. Sequence features that drive human promoter function and tissue specificity. *Genome Res,* 20**,** 890-8.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LEE, J. T. 2005. Regulation of X-chromosome counting by Tsix and Xite sequences. *Science,* 309**,** 768-71.

LEE, J. T. & LU, N. 1999. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell,* 99**,** 47-57.

LEE, K. K. & WORKMAN, J. L. 2007. Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol,* 8**,** 284-95.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, X., LI, L., PANDEY, R., BYUN, J. S., GARDNER, K., QIN, Z. & DOU, Y. 2012. The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. *Cell Stem Cell,* 11**,** 163-78.

LUIKENHUIS, S., WUTZ, A. & JAENISCH, R. 2001. Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. *Mol Cell Biol,* 21**,** 8512-20.

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol,* 28**,** 495-501.

MENDJAN, S., TAIPALE, M., KIND, J., HOLZ, H., GEBHARDT, P., SCHELDER, M., VERMEULEN, M., BUSCAINO, A., DUNCAN, K., MUELLER, J., WILM, M., STUNNENBERG, H. G., SAUMWEBER, H. & AKHTAR, A. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. *Mol Cell,* 21**,** 811-23.

MOREY, C., ARNAUD, D., AVNER, P. & CLERC, P. 2001. Tsix-mediated repression of Xist accumulation is not sufficient for normal random X inactivation. *Hum Mol Genet,* 10**,** 1403-11.

NAGY, Z., RISS, A., FUJIYAMA, S., KREBS, A., ORPINELL, M., JANSEN, P., COHEN, A., STUNNENBERG, H. G., KATO, S. & TORA, L. 2010. The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. *Cellular and Molecular Life Sciences,* 67**,** 611-628.

NAVARRO, P., CHAMBERS, I., KARWACKI-NEISIUS, V., CHUREAU, C., MOREY, C., ROUGEULLE, C. & AVNER, P. 2008. Molecular coupling of Xist regulation and pluripotency. *Science,* 321**,** 1693-5.

NAVARRO, P., OLDFIELD, A., LEGOUPI, J., FESTUCCIA, N., DUBOIS, A., ATTIA, M., SCHOORLEMMER, J., ROUGEULLE, C., CHAMBERS, I. & AVNER, P. 2010. Molecular coupling of Tsix regulation and pluripotency. *Nature,* 468**,** 457-60.

NAVARRO, P., PICHARD, S., CIAUDO, C., AVNER, P. & ROUGEULLE, C. 2005. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. *Genes Dev,* 19**,** 1474-84.

NESTEROVA, T. B., SENNER, C. E., SCHNEIDER, J., ALCAYNA-STEVENS, T., TATTERMUSCH, A., HEMBERGER, M. & BROCKDORFF, N. 2011. Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. *Epigenetics Chromatin,* 4**,** 17.

NORA, E. P., LAJOIE, B. R., SCHULZ, E. G., GIORGETTI, L., OKAMOTO, I., SERVANT, N., PIOLOT, T., VAN BERKUM, N. L., MEISIG, J., SEDAT, J., GRIBNAU, J., BARILLOT, E., BLUTHGEN, N., DEKKER, J. & HEARD, E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature,* 485**,** 381-5.

OHHATA, T., HOKI, Y., SASAKI, H. & SADO, T. 2006. Tsix-deficient X chromosome does not undergo inactivation in the embryonic lineage in males: implications for Tsix-independent silencing of Xist. *Cytogenet Genome Res,* 113**,** 345-9.

PAULI, F. 2010. Myers Lab ChIP-seq Protocol, v041610.1 and v041610.2. *In:* MYERS, R. (ed.). http://www.hudsonalpha.org/myers-lab.

PEEDICAYIL, A., VIERKANT, R. A., HARTMANN, L. C., FRIDLEY, B. L., FREDERICKSEN, Z. S., WHITE, K. L., ELLIOTT, E. A., PHELAN, C. M., TSAI, Y. Y., BERCHUCK, A., IVERSEN, E. S., JR., COUCH, F. J., PEETHAMABARAN, P., LARSON, M. C., KALLI, K. R., KOSEL, M. L., SHRIDHAR, V., RIDER, D. N., LIEBOW, M., CUNNINGHAM, J. M., SCHILDKRAUT, J. M., SELLERS, T. A. & GOODE, E. L. 2010. Risk of ovarian cancer and inherited variants in relapse-associated genes. *PLoS One,* 5**,** e8884.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

RAJA, S. J., CHARAPITSA, I., CONRAD, T., VAQUERIZAS, J. M., GEBHARDT, P., HOLZ, H., KADLEC, J., FRATERMAN, S., LUSCOMBE, N. M. & AKHTAR, A. 2010. The nonspecific lethal complex is a transcriptional regulator in Drosophila. *Mol Cell,* 38**,** 827-41.

RAMIREZ, F., DUNDAR, F., DIEHL, S., GRUNING, B. A. & MANKE, T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*.

ROUGEULLE, C. & AVNER, P. 2004. The role of antisense transcription in the regulation of X-inactivation. *Curr Top Dev Biol,* 63**,** 61-89.

SAPOUNTZI, V. & COTE, J. 2011. MYST-family histone acetyltransferases: beyond chromatin. *Cell Mol Life Sci,* 68**,** 1147-56.

SHIN, J., BOSSENZ, M., CHUNG, Y., MA, H., BYRON, M., TANIGUCHI-ISHIGAKI, N., ZHU, X., JIAO, B., HALL, L. L., GREEN, M. R., JONES, S. N., HERMANS-BORGMEYER, I., LAWRENCE, J. B. & BACH, I. 2010. Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. *Nature,* 467**,** 977-81.

SMITH, E. R., CAYROU, C., HUANG, R., LANE, W. S., COTE, J. & LUCCHESI, J. C. 2005. A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol,* 25**,** 9175-88.

SPLINTER, E., DE WIT, E., NORA, E. P., KLOUS, P., VAN DE WERKEN, H. J., ZHU, Y., KAAIJ, L. J., VAN IJCKEN, W., GRIBNAU, J., HEARD, E. & DE LAAT, W. 2011. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev,* 25**,** 1371-83.

STADLER, M. B., MURR, R., BURGER, L., IVANEK, R., LIENERT, F., SCHOLER, A., VAN NIMWEGEN, E., WIRBELAUER, C., OAKELEY, E. J., GAIDATZIS, D., TIWARI, V. K. & SCHUBELER, D. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature,* 480**,** 490-5.

STAVROPOULOS, N., LU, N. & LEE, J. T. 2001. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc Natl Acad Sci U S A,* 98**,** 10232-7.

SUN, B. K., DEATON, A. M. & LEE, J. T. 2006. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell,* 21**,** 617-28.

SUN, S., DEL ROSARIO, B. C., SZANTO, A., OGAWA, Y., JEON, Y. & LEE, J. T. 2013. Jpx RNA activates Xist by evicting CTCF. *Cell,* 153**,** 1537-51.

TAIPALE, M., REA, S., RICHTER, K., VILAR, A., LICHTER, P., IMHOF, A. & AKHTAR, A. 2005. hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol Cell Biol,* 25**,** 6798-810.

TAYLOR, G., ESKELAND, R., HEKIMOGLU-BALKAN, B., PRADEEPA, M. & BICKMORE, W. A. 2013. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res*.

THOMAS, T., DIXON, M. P., KUEH, A. J. & VOSS, A. K. 2008. Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. *Mol Cell Biol,* 28**,** 5093-105.

THOMAS-CHOLLIER, M., HUFTON, A., HEINIG, M., O'KEEFFE, S., MASRI, N. E., ROIDER, H. G., MANKE, T. & VINGRON, M. 2011. Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nat Protoc,* 6**,** 1860-9.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.

TIAN, D., SUN, S. & LEE, J. T. 2010. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell,* 143**,** 390-403.

TRAPNELL, C., HENDRICKSON, D. G., SAUVAGEAU, M., GOFF, L., RINN, J. L. & PACHTER, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol,* 31**,** 46-53.

VIGNEAU, S., AUGUI, S., NAVARRO, P., AVNER, P. & CLERC, P. 2006. An essential role for the DXPas34 tandem repeat and Tsix transcription in the counting process of X chromosome inactivation. *Proc Natl Acad Sci U S A,* 103**,** 7390-5.

VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature,* 457**,** 854-8.

WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.

WU, H., CAFFO, B., JAFFEE, H. A., IRIZARRY, R. A. & FEINBERG, A. P. 2010. Redefining CpG islands using hidden Markov models. *Biostatistics,* 11**,** 499-514.

WU, L., ZEE, B. M., WANG, Y., GARCIA, B. A. & DOU, Y. 2011. The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. *Mol Cell,* 43**,** 132-44.

XIE, X., LU, J., KULBOKAS, E. J., GOLUB, T. R., MOOTHA, V., LINDBLAD-TOH, K., LANDER, E. S. & KELLIS, M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature,* 434**,** 338-45.

YOSHIDA, K., TOKI, T., OKUNO, Y., KANEZAKI, R., SHIRAISHI, Y., SATO-OTSUBO, A., SANADA, M., PARK, M. J., TERUI, K., SUZUKI, H., KON, A., NAGATA, Y., SATO, Y., WANG, R., SHIBA, N., CHIBA, K., TANAKA, H., HAMA, A., MURAMATSU, H., HASEGAWA, D., NAKAMURA, K., KANEGANE, H., TSUKAMOTO, K., ADACHI, S., KAWAKAMI, K., KATO, K., NISHIMURA, R., IZRAELI, S., HAYASHI, Y., MIYANO, S., KOJIMA, S., ITO, E. & OGAWA, S. 2013. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat Genet*.

