The cohesin regulator Stag1 promotes cell plasticity through heterochromatin regulation.

Dubravka Pezic[1], Sam Weeks[1], Wazeer Varsally[1], Pooran S. Dewari[2], Steven Pollard[2], Miguel R. Branco[3], Suzana Hadjur[1]*

1 Research Department of Cancer Biology, Cancer Institute, University College London, 72 Huntley Street, London, United Kingdom

2 MRC Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, United Kingdom

3 Blizard Institute, Barts and The London School of Medicine and Dentistry, QMUL, London, United Kingdom

* Correspondence: s.hadjur@ucl.ac.uk

SUMMARY

Fundamental biological processes such as embryo development and stem cell control rely on cellular plasticity. We present a role for the cohesin regulator, Stag1 in cellular plasticity control via heterochromatin regulation. Stag1 localises to heterochromatin domains and repetitive sequences in embryonic stem (ES) cells and contains intrinsically disordered regions in its divergent terminal ends which promote heterochromatin compaction. ES cells express Stag1 protein isoforms lacking the disordered ends and fluctuations in isoform abundance skews the cell state continuum towards increased differentiation or reprogramming. The role for Stag1 in heterochromatin condensates and nucleolar function is dependent on its unique N-terminus. Stag1$^{N\Delta}$ ESCs have decompacted chromatin and reprogram towards totipotency, exhibiting MERVL de-repression, reduced nucleolar transcription and decreased translation. Our results move beyond protein-coding gene regulation via chromatin loops into a new role for Stag1 in heterochromatin and nucleolar function and offer fresh perspectives on its contribution to cell identity and disease.

KEY WORDS

Cellular plasticity, heterochromatin, repetitive elements, nucleolus, Stag1, translation

INTRODUCTION

Cellular populations consist of mixtures of cells across a continuum of states and such heterogeneity underlies responsiveness to changing conditions. Fundamental biological processes including embryo development, stem cell control and cancer rely on conversion between states. While transcriptional states are ever-increasingly well described (Wagner et al., 2016), the mechanisms that control the stability of a given state are less well understood. The chromatin landscape with its inherent dynamics (Cho et al., 2018; Finn et al., 2019; Nozaki et al., 2017; Ricci et al., 2015) and complex 3-dimensional (3D) organization (Bonev et al., 2017; Cardozo Gizzi et al., 2019; Dekker and Mirny, 2016; Nagano et al., 2017; Schlesinger and Meshorer, 2019) can act as a mechanism to support transcriptional heterogeneity (Mateo et al., 2019; Rodriguez et al., 2019) and the chromatin proteins that regulate this landscape may play key roles in cell plasticity (Meshorer et al., 2006).

Genomes are partitioned into distinct functional domains within the nucleus (Bickmore and van Steensel, 2013; Bonev and Cavalli, 2016). A prominent feature of nuclear organization is the compact heterochromatin that accumulates at the periphery of the nucleus, around the nucleolus and at distinct foci within the nucleoplasm (Guelen et al., 2008; Németh et al., 2010; Padeken and Heun, 2014; Quinodoz et al., 2018). Heterochromatin is formed at repetitive sequences, is tightly associated with repressive histone modifications like methylation of histone H3 at lysine 9 (H3K9me2,3), and a specific set of proteins, including Heterochromatin Protein 1 (HP1) that together condense chromatin to maintain repression (Allshire and Madhani, 2018). Further, sequestration of condensed chromatin into dynamic phase-separated condensates (Larson et al., 2017; Strom et al., 2017) contributes to heterochromatin-mediated silencing. Heterochromatin organization plays a key role in cell identity and is rapidly remodelled during ES cell differentiation and embryo development. Decompaction of heterochromatin and subsequent de-repression of repetitive elements drives reprogramming towards totipotent embryos, while progressive compaction is associated with terminal differentiation (Ahmed et al., 2010; Borsos and Torres-Padilla, 2016; Martin et al., 2006; Meshorer et al., 2006; Novo et al., 2016).