YOUNG, R. A. 2011. Control of the embryonic stem cell state. *Cell,* 144**,** 940-54.

ZAWEL, L., DAI, J. L., BUCKHAULTS, P., ZHOU, S., KINZLER, K. W., VOGELSTEIN, B. & KERN, S. E. 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell,* 1**,** 611-7.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHAO, W., LI, Q., AYERS, S., GU, Y., SHI, Z., ZHU, Q., CHEN, Y., WANG, H. Y. & WANG, R. F. 2013a. Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination. *Cell,* 152**,** 1037-50.

ZHAO, X., SU, J., WANG, F., LIU, D., DING, J., YANG, Y., CONAWAY, J. W., CONAWAY, R. C., CAO, L., WU, D., WU, M., CAI, Y. & JIN, J. 2013b. Crosstalk between NSL histone acetyltransferase and MLL/SET complexes: NSL complex functions in promoting histone H3K4 di-methylation activity by MLL/SET complexes. *PLoS Genet,* 9**,** e1003940.

ZOLLINO, M., ORTESCHI, D., MURDOLO, M., LATTANTE, S., BATTAGLIA, D., STEFANINI, C., MERCURI, E., CHIURAZZI, P., NERI, G. & MARANGI, G. 2012. Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat Genet,* 44**,** 636-8.

Figure legends

Figure 1: Distinct dynamics of MOF, MSL and NSL complexes during differentiation from ESCs to NPCs.

(A) We monitored the cell morphology during differentiation of mouse embryonic stem cells (Zollino et al.) into neuronal progenitor cells (NPC) via embryoid body formation (EB) with bright field microscopy. The day of differentiation is indicated in white boxes.

(B) Western blot analysis for ESC to NPC differentiation. Stages of differentiation together with the day of differentiation (d0 to d15) are indicated on top. GAPDH and histone 3 (H3) were used as loading controls. For expression analysis see Figure 1—figure supplement 1.

Figure 2: Distinct and shared binding sites of MOF and its complexes in mouse ESCs and NPCs.

We applied unsupervised clustering on the union of peaks from all ChIP-seq samples and thereby identified 5 distinct groups of binding for MOF, MSL and NSL proteins in ESCs and NPCs. Shown here are the input-normalized ChIP signals for each cluster of peaks including a size-matched control set of random genomic regions. The order of the regions is the same for all columns. The pie charts on the left indicate the number of regions from each cluster that overlap with gene bodies, the region 1 kb upstream of genes' TSS or intergenic regions.

Figure 3: Both MOF-complexes bind to the TSS of broadly expressed genes in mouse ESCs and NPCs.

(A) Genome browser snapshots of genes targeted by MSL and NSL complexes or by the NSL complex only. Signals were sequencing-depth-normalized and from ESCs. For ChIP-qPCR-based validation of the signals see (Figure 3—figure supplement 4B).

(B) Venn diagrams of genes whose promoter regions (TSS +/- 500 bp) overlapped with ChIP-seq peaks of NSL complex members (KANSL3 and/or MCRS1), MOF and MSL complex members (MSL1 and/or MSL2). The right-most panel depicts the overlap of genes bound by at least one factor in ESCs and NPCs.

(C) The heatmaps display the input-normalized ChIP enrichments of MOF, MSL2 and KANSL3 around the TSS of genes that were active in ESCs as well as NPCs based on RNA-seq data that we generated for both cell types.

(D) Summary plots of genes bound by the NSL complex in *D. melanogaster* for which mouse homologues were found. The input-normalized ChIP-seq signals around the TSS reveal markedly increased binding of MOF for male X-linked fly genes (left panels) that was not recapitulated in the mouse (right panels; ChIP-seq signals from ESCs). Fly genes were scaled to 1.2 kb and values were extracted from published data sets, mouse genes were scaled to 30 kb.

(E) Heatmap depicting results of RNA-seq experiments from different shRNA-treated cells. The colors correspond to $\log_2$ fold changes (shRNA-treated cells/scrambled control) for genes whose expression was significantly affected in all knockdown conditions. Values were ordered using hierarchical clustering.

(F) Bar plot of gene counts for different gene classes. We determined significantly up- and downregulated genes for each knockdown condition and binned them according to their expression strength in wild-type ESCs (high, intermediate, low). Then, for each gene, information about the TSS-targeting was extracted from the corresponding ChIP-seq sample. Non-target genes are neither bound at the promoter or the gene body and were not predicted to be regulated via TSS-distal binding sites in any of the 5 ChIP-seq ESC samples. For details on the target classification see Methods and Materials.

(G) Western blot analysis of MSL and NSL complex members and H4K16 acetylation in scrambled-, *Mof*-, *Msl1*- and *Kansl3*-shRNA treated male ESCs. 3 concentrations (100 %,

30 %, 10 %) of RIPA extract were loaded per sample. Asterisks mark the position of unspecific bands; triangles indicate the protein of interest.

Figure 4: MSL and NSL complex members are enriched at regions with enhancer marks in ESCs.

(A) Shown here are the fractions of methylated cytosines and ChIP-seq read densities of enhancer markers for regions of ESC-specific enrichments of our proteins of interest. We downloaded the different data from public repositories (see Supplementary file 3A for details) and calculated the values for the regions of the ESC-specific clusters D and E and random genomic loci. Most data sets used here were from mouse ESC except one RNA Polymerase II (Pol II) sample from NPC. All heatmaps were sorted according to the DNase hypersensitivity values except for CpG methylation heatmaps which were sorted according to their own values.

(B) Summary plots of input-normalized ChIP-seq signals along typical (TE) and super enhancers (SE) (Whyte et al., 2013). Note that we show the ESC-specific TE only while on the right-hand side we show the signal for SE regions from several cell types. Enhancer regions were scaled to 30 kb (SE) and circa 700 bp (TE). The heatmaps between the summary plots depict how much of each enhancer region overlaps with ChIP-seq peaks of MSL2 or KANSL3. ESC = embryonic stem cells (n = 232), pro-B = progenitor B cells (n = 396), Th = T helper cells (n = 437), C2C12 = myotube cells (n = 536).

(C) Exemplary genome browser snapshots of annotated super enhancers (SE, pink boxes) for 3 pluripotency factors displaying the sequencing-depth normalized ESC ChIP-seq signals of MSL2, MOF and KANSL3. See Figure 4—figure supplement 4C for additional examples.

(D) Luciferase assays demonstrate the biological activity of regions bound by MOF-associated proteins in ESCs (“in” stands for intronic region, “us” indicates that the cloned region is upstream of the gene). The firefly luciferase gene was cloned under a minimal promoter together with the putative enhancer region in ESCs, NPCs and 3T3 cells. The graphs represent at least 3 independent experiments performed in technical triplicates; error bars represent S.E.M.

(E) Bar plots depicting the fraction of significantly up- and downregulated genes per chromosome in the different shRNA-treated cells compared to shScrambled controls (total number of significantly affected genes per sample and chromosome labels are indicated). All genes counted here were classified as TSS-distal target genes in the respective ChIP-seq experiments. See Methods and Materials for details of the classifications.

(F) Western blot analyses of the pluripotency factors in scrambled-, *Mof*-, *Kansl3*-, *Msl1*- and *Msl2*-shRNA-treated male ESCs. For additional analyses in female ESCs see Figure 6C. The respective dilution (100 %, 30 %, 10 %) of loaded RIPA extract is indicated above each panel. Asterisks mark the position of unspecific bands; triangles indicate the protein of interest. GAPDH was used as the loading control. For antibodies see Methods and Materials.

Figure 5: The MSL complex binds multiple loci within the X inactivation center including the *Tsix DXPas34* minisatellite enhancer.

(A) Genome browser snapshots of the mouse X inactivation center (approx. 0.9 Mb) (upper panel) plus enlargement of the 164 kb region between *Chic1* and *Jpx/Enox* (lower panel). The signals shown are the sequencing-depth normalized profiles for ChIP-seq from ESCs (for corresponding profiles in NPCs see Figure 5 — figure supplement 1A); colored arrows indicate genes of lncRNAs. The schematic representation of the *DXPas34* locus depicts the locations of the primer pairs that were used for ChIP-qPCR analyses (Supplementary file 3B).
(B) Genome browser snapshots of the *DXPas34* minisatellite of sequencing-depth normalized ChIP-seq profiles in ESCs and NPCs.
(C) ChIP-qPCR analyses of MSL1 (blue), MSL2 (red), MOF (green) and H4K16 acetylation (purple) across the *Tsix* major promoter (*P2*) and the *DXPas34* enhancer in male ESCs treated with the indicated shRNAs. For corresponding ChIP-qPCR in female ESCs see Figure 5–figure supplement 1C. Panels in the middle show the effects of MOF depletion on the recruitment of MSL1 and MSL2 to *DXPas34* and vice versa. The bottom panel shows effects of depletion of control (dark pink), MOF (light pink) and MSL2 (purple) on the H4K16 acetylation signal. The labels of the x axes correspond to the arrowheads in (A). Results are expressed as mean +/- S.D. of 3 biological replicates; cells were harvested on day 4 (*Msl1*, *Msl2*) or 5 (*Mof*) after shRNA treatment. For primer pairs see Supplementary file 3C.

Figure 6: Depletion of MSL1 and MSL2 leads to downregulation of *Tsix* with concomitant upregulation of *Xist*

(A) Gene expression analysis for the indicated genes in male and female ESCs treated with scrambled RNA (shScram) or shRNA against *Msl1*, *Msl2* or *Mof*. All results are represented as relative values normalized to expression levels in shScram (normalized to *Hprt*) and expressed as means +/- S.D. in 3 biological replicates.

(B) RNA-FISH for *Huwe1* (red) and *DXPas34* (green) in: scrambled control, sh*Msl1*-, sh*Msl2*- and sh*Mof*-treated female ESCs. Nuclei were counterstained with DAPI (blue). White arrows denote foci corresponding to *Huwe1* or *Tsix*; dashed lines indicate nuclei borders. For additional images, phenotypes and quantifications see Figure 6 – figure supplement 1A-C. For probe references see Methods and Materials.

(C) Western blot analyses of the pluripotency factors in scrambled-, *Mof*-, *Msl1*- and *Msl2*-shRNA-treated female ESCs. For corresponding expression analyses see Figure 6 — figure supplement 1D and 1E. The respective dilution (100 %, 30 %, 10 %) of loaded RIPA extracts is shown above each panel. GAPDH was used as the loading control. For antibodies see Methods and Materials.

(D) Western blot analyses of the transcription factors involved in regulation of the XIC in scrambled-, *Mof*-, *Msl1* and *Msl2*-shRNA treated female ESCs. The respective dilution (100 %, 30 %, 10 %) of loaded RIPA extracts is shown above each panel. GAPDH was used as the loading control.

(E) ChIP-qPCR analysis of REX1 (left panel) and YY1 (right panel) across the *Tsix* major promoter (P2) and *DXPas34* in male ESCs treated with the indicated shRNAs. The labels of the x axes correspond to the arrowheads in Figure 5A. For all ChIP experiments, 3 biological replicates were used; results are expressed as mean +/- S.D.; cells were harvested on day 4 (*Msl2*) or 5 (*Mof*) after shRNA treatment.

Figure 7: MSL1 and MSL2 depletion leads to enhanced and chaotic *Xist* accumulation in early differentiation

(A) RNA-FISH for *Huwe1* (red) and *Xist* (green) in: scrambled control, sh*Msl1*-, sh*Msl2*- and sh*Mof*-treated female ESCs. Nuclei were counterstained with DAPI (blue). White arrows denote foci corresponding to *Huwe1* or *Xist*; dashed lines indicate nuclei borders. For additional images, phenotypes and quantifications see Figure 7 — figure supplement 1B-D. For probe references see Methods and Materials.

(B) Expression analysis for *Xist* in undifferentiated, day 2 (D2) and day 3 (D3) differentiating female ESCs treated with scrambled RNA (shScram) or shRNA against *Mof*, *Msl1* and *Msl2*. All results are represented as arbitrary units (*Xist* expression in undifferentiated ESCs = 1) normalized to expression levels in shScram (normalized to *Hprt*) and expressed as means +/- S.D. in 3 biological replicates. P-values for D2-to-D3 expression change were obtained using unpaired t-test.

(C) RNA-FISH for *Huwe1* (red) and *Xist* (green) in: scrambled control, sh*Msl1*-, sh*Msl2*- and sh*Mof*-treated differentiating female ESCs. Nuclei were counterstained with DAPI (blue). RNA-FISH was performed on the sixth day of knockdown (after 72 hours of differentiation). Percentages indicate number of cells with at least one *Xist* cloud for each of the knockdowns. For additional images of multicellular colonies see Figure 7 — figure supplement 2A.

(D) Bar plot summarizing the percentage of *Xist* clouds for individual knockdowns in differentiating (DAY3) female ESCs for individual knockdowns. Cells were divided into three categories: cells carrying no *Xist* clouds (Peedicayil et al.), single *Xist* cloud (light green) or two *Xist* clouds (dark green). For quantifications, see Figure 7 — figure supplement 2B.

(E) RNA-FISH for *Xist* (green) in: scrambled control, sh*Msl1*-, sh*Msl2*- and sh*Mof*-treated differentiating (DAY3) female ESCs. Here we show examples of individual nuclei carrying different patterns of *Xist* accumulation. Percentages correspond to the frequency of the shown *Xist* pattern within the population of cells. White arrows denote *Xist* foci; dashed lines indicate nuclei borders. For quantifications see Figure 7 – figure supplement 2B.

Figure 8: A summary model. Shared and distinct pathways by which MOF, MSLs and NSLs regulate gene expression, pluripotency and the X inactivation center.

(A) In this study, we have identified several modes of concurrent and independent binding of mammalian MOF, MSL and NSL proteins. We find that all complexes bind to promoters of housekeeping genes in ESCs and NPCs with NSL complex members occupying the majority of the target genes while MOF and MSL proteins bind NSL-bound genes in a more restricted manner. Furthermore, we observe that upon differentiation, KANSL3 and MSL2 additionally occupy TSSs of different sets of cell-type-specific genes in the absence of MOF.
(B) When we studied the functions of MSL and NSL complexes at the murine X inactivation center, we determined two basic mechanisms by which the different proteins affect the maintenance of two active X chromosomes in ESCs. (1) MSLs bind to the promoter and enhancer of *Tsix* whose transcription represses *Xist* expression. Upon depletion of MSLs, *Tsix* expression is compromised, so is REX1 recruitment to the *Tsix* locus. Consequently, *Xist* is increasingly transcribed and can occasionally accumulate. (2) In addition, MOF, MSLs and NSLs bind to ESC enhancers (E) and super enhancers (SE) of pluripotency factors. In WT ESCs, the high expression of pluripotency factors is another layer of *Xist* repression. The depletions of MOF or KANSL3, but not of MSL1 or MSL2 reduce the expression of pluripotency factors involved in *Xist* repression causing a *Tsix*-independent increase of *Xist* expression.

Supplementary Figure Legends

Figure 1—figure supplement 1: Monitoring RNA and protein levels in ESCs and NPCs.

(A) We monitored the expression dynamics during ESC differentiation for markers of pluripotency (*Oct4, Nanog, Rex1, Klf4*), embryoid body formation (*Fgf*5), differentiation (*Sox2*), and NPC (*Nestin*). Panels 3 and 4 contain the expression profiles for members of the MSL complex (*Msl1*, *Msl2*), *Mof*, and the NSL complex (*Kansl1*, *Kansl3*, *Mcrs1*), respectively. All results are represented as relative values individually normalized to *Rplp0* expression levels (panel 2) on a given day and to the highest expression level of a given gene during the entire differentiation process (highest expression level of each gene = 1). The x-axes show days of differentiation. All results are expressed as means +/- S.D. for technical replicates. For primers see Supplementary File 3C.

(B) Bright field images illustrate the cell morphology before and after the process of differentiation. The immunofluorescence analysis indicates the specific staining for the NESTIN protein (green) in neuronal progenitors (NPC); DNA is counterstained with DAPI (blue).

(C) Expression changes for selected ESC-specific and NPC-specific markers before and after differentiation of wild-type WT26 cells in using RT-PCR analysis and RNA-seq.

(D) Western blots for proteins from two ES cell lines and their NPC derivatives. Different dilutions were loaded (100 %, 30 %, 10 %) with the order indicated on top of the blots. Anti-GAPDH was used as loading control; arrows indicate the protein of interest.

Figure 1—figure supplement 2: Verification of antibodies used in this study.

(A) Immunoprecipitations from mouse ESC nuclear extracts with antibodies specific for KANSL1, KANSL3 or MOF and rabbit or rat antisera. The blot was probed with indicated antibodies showing the co-immunoprecipitation of several NSL complex members. Asterisks represent the IgG signal. Pol II = RNA Polymerase II.

(B) and (C) same as (A) except that immunoprecipitations were performed with antibodies specific to MSL1 (B) and MSL2 (C). Asterisks represent the IgG signal.

Figure 2—figure supplement: ChIP-seq quality measures.

(A) Correlation plot for all individual ChIP-seq and input samples from ESCs (left) and NPCs. The genome was sampled in windows of 10 kb length; the numbers of reads per bin were counted for each ChIP sample and correlated using Pearson correlation. The calculation and heatmap visualization were done with the bamCorrelate module from the deepTools suite (Ramirez et al., 2014).

(B) The bar chart depicts the fraction of ChIP-seq peaks for each protein that reside within each cluster shown in Figure 2, i.e. approximately 30 % of MSL1 peaks in ESCs locate in cluster E. Note that the absolute numbers of peaks differ between the samples (see Supplementary file 1B for absolute peak numbers and Methods and Materials for peak calling details).

Figure 3—figure supplement 1: MSL and NSL complexes target promoters of broadly expressed genes in ESCs and NPCs.

(A) The heatmap is related to Figure 3B as it is based on all genes that are bound by at least 1 ChIPed factor in ESCs or NPCs. The intensity of the color depicts the fraction of the 1 kb TSS-region that was covered by a binding site of MOF, MSL1, MSL2, KANSL3 or MCRS1. Rows and columns were sorted using hierarchical clustering on the Euclidean distances of the overlap fractions using R. The left color bar indicates which genes are targeted in 1 or both cell types.
(B) Distribution of expression values from RNA-seq data in ESCs and NPCs for genes targeted by MSL and NSL complex members together or by the NSL complex only. P-values were calculated using Welch t-test.
(C) Results of the GO term analysis using DAVID (Huang da et al., 2009) on genes that were bound at the TSS in ESCs by NSL complex members only or both MSL and NSL complexes.
(D) The pie charts depict how many times annotated TSSs overlapped with a CpG island. The vast majority of genes that were bound in ESCs by MSL and NSL together or by NSL complex members alone overlapped with at least 1 CpG island (dark and medium blue) while approximately 2/3 of the non-target-TSS did not overlap with any CpG island (light blue for 0 CpG islands within the queried regions).

Figure 3—figure supplement 2: The NSL-, but not the MSL-binding mode of *D. melanogaster* is present in mammalian cells.