Cohesin is a ubiquitously expressed, multi-subunit protein complex that has fundamental roles in cell biology including sister chromosome cohesion, 3D chromatin topology and regulation of cell identity (Cuartero et al., 2018; Horsfield et al., 2007; Kline et al., 2018; Leiserson et al., 2015; Romero-Pérez et al., 2019; Viny et al., 2019). Much of our understanding of how cohesin contributes to cell identity has been studied in the context of its roles in protein-coding gene expression and 3D organization of interphase chromatin structure (Hadjur et al., 2009; Kagey et al., 2010; Mishiro and Tsutsumi, 2009; Misulovin et al., 2007; Parelho et al., 2008; Phillips-Cremins et al., 2013; Rao et al., 2014; Vietri Rudan et al., 2015; Wendt et al., 2008). Indeed, loss of cohesin and its regulators results in a dramatic loss of chromatin topology at the level of Topologically Associated Domains (TAD) and chromatin loops, with only modest changes to gene expression (Haarhuis et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Seitan et al., 2013; Sofueva et al., 2013; Wutz et al., 2017; Zuin et al., 2014). This suggests that cohesin's roles in development and disease extend beyond gene expression regulation and highlight the need to re-evaluate how cohesin regulators shape the structure and function of the genome.

The association of cohesin with chromosomes is tightly controlled by several regulators, including the Stromalin Antigen protein (known as Stag or SA), which has been widely implicated in cell identity regulation and disease development (Cuadrado et al., 2019; Lehalle et al., 2017; Leiserson et al., 2015; Soardi et al., 2017; Viny et al., 2019; Yuan et al., 2019). Stag proteins interact with the Rad21 subunit of cohesin and mediate its association with DNA and CTCF (Hara et al., 2014; Li et al., 2020; Orgil et al., 2015; Xiao et al., 2011). Mammalian cells have three Stag paralogs, Stag1, 2 and 3. These show >90% conservation of sequence in their central domain yet perform distinct functions (Canudas and Smith, 2009; Kojic et al., 2018; Remeseiro et al., 2012a; Winters et al., 2014). It is likely that the divergent N- and C-terminal regions provide functional specificity. For example, the N-terminus of Stag1 contains a unique AT-hook (Bisht et al., 2013) which is required for its preferential participation in telomere cohesion (Canudas and Smith, 2009). The underlying mechanisms by which Stag proteins and their divergent ends influence cell identity are still largely unknown.

Here we report that Stag1 is the dominant paralogue in ES cells and supports pluripotency by regulating heterochromatin organization. We discover that ES cells regulate the level of Stag1 protein and the proportion of its divergent N and C-terminal ends, which contain disordered regions. This naturally occurring Stag1 protein heterogeneity supports a continuum of functionally distinct cellular states within the population. Changing the balance in the levels of Stag1 isoforms leads to conversion between cell states, with the loss of the N-terminus favoring a reprogrammed, totipotent state and decompaction of heterochromatin condensates and the loss of the C-terminus priming cells towards exit from pluripotency through gene expression deregulation. These results define specialised, non-redundant roles for the divergent ends. Mechanistically, the N-terminus of Stag1 represses reprogramming to totipotent 2-cell-like (2C-L) cells by maintaining nucleolar structure and function. We uncover Nucleolin and Trim28 as direct interactors of Stag1 and show that cells selectively expressing Stag1 isoforms lacking the N-terminal AT-hook domain exhibit reduced nascent nucleolar transcription and a decrease in global translation. Our results take us beyond protein coding gene regulation via chromatin loops into a new role for Stag1 in the regulation of heterochromatin and nucleolar structure and function. Importantly, by identifying changes to translation control upon Stag1 loss in stem cells, we open a new perspective by which Stag proteins and cohesin regulation can impact cell identity and disease.

RESULTS

A functional change in cohesin regulation in cells of different potential.