(A) Exemplary genome browser snapshots of the X-linked fly gene CG4419. Shown here are the sequencing-depth normalized profiles for ChIP and corresponding input samples, clearly showing a broad enrichment of MOF and MSL1 along the entire gene body in male (m) *D. melanogaster* while all other marks show sharp enrichments around the TSS (including MSL1 and MOF in female (f) *D. melanogaster*) which are similar to those seen for both complexes in mouse cells (Figure 3A and 3D).

(B) Comparison of expressed (FPKM > 4) mouse genes whose homologous genes are either bound or not bound by MOF and its complexes in the fly. We extracted the input-normalized ChIP-seq values for 6 kb regions around the TSS using the computeMatrix module of deepTools (Ramirez et al., 2014). H3K4me3 signal is from a published data set, see Supplementary file 2 for the corresponding accession number.

Figure 3—figure supplement 3: Effects of shRNA-mediated depletion of MOF, MSL1, MSL2, and KANSL3.

(A) Time course of knockdown experiments. For experimental details see Methods and Materials. Samples for RNA-sequencing and AP staining (see Figure 4—figure supplement 4) were extracted 4 days after puromycin selection of shRNA-treated cells.

(B) Proliferation assay for shRNA-treated cells, starting at day 4 after puromycin selection (see Figure 3—figure supplement 3A).

(C) Bar plots depicting the fractions of genes (per chromosome) that were significantly up- or downregulated in RNA-seq experiments from shRNA treated cells. The left plot contains genes which were defined as TSS-targets in the respective ChIP-seq samples, the right plot contains genes that were neither classified as TSS- nor as TSS-distal targets. The labels on each bar indicate the chromosome name and the total number of genes that fulfilled the criteria for this chromosome (significantly affected, TSS-bound or non-targeted). See Methods and Materials for details of the classification.

Figure 3—figure supplement 4: Assessment of ChIP signals around the TSSs of putative target genes as determined by ChIP-seq.

(A) Genome Browser snapshots of several MSL/NSL (left) or NSL-only (Visel et al.) target genes and respective sequencing-depth-normalized ChIP-seq and input signals from ESCs. The exact genomic coordinates are indicated on top of each panel. Gene names are indicated on the bottom.

(B) ChIP-qPCR validation for MOF (green) and KANSL3 (purple) signals. Immunoprecipitated DNA was amplified by qPCR with primer sets positioned at the promoter (P) and end (E) of the coding sequence (Supplementary file 3A). Results are expressed as mean +/- S.D. of 3 biological replicates; cells were harvested for experiments on day 4 (*Kansl3*) or 5 (*Mof*) of knockdown.

(C) ChIP-qPCR for MSL1 (blue), MSL2 (red) and KANSL3 (purple) in ESCs treated with sh-RNA (scrambled or against a specific transcript). Signals on genes were evaluated using primers at the promoter (P), and end (E) of the coding sequence. Results are expressed as mean +/- S.D. of 3 biological replicates; cells were harvested for experiments on day 5 of *Mof* knockdown.

Figure 4—figure supplement 1: MSL2 and KANSL3 show strong enrichments at typical and super enhancers in ESCs.

(A) Boxplots demonstrating the distribution of mean ChIP enrichments for enhancer regions defined by H3K4me1 and H3K27ac marks in ESCs (see Creyghton et al., 2010 for details) that overlap with the clusters of binding defined by our ChIP-seq samples. Mean values were extracting using the UCSCtool bigWigAverageOverBed.

(B) Summary plots for typical enhancer regions (Whyte et al., 2013) that overlapped with either MSL2 (top) or KANSL3 (bottom) peaks. Different colors indicate different ChIP-seq signals. Related to the heatmaps of Figure 4B.

(C) Genome browser snapshots of sequencing-depth normalized ChIP-seq and input profiles for super enhancers of key pluripotency factors.

Figure 4—figure supplement 2: MOF is moderately enriched at non-canonical enhancers

(A) Summary plots of ChIP-seq values for binding sites belonging to cluster D. The regions were divided based on the presence or absence of annotated ESC enhancers (Whyte et al., 2013, Creyghton et al., 2010).

(B) Heatmaps of ChIP-seq read densities of known enhancer markers for the ESC-specific binding sites of our proteins of interest (cluster D, see Figure 2) and random genomic regions. The binding sites of cluster D (excluding regions with TSSs) were divided into 2 basic groups based on the presence or absence of known ESC enhancers (Whyte et al., 2013, Creyghton et al., 2010). The latter group was further divided into 3 (arbitrarily numbered) sub-clusters based on hierarchical clustering of the values from DNase hypersensitivity sites, p300, H3K4me1 and our MOF sample (in ESCs). Heatmaps of the ESC-enhancer-containing regions were sorted according to p300, those of the sub-clustered regions were sorted according to the MOF signal.

(C) Related to (B), shown here are the corresponding summary plots of ChIP-seq values for cluster D binding sites that do not overlap with annotated enhancer regions (bottom part of the heatmaps in the figure above). The 3 indicated groups are based on the hierarchical clustering that was performed on p300, H3K4me1 and MOF values ("Regions without annotated ESC enhancers" in (B)).

Figure 4—figure supplement 3: MSL2 has intergenic binding sites in DNA-hypomethylated regions that are enriched for SMAD3 binding sites.

(A) We extracted the percentage of methylated CpGs and the input-normalized ChIP-seq values from KANSL3 and MSL2 and 5 kb surrounding the center of the regions belonging to cluster C (Figure 2) and random genomic control regions. All heatmaps were sorted according to the percentages of methylated CpGs (Stadler et al., 2011).

(B) Motif obtained by MEME analysis on the top 200 MSL2 peaks within cluster C.

(C) Same as for (A), except that the score was the motif hit score for SMAD3 for 1 kb. See Methods and Materials for details.

Figure 4—figure supplement 4: Biological significance of the TSS-distal binding sites of the investigated proteins.

(A) Genome browser snapshots of sequencing-depth normalized ChIP-seq and input profiles. Pink boxes mark the regions cloned and transfected into ESCs and NPCs for luciferase assays (Figure 4D).

(B) Genes that were significantly up- or downregulated in the respective shRNA-treatments compared to shScrambled were classified according to ChIP-seq peak overlaps (TSS-distal, no target) and expression strength in wild type ESCs (high, intermediate, low). See Methods and Materials for details of the classifications.

(C) Distribution of absolute $\log_2$ fold changes (sh*Kansl3* or sh*Msl2* compared to shScrambled) for significantly downregulated genes. Different shades of orange indicate different target classes based on ChIP-seq experiments for KANSL3 or MSL2, respectively. P-values were calculated with Welch t-test.

(D) Alkaline phosphatase staining and morphology of ESC colonies in indicated knockdowns after 4 days growth under puromycin selection (see Figure 3—figure supplement 3A). MOF- and KANSL3-depleted cells demonstrate reduced alkaline phosphatase positive colonies with increased differentiation compared with MSL1- and MSL2-depleted cells and scrambled control.

Figure 5—figure supplement 1: The MSL proteins bind to multiple loci within the X inactivation center (XIC)

(A) Genome browser snapshots of the mouse XIC (top panel) with three enlargements on *Jpx*, *Ftx* and Rnf12 genes (lower panels). Red boxes with corresponding numbers mark the enlarged regions presented in the lower panels. The exact genomic coordinates are indicated on top of each panel, arrows represent genes. The signals shown are the sequencing-depth normalized ChIP-seq profiles in NPCs.

(B) ChIP analysis of MSL1, MSL2 and MOF across the *DXPas34* minisatellite in female ESCs. The x-axis labels indicate the genomic coordinates corresponding to the arrowheads in Figure 5A. The y-axes show the percentage of ChIP recovery for MSL1 and MSL2 (left-hand side) and MOF (right-hand side) normalized to input. For all ChIP experiments, 3 biological replicates were used; all results are expressed as mean +/- S.D.

Figure 6—figure supplement 1: Cells depleted of MSL1 or MSL2, but not MOF show loss of *DXPas34* foci

(A) RNA-FISH for *Huwe1* RNA (red) and *DXPas34* RNA (green) in shScrambled-, sh*Msl1*-, sh*Msl2*- and sh*Mof*-treated female ESCs. Shown here are examples of RNA-FISH signals for multicellular colonies and loss of *DXPas34* signal in MSL1- and MSL2-depleted cells. White boxes indicate cells enlarged and resented in Figure 6B. For all experiments, nuclei were counterstained with DAPI (blue).

(B) Summary of RNA-FISH for *DXPas34* and *Huwe1*. Red dots indicate the number of X chromosomes and green dots, *DXPas34* foci (smaller dot = reduced signal). Phenotypes that we observed in knockdowns are categorized into 4 groups containing cells with equal *Huwe1*/*DXPas34* ratio and with *DXPas34* loss. The percentages indicate how many cells per population showed the respective phenotype.

(C) Corresponding to Figure 6B. Summary of total cell counts from RNA-FISH for (*DXPas34*) and *Huwe1* in MSL1-, MSL2- or MOF-depleted female ESCs.

(D) Gene expression analysis for the indicated genes in female ESCs treated with scrambled RNA (shScram) or shRNA against *Mof*, *Msl1* and *Msl2*. All results are represented as relative values normalized to expression levels in shScram (normalized to *Hprt*) and expressed as means +/- S.D. in 3 biological replicates.

(E) Gene expression analysis for genes of the XIC in female ESCs treated with scrambled RNA or shRNA against Msl1, Msl2 or Mof. All results are represented as relative values normalized to expression levels in shScrambled (normalized to Hprt) and expressed as means +/- S.D. for 3 biological replicates.

Figure 7—figure supplement 1: Depletion of MSL1 and MSL2 leads to occasional accumulation and spreading of *Xist* in undifferentiated ESCs

(A) RNA-FISH for *Huwe1* RNA (red) and *Xist* RNA (green) in shScrambled- (top left) and sh*Mof*- (top right), sh*Msl1*- (bottom left) and sh*Msl2*-treated (bottom right) female ESCs. Shown here are additional examples of RNA-FISH for multicellular colonies and individual cells exhibiting *Xist*-mediated coating (see Figure 7A). White boxes indicate cells enlarged in Figure 7A. White arrows denote *Huwe1* and *Xist* foci. Dashed lines indicate nuclei borders. For all experiments, nuclei were counterstained with DAPI (blue).

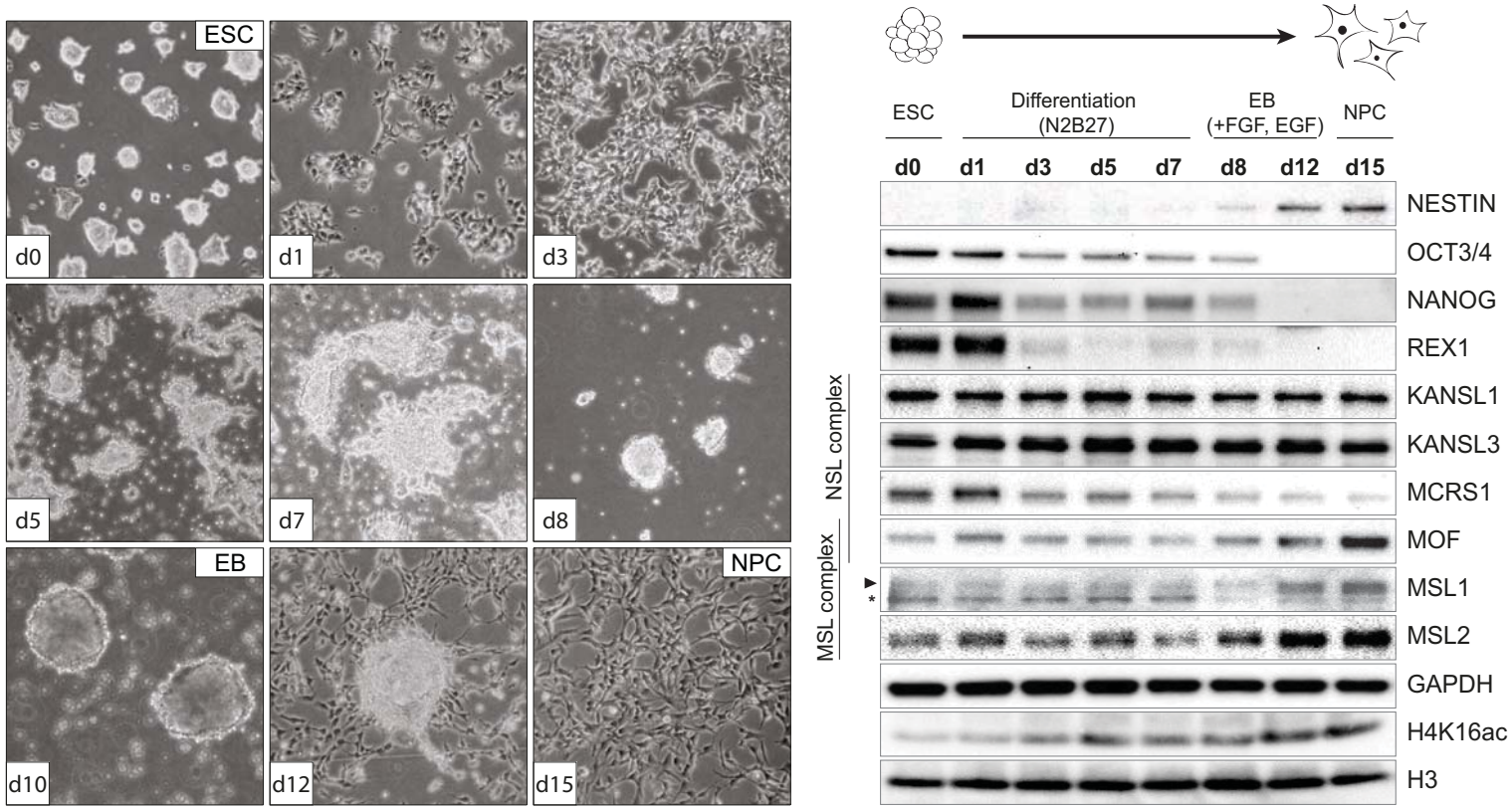
(B) Summary of RNA-FISH for *Xist* and *Huwe1*. The number of green dots indicates the number of X chromosomes within the cell while the larger dot indicates *Xist* accumulation. Cells were classified into three phenotypic groups with cells showing sharp, localized *Xist* signals (once or twice) or *Xist* "clouds". The percentages indicate how many cells per population showed the respective phenotype.

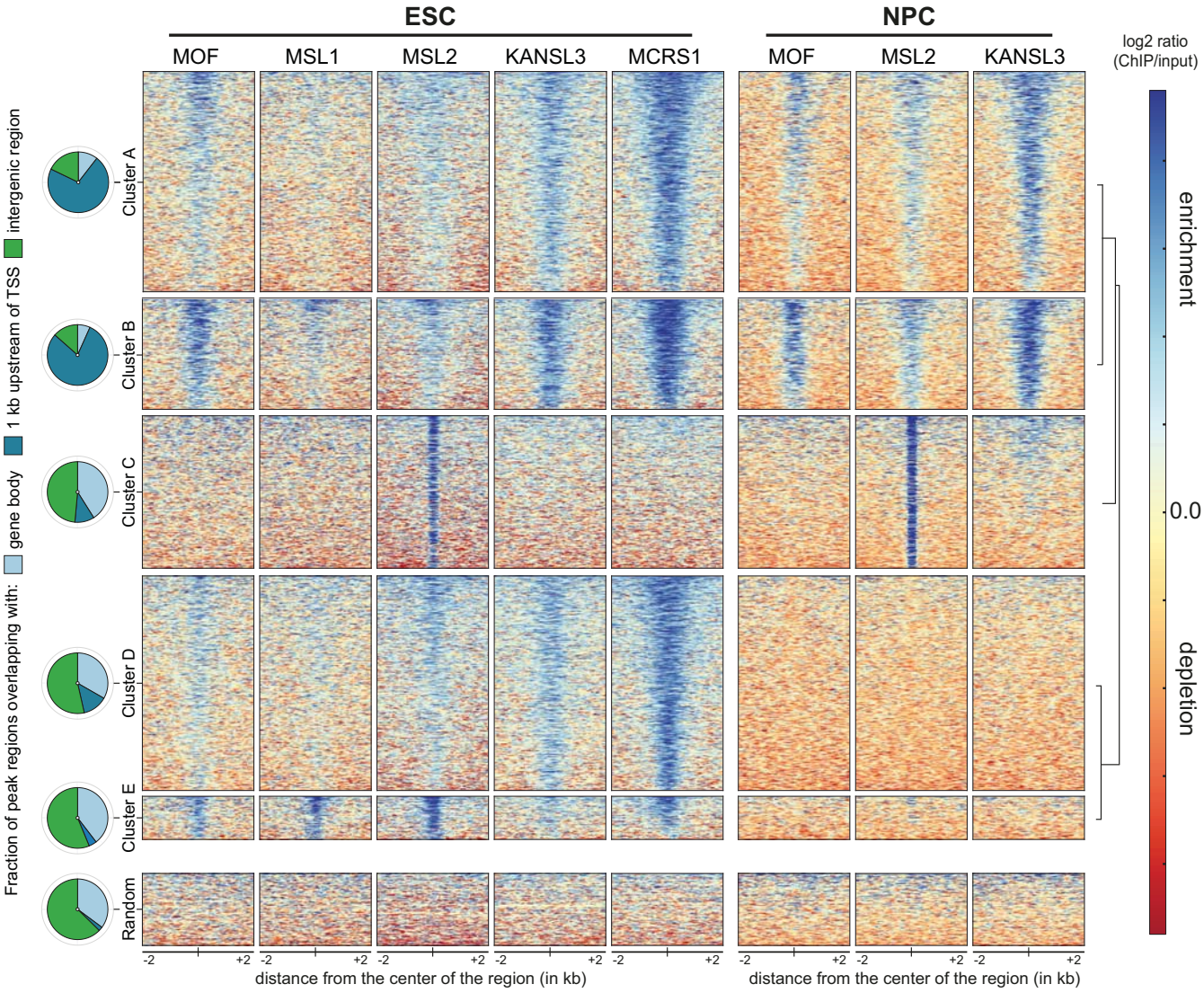
(C) Corresponding to Figure 7A. Summary of the total cell counts from *Xist* and *Huwe1* RNA-FISH in indicated knockdowns.