We analyzed the expression levels of cohesin regulators in embryonic stem cells (ES) at different stages of pluripotency. During the transition between naïve (2i) and primed (EpiLC) pluripotency *in vitro*, levels of the core cohesin subunits Smc1 and Smc3 do not change, while Stag1 becomes downregulated and Stag2 becomes upregulated (Figures 1A, S1A, B). This was confirmed at the protein level, where we observe a 2-3-fold higher level of chromatin-associated Stag1 compared to Stag2 in naïve ES cells, while Stag2 levels are 5-10-fold higher in EpiLCs (Figures 1B, S1C). These results, together with similar observations (Cuadrado et al., 2019), identify Stag1 as the dominant paralog in naïve ES cells and suggest that a switch between Stag1 and Stag2 may represent a functionally relevant change in cohesin regulation at different stages of pluripotency.

Stag1 supports the naïve pluripotent state.

To investigate the functional importance of Stag1 in the regulation of pluripotency, we established a Stag1 RNA knockdown (KD, 'siSA1-SP', Methods) strategy using siRNAs. This resulted in a significant reduction of Stag1 at the mRNA and protein levels (4-5-fold, 8-10-fold, respectively), in both serum-grown (FCS) and naïve (2i) ES cells without affecting the cell cycle (Figures 1C, S1D, E). Using Nanog as a marker of naïve pluripotency, we observed a significant downregulation of Nanog mRNA and protein levels within 24hrs of Stag1 KD in all ES populations (Figures 1D, E, S1F), suggesting that Stag1 may be functionally required for pluripotency. Indeed, a global analysis of the ES transcriptome upon siRNA-mediated Stag1 KD revealed that 375 genes were up- and 205 genes were down-regulated by at least 2-fold (Figure 1F). Among the downregulated group were several genes known to have key roles in the maintenance of pluripotency, including Nanog, Tbx3, Esrrb, Klf2, Klf4, Prdm14, Tfcp2l1, Lefty2. Notably, we did not detect a change in expression of Oct4 or Sox2, and Zfp42 was minimally affected. Moreover, we also observed both an upregulation of genes associated with exit from the pluripotent state (Dppa3, Fgf5) as well as differentiation-specific genes such as Pou3f1

(Oct6) and Sox11 (Figure 1F). Single-gene analysis revealed consistent, albeit low fold-change trends across our biological replicates, thus we used Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) to detect modest but coordinate changes in the expression of groups of functionally related genes. This revealed a robust gene signature of exit from pluripotency and enrichment for genes associated with primed pluripotency upon Stag1 KD across all biological replicates (Figures 1G, S1G).

The loss of the naïve transcriptional programme upon Stag1 KD suggests that ES cells may require Stag1 for the maintenance of self-renewal. To test this, we plated cells in self-renewal conditions at clonal density and determined the proportion of undifferentiated cells upon Stag1 KD by measuring the area occupied by the colonies with high alkaline phosphatase activity (AP+). In scrambled siRNA-treated controls, 52% of plated cells retain their naïve state, identified by AP+ colonies (Figure S1H). However, upon Stag1 KD, both the proportion of AP+ colonies and the area they occupy decreased by 20%, indicating that ES cells have a reduced self-renewal ability in the absence of Stag1 (Figures 1H, S1H). As independent validation of these results, we used CRISPR/Cas9 to knock-in an mNeonGreen-FKBP12^{F36V} tag (Nabet et al., 2018) at the C-terminus of both alleles of the endogenous Stag1 locus (SA1^{NG_FKBP}) in ES cells (Figures 1I, S1I-K). Upon dTAG addition, Stag1 protein is robustly degraded in SA1^{NG_FKBP} ES clones (Figure 1I, S1K). As was observed with siRNA treatment, dTAG-mediated degradation of Stag1 led to a 50% reduction of self-renewal potential (Figure 1J). Together, our results are consistent with a requirement for Stag1 in the control of naïve pluripotency and provide an opportunity to discover the mechanisms of Stag1 actions.

STAG1 localizes to AT-rich heterochromatin.