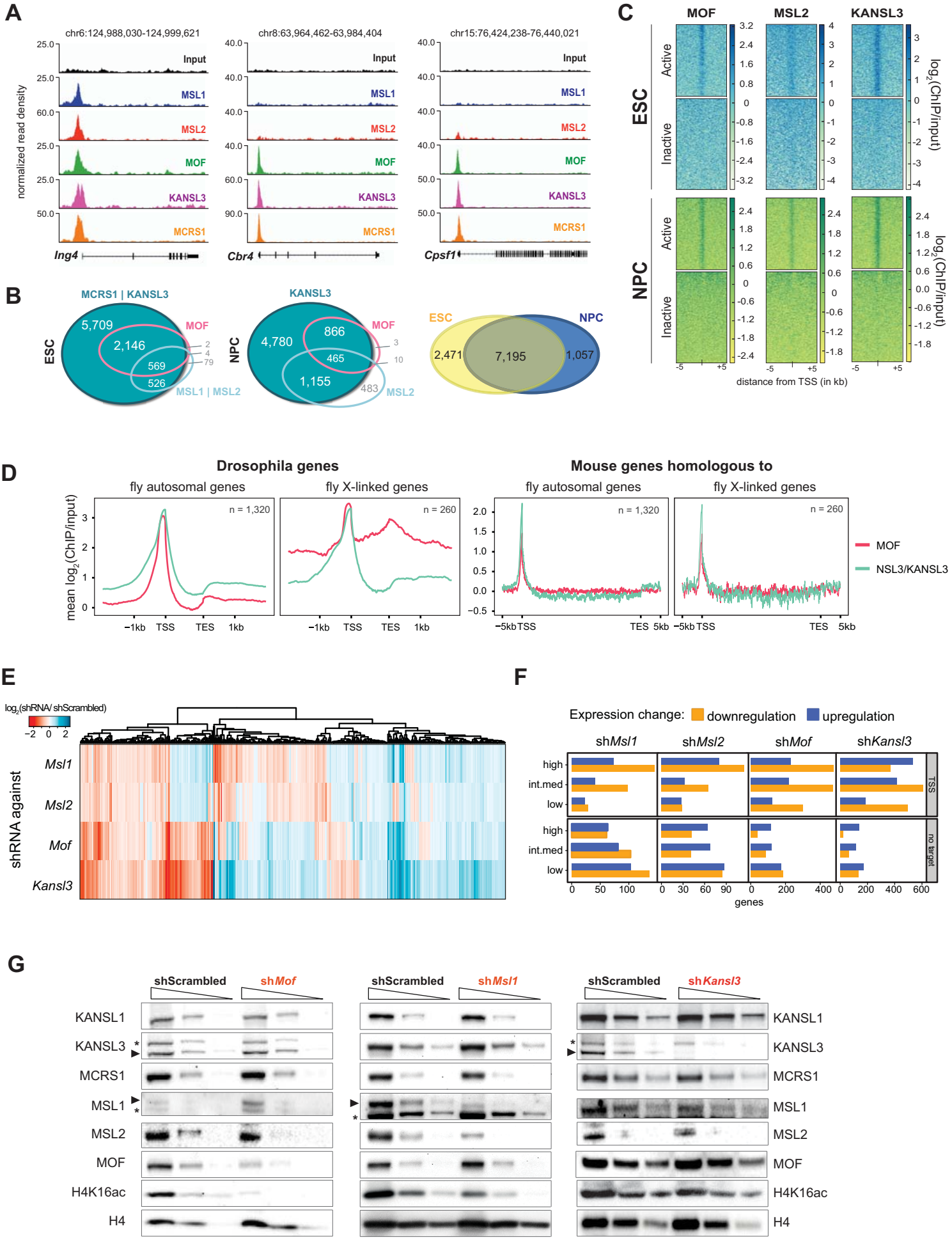
Figure 7—figure supplement 2: Depletion of MSL1 and MSL2 lead to enhanced *Xist* accumulation in differentiating ESCs

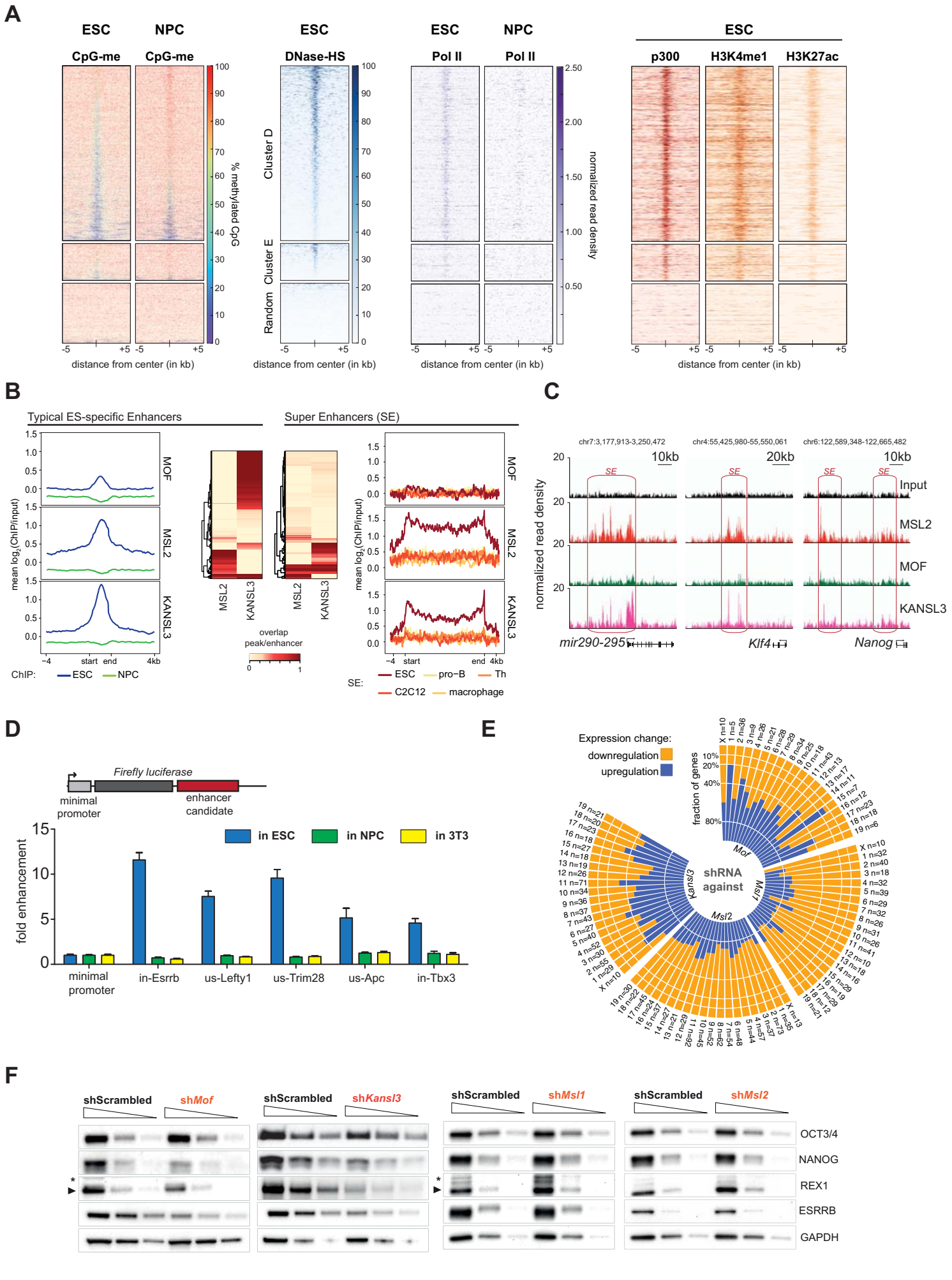
(A) RNA-FISH for *Huwe1* RNA (red) and *Xist* RNA (green) in shScrambled-, sh*Msl1*- and sh*Msl2*-treated differentiating (DAY3) female ESCs. Shown here are additional examples of RNA-FISH for multicellular colonies (see Figure 7C). Dashed lines indicate nuclei borders. For all experiments, nuclei were counterstained with DAPI (blue).

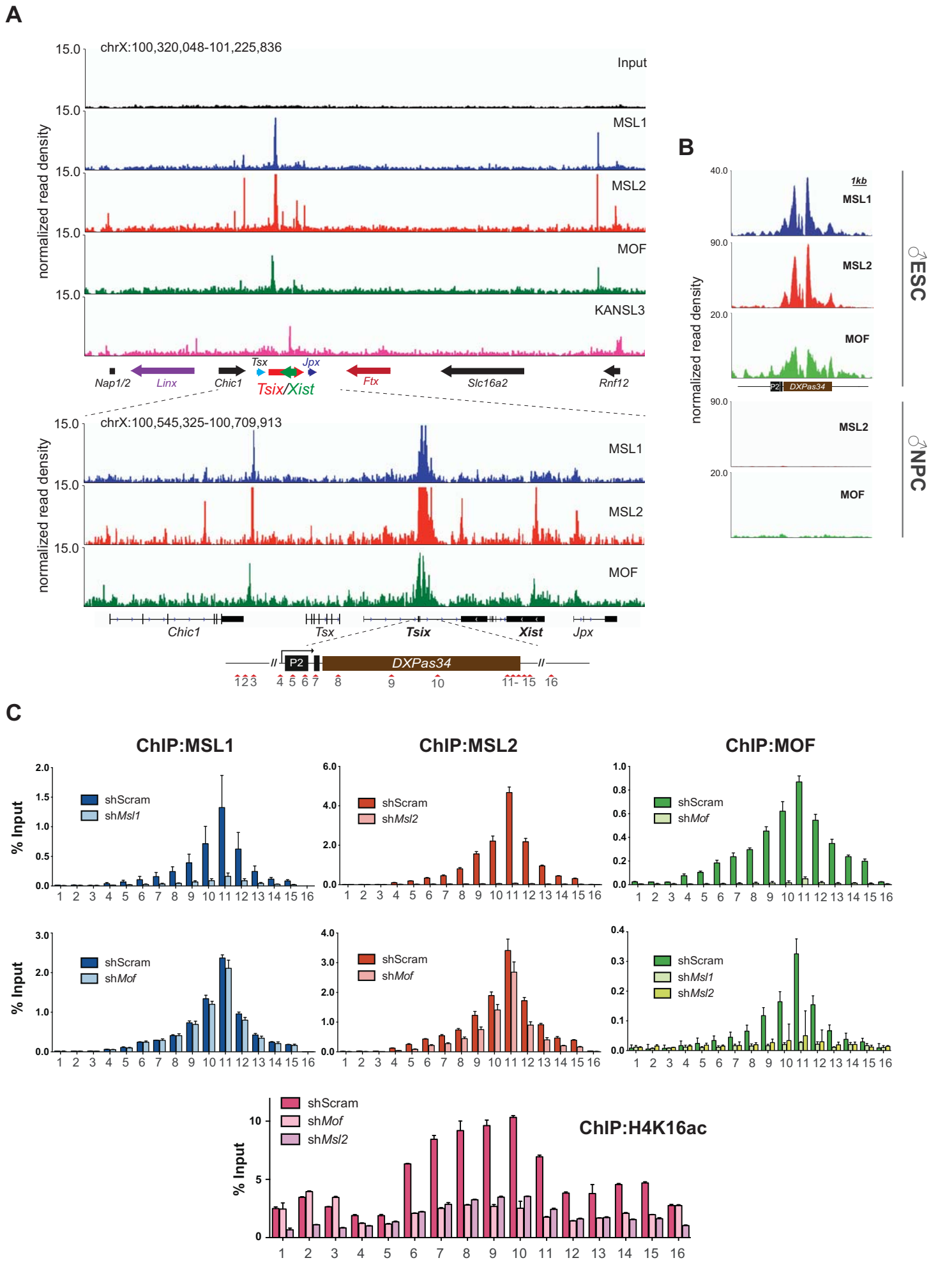
(B) Corresponding to Figure 7C-E. Summary of the total cell counts from *Xist* RNA-FISH in indicated knockdowns. Percentage of cells with respective phenotype indicated in red.

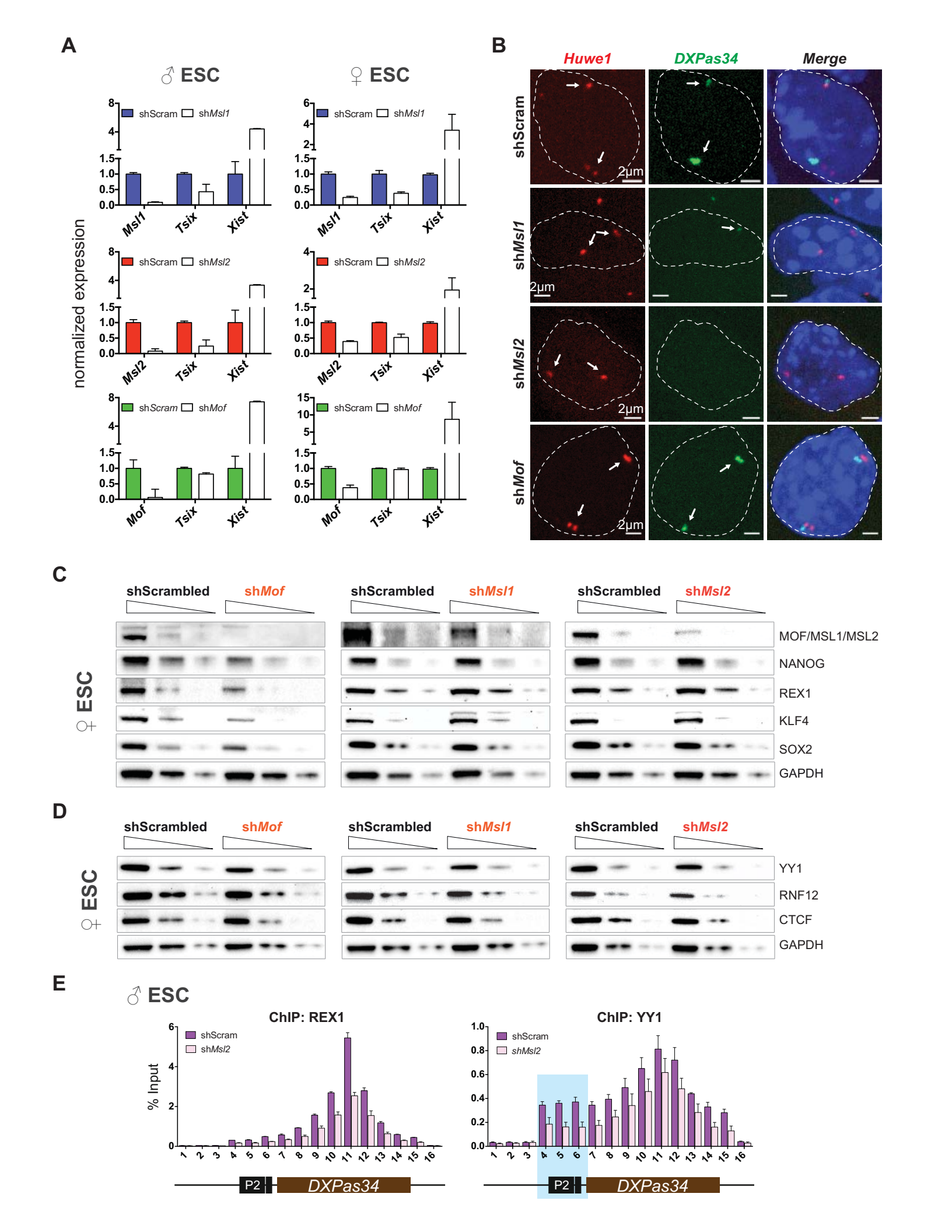

A
ESC
d0
d1
d3
d5
d7
d8
EB
NPC
d10
d12
d15
B
ESC
Differentiation
(N2B27)
EB
(+FGF, EGF)
NPC
d0
d1
d3
d5
d7
d8
d12
d15
NSL complex
MSL complex
NESTIN
OCT3/4
NANOG
REX1
KANSL1
KANSL3
MCRS1
MOF
MSL1
MSL2
GAPDH
H4K16ac
H3

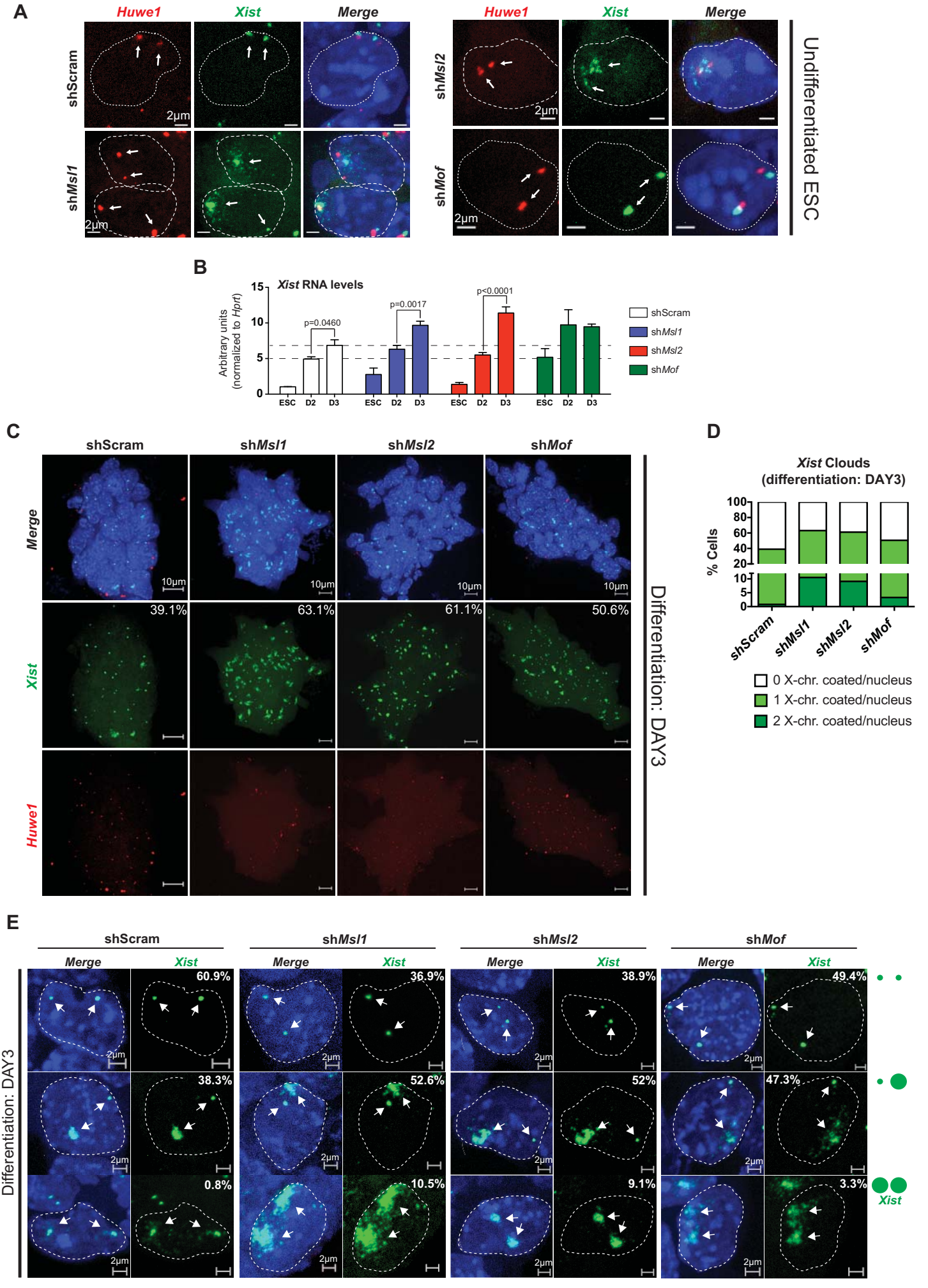

ESC
NPC
MOF
MSL1
MSL2
KANSL3
MCRS1
MOF
MSL2
KANSL3
log2 ratio
(ChIP/input)
enrichment
0.0
depletion
Fraction of peak regions overlapping with:
gene body
1 kb upstream of TSS
intergenic region
Cluster A
Cluster B
Cluster C
Cluster D
Cluster E
Random
-2
+2
distance from the center of the region (in kb)
distance from the center of the region (in kb)