To begin to understand how STAG1 contributes to pluripotency, we investigated the subcellular localization of endogenous STAG1. Live cell imaging of Hoechst-labelled Stag1^{NG_FKBP} ES cells revealed the expected and predominant localisation of STAG1 in

the nucleus (Figures 2A, B). Interestingly, STAG1 was not uniformly distributed within the nucleoplasm. In addition to a dispersed nucleoplasmic localisation pattern, we observed STAG1 colocalization with Hoechst-dense regions. These included colocalization at large Hoechst-dense foci (Figure 2A, top cell), within the interior of the nucleolus (Figure 2A, top cell) and at the periphery of the nucleolus (Figure 2A, all cells). The mean intensity of STAG1 (as measured by mNeonGreen signal) was significantly enriched within Hoechst-dense foci compared to the whole nucleus, and the signal was sensitive to treatment with dTAG (Figure 2B). We made similar observations of STAG1 localization at DAPI-dense foci in cells expressing Dox-inducible GFP-tagged full-length Stag1 ($SA1^{FL}$-GFP) (Figure S2A, B). We note that repressive heterochromatin domains are readily observed by staining with AT-rich DNA dyes (ie. DAPI and Hoechst) and are organized around the nucleolus, in discreet foci within the nucleoplasm, or tethered to the nuclear periphery (Padeken and Heun, 2014; Quinodoz et al., 2018). Thus, the profile of STAG1 within ES cells is consistent with its localization to AT-rich heterochromatin. Given the presence of an AT-hook within STAG1 and the importance of heterochromatin regulation in development, we investigated whether STAG1 may have a role in heterochromatin structure.

STAG1 interacts with heterochromatin proteins and repetitive DNA.

Since STAG1 was localised to nuclear heterochromatin domains, we investigated whether it was also interacting with heterochromatin proteins and bound to genomic sequences known to form heterochromatin, such as repeats. Constitutive heterochromatin is characterized by the binding of HP1α to H3K9me2/) and plays a critical role in silencing of repetitive DNA elements (Allshire and Madhani, 2018) and nuclear organization (Larson et al., 2017). The periphery of the nucleolus accumulates marks of constitutive heterochromatin coincident with transcriptionally inactive rDNA repeats. Nucleolin is a major nucleolar protein which controls the organization of nucleolar chromatin, rDNA transcription and ribosome assembly and plays important roles during development and in ES cells (Kresoja-Rakic and Santoro, 2019; Percharde et al., 2018). We observed nuclear colocalization between STAG1 and HP1α in dox-

186 induced SA1^{FL}-GFP cells around DAPI-dense foci (Figures S2A, B). Further, using
187 chromatin coImmunoprecipitation (coIP), we show that STAG1 interacts with both HP1α
188 and Nucleolin in ES cells (Figure 2C, S2C).
189 Previous studies have analysed STAG1 binding profiles in mouse ES cells and
190 have primarily focused on its association with protein-coding genes (Cuadrado et al.,
191 2019). A thorough investigation of STAG1 binding to repetitive sequences has not been
192 conducted. Thus, we re-analysed STAG1 chromatin immunoprecipitation followed by
193 sequencing (ChIP-seq) experiments to calculate the proportion of STAG1 peaks that
194 overlapped genes (based on promoter and exon features), repeats (within the Repeat
195 Masker annotation) introns and intergenic regions not already represented by repeats or
196 genes. Of the 18,600 STAG1 peaks identified, the majority (76%) are bound to genomic
197 elements that are distinct from protein-coding genes with a significant proportion of
198 binding sites at repetitive elements and intergenic regions (Figure 2D). Together with the
199 localizations observed by microscopy, this suggests that the role of STAG1 in ES cells
200 may extend beyond protein-coding gene regulation. While STAG1 binding is not enriched
201 compared to all genomic repeats, we asked whether specific repeat families might be
202 enriched for STAG1 binding above random expectation (Deniz et al., 2020). We found
203 several repeat families to be significantly enriched for STAG1 peaks, including those
204 within the DNA transposon and Retrotransposon classes, both known to form constitutive
205 heterochromatin. Specifically, STAG1 was enriched at SINE B2-Mm2, (previously shown
206 to be enriched at TAD borders (Dixon et al., 2012)) and B3 elements, LINE1 elements
207 (L1Tf, L1A), and several LTR families, two of which have been previously shown to be
208 associated with CTCF (LTR41, LTR55) (Schwalie et al., 2013) (Figures 2E and S2D). The
209 enrichment of Stag1 at non-coding and repetitive sequences has not been previously
210 described and points to a novel role for STAG1 in the regulation of repeats in ES cells.
211
212 **Stag1 regulates nuclear organization of heterochromatin.**
213 Since H3K9me3 is the defining histone modification of silent heterochromatin, we
214 assessed the impact of STAG1 loss and overexpression on H3K9me3 in ES cells. While
215 global levels of H3K9me3 were unchanged upon Stag1 KD (Figure S2E),