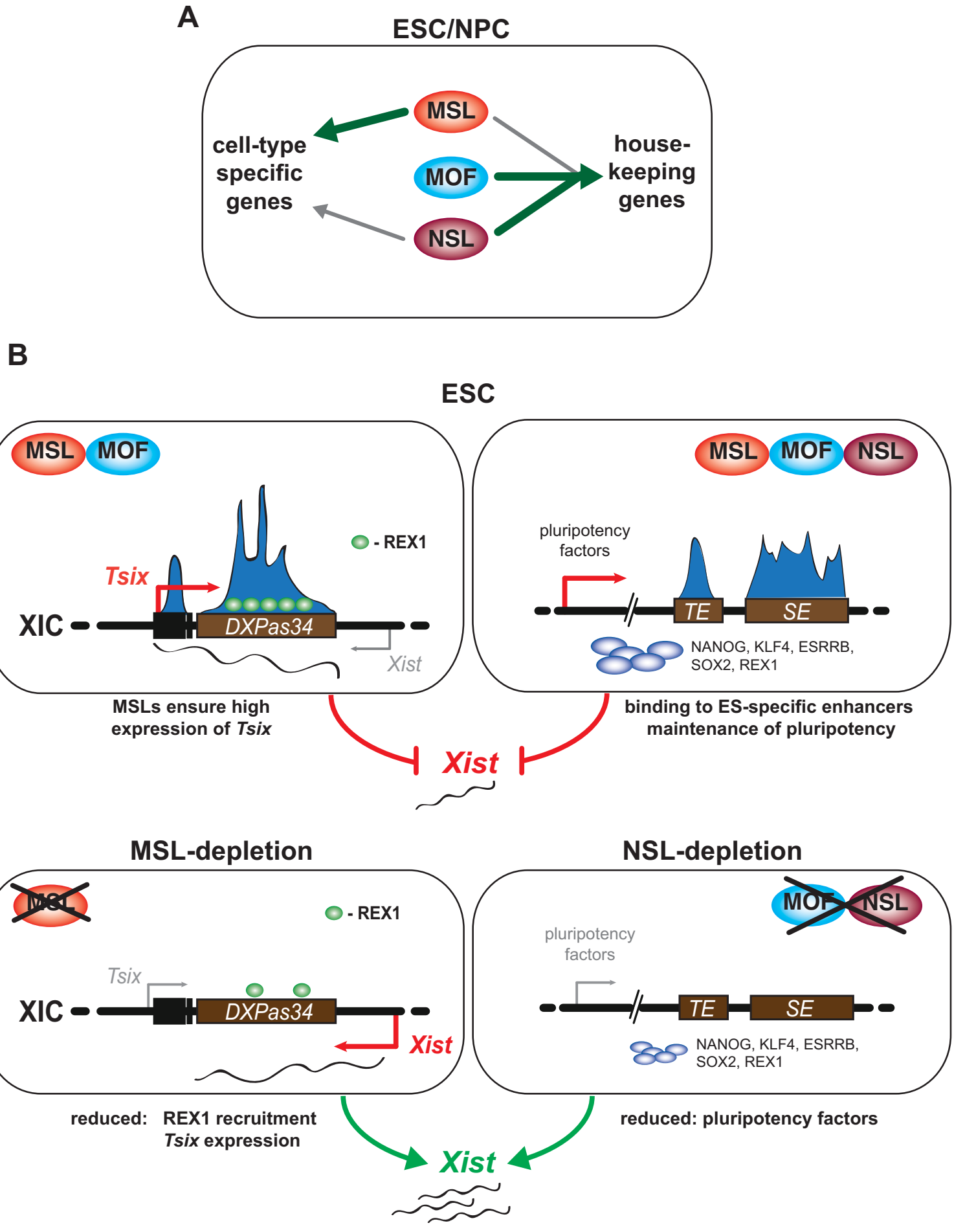

A
chr6:124,988,030-124,999,621
chr8:63,964,462-63,984,404
chr15:76,424,238-76,440,021
normalized read density
Input
MSL1
MSL2
MOF
KANSL3
MCRS1
Ing4
Cbr4
Cpsf1
B
MCRS1 | KANSL3
KANSL3
ESC
NPC
5,709
2,146
569
526
MOF
MSL1 | MSL2
4,780
866
465
1,155
483
2,471
7,195
1,057
C
MOF
MSL2
KANSL3
Active
Inactive
log2(ChIP/input)
distance from TSS (in kb)
D
Drosophila genes
Mouse genes homologous to
fly autosomal genes
fly X-linked genes
n = 1,320
n = 260
mean log2(ChIP/input)
−1kb TSS TES 1kb
−5kb TSS TES 5kb
MOF
NSL3/KANSL3
E
log2(shRNA/ shScrambled)
shRNA against
Msl1
Msl2
Mof
Kansl3
F
Expression change: downregulation upregulation
shMsl1
shMsl2
shMof
shKansl3
high
int.med
low
TSS
no target
genes
G
shScrambled
shMof
shMsl1
shKansl3
KANSL1
KANSL3
MCRS1
MSL1
MSL2
MOF
H4K16ac
H4


A
ESC
NPC
CpG-me
CpG-me
% methylated CpG
distance from center (in kb)
ESC
DNase-HS
Cluster D
Cluster E
Random
distance from center (in kb)
ESC
NPC
Pol II
Pol II
normalized read density
distance from center (in kb)
ESC
p300
H3K4me1
H3K27ac
distance from center (in kb)
B
Typical ES-specific Enhancers
Super Enhancers (SE)
MOF
MSL2
KANSL3
mean log2(ChIP/input)
overlap peak/enhancer
ChIP: ESC NPC
SE: ESC pro−B Th C2C12 macrophage
C
chr7:3,177,913-3,250,472
chr4:55,425,980-55,550,061
chr6:122,589,348-122,665,482
10kb
20kb
10kb
SE
Input
MSL2
MOF
KANSL3
normalized read density
mir290-295
Klf4
Nanog
D
Firefly luciferase
minimal promoter
enhancer candidate
in ESC
in NPC
in 3T3
fold enhancement
minimal promoter
in-Esrrb
us-Lefty1
us-Trim28
us-Apc
in-Tbx3
E
Expression change:
downregulation
upregulation
fraction of genes
shRNA against
Mof
Msl1
Msl2
Kansl3
F
shScrambled
shMof
shScrambled
shKansl3
shScrambled
shMsl1
shScrambled
shMsl2
OCT3/4
NANOG
REX1
ESRRB
GAPDH


A
chrX:100,320,048-101,225,836
Input
MSL1
MSL2
MOF
KANSL3
normalized read density
Nap1/2
Linx
Chic1
Tsx
Jpx
Tsix/Xist
Ftx
Slc16a2
Rnf12
chrX:100,545,325-100,709,913
Chic1
Tsx
Tsix
Xist
Jpx
P2
DXPas34
B
1kb
ESC
NPC
C
ChIP:MSL1
ChIP:MSL2
ChIP:MOF
shScram
shMsl1
shMsl2
shMof
% Input
ChIP:H4K16ac


A
♂ ESC
♀ ESC
normalized expression
shScram
shMsl1
shMsl2
shMof
Msl1
Msl2
Mof
Tsix
Xist
B
Huwe1
DXPas34
Merge
shScram
shMsl1
shMsl2
shMof
2μm
C
shScrambled
shMof
shMsl1
shMsl2
♂ ESC
MOF/MSL1/MSL2
NANOG
REX1
KLF4
SOX2
GAPDH
D
YY1
RNF12
CTCF
E
♂ ESC
ChIP: REX1
ChIP: YY1
% Input
P2
DXPas34


A
Huwe1
Xist
Merge
shScram
shMsl1
shMsl2
shMof
2µm
Undifferentiated ESC
B
Xist RNA levels
Arbitrary units (normalized to Hprt)
p=0.0460
p=0.0017
p<0.0001
ESC
D2
D3
shScram
shMsl1
shMsl2
shMof
C
shScram
shMsl1
shMsl2
shMof
Merge
Xist
Huwe1
10µm
39.1%
63.1%
61.1%
50.6%
Differentiation: DAY3
D
Xist Clouds
(differentiation: DAY3)
% Cells
0 X-chr. coated/nucleus
1 X-chr. coated/nucleus
2 X-chr. coated/nucleus
E
Differentiation: DAY3
60.9%
36.9%
38.9%
49.4%
38.3%
52.6%
52%
47.3%
0.8%
10.5%
9.1%
3.3%


A
ESC/NPC
MSL
MOF
NSL
cell-type specific genes
house-keeping genes
B
ESC
MSL
MOF
- REX1
Tsix
XIC
DXPas34
Xist
MSL
MOF
NSL
pluripotency factors
TE
SE
NANOG, KLF4, ESRRB, SOX2, REX1
MSLs ensure high expression of Tsix
binding to ES-specific enhancers maintenance of pluripotency
Xist
MSL-depletion
NSL-depletion
MSL
- REX1
Tsix
XIC
DXPas34
Xist
MOF
NSL
pluripotency factors
TE
SE
NANOG, KLF4, ESRRB, SOX2, REX1
reduced: REX1 recruitment Tsix expression
reduced: pluripotency factors
Xist