immunofluorescence (IF) of H3K9me3 revealed changes to heterochromatin organization. H3K9me3 foci displayed a greater variation in volume compared to scrambled control treated cells, suggesting that Stag1 is required for proper H3K9me3 compaction (Figures 2F, G). We detected similarly variable changes to global chromatin accessibility, as measured by DNase I digestion. In four out of six experiments, ES cells treated with Stag1 siRNAs revealed a tendency towards increased accessibility, whereas two experiments showed increased compaction (Figure 2H, S2F). On the other hand, Dox-inducible STAG1 expression led to a dramatic condensation of H3K9me3 into large nuclear puncta compared to non-induced cells (Figures 2I, J), where STAG1 could also be seen to be colocalized with H3K9me3 in the condensate. This phenotype led us to investigate whether STAG1 may have characteristics found in other proteins known to play a role in heterochromatin phase separation(Larson et al., 2017). We used the PONDR tool (Obradovic et al., 2003) to assess potential intrinsically disordered regions (IDR) (Banani et al., 2017) within STAG1. STAG1 has an overall PONDR score of 0.4397, and both the N-and C-terminal divergent regions contain sequences with a high propensity for intrinsic disorder (Figure 2K). Interestingly, these coincide with known STAG1 domains, most notably the N-terminal AT-hook. Together our results uncover a novel role for STAG1 in forming or maintaining heterochromatin structures in ES cells and suggest that the terminal ends may play important roles therein.

Stag1 expression is highly regulated in ES cells.

STAG1 levels are highest in naïve 2i-grown and lower in FCS-grown ES cells, a culture condition that supports a mix of naïve and primed cells (Figure S1B). This prompted us to investigate whether STAG1 is regulated at the transcriptional level. To address this, we employed a series of approaches to comprehensively characterize Stag1 mRNAs and discovered widespread regulation of Stag1 transcription in ES cells. First, we used RACE (Rapid Amplification of cDNA Ends) to characterize the starts and ends of Stag1 mRNAs. Using 5' RACE, we uncovered four novel alternative transcription start sites (TSS) in ES cells; one located 50kb upstream of the canonical Stag1 TSS (referred to as 'SATS', and previously identified in (Feng et al., 2016)) (Figures 3A, D, S3A), one between canonical

exon 1 and exon 2 (referred to as alternative exon 1 or altex1) (Figure 3D), one at exon 6, and one at exon 7 (Figures 3B, D, S3A, B (for increased exposure). Interestingly, the novel TSS located at exon 7 (e7) was preceded by a sequence located *in trans* to the Stag1 gene, carrying simple repeats and transcription factor binding sites (Figure S3C). While the frequency of this alternative TSS was significantly lower than the other TSSs, it was identified in multiple RACE replicates, indicating it may have important functions in a subset of the ES population. We also discovered widespread alternative splicing in the 5' region of Stag1, with particularly frequent skipping of exons 2 and 3 (e2/3Δ) and exon 5 (e5Δ) (Figures 3D, S3A). Using 3' RACE, we detected an early termination site in intron 25 and inclusion of an alternative exon 22 introducing an early STOP codon, as well as several 3'UTRs (Figures 3C, D, S3D).

Next, PCR- and Sanger sequencing-based clonal screening confirmed that the newly discovered 5' and 3' ends represent true Stag1 transcript ends, validated the existence of the e2/3Δ and e5Δ isoforms, and uncovered an isoform lacking exon 31 (e31Δ) (Figures 3D, S3E). To determine the complete sequences of the Stag1 transcript isoforms and to use a non-PCR-based approach, we performed long-read PacBio Iso-seq from 2i ES RNA. This confirmed the diversity of the Stag1 5' and 3'UTRs, the e31Δ isoform, multiple TSSs, including SATS, and early termination events, including in i22 and i25 (Figure S3F). Importantly, these transcripts all had polyA tails, in support of their protein-coding potential. Finally, we validated and quantified the newly discovered splicing events by calculating the frequency (percentage spliced in (PSI)) of exon splicing in our and published RNA-seq data using VAST-tools, a previously developed computational method (Tapial et al., 2017) (Figure 3E, Table S2). Together, these results point to a previously unappreciated diversity of endogenous Stag1 transcripts in ES cells, highlighting the importance of Stag1 regulation in stem cell populations.

The chromatin landscape reflects Stag1 transcriptional regulation.

Visual inspection of the genome topology around the *Stag1* locus in our existing 2i ES and neural stem (NS) cell Hi-C datasets (Barrington et al.) revealed that the *STAG1* gene undergoes significant 3-D reorganization as cells differentiate (Figure 3F). First, the entire

STAG1 TAD switches from the active to the repressive compartment during differentiation, in line with the decrease in Stag1 levels described above. Furthermore, we observed several changes to the sub-TAD architecture which corresponded to the newly discovered Stag1 TSSs and TTSs described above (Figure 3F, compare 'SA1 transcripts' track with Hi-C maps). We quantified these changes by designing site-specific UMI-4C baits guided by the Hi-C topology. In naïve ES cells, the genomic region containing the alternative SATS TSS has several Nanog binding sites and makes numerous contacts with the *Stag1* gene body (Figure 3F, 'Nanog bait 1'). Meanwhile, in NS cells, the SATS TSS is no longer active and becomes isolated away from the *STAG1* gene body due to the reinforcement of the TAD border at the canonical promoter of *Stag1* (Figure 3F, 'CTCF bait'). In addition, a cluster of low-occupancy CTCF binding sites within the *Stag1* gene mark an ES-specific contact insulation point giving rise to a sub-TAD. We note that the observed Stag1 early termination events align with these CTCF sites. These results suggest that 3D chromatin topology may play a direct role in facilitating the transcriptional output of *Stag1*.

Multiple Stag1 protein isoforms are expressed in ES cells.

Stag1 transcript diversity was intriguing because many of the events were either specific to ES cells or enriched compared to MEFs and NSCs (Figure 3A, E, S3E) and further, because the transcript variants are predicted to produce STAG1 protein isoforms with distinct structural features and molecular weights (Figure 3D, G). For example, the truncation of the N terminus (e2/3Δ, e5Δ, e6 TSS and e7 TSS), and thus loss of the AT hook (amino acid 3-58), could impact STAG1 association with DNA. Meanwhile, C-terminal truncated Stag1 isoforms (altex22, i25 end, e31Δ) could affect STAG1-cohesin interactions. Interestingly, the evolutionarily conserved Stag-domain ('SCD', AA 296-381) (Orgil et al., 2015), shown to play a role in CTCF interaction (Li et al., 2020), would be retained in all the transcripts identified here. Importantly, these events would yield STAG1 isoforms lacking either the N- or C-terminal disordered regions and could thus impact the ability of STAG1 to form condensate-like structures (Figure 3H).

Immunoprecipitation (IP) of endogenous STAG1 revealed multiple bands corresponding to the predicted molecular weights for several protein isoforms and identified by mass spectrometry to contain Stag1 peptides (Figure 3I and Table S3), whose levels were reduced between naïve and primed cells (Figure S3G) and sensitive to Stag1 KD, alongside the canonical, full-length isoform (Figure 3J). Treatment of Stag1^{NG_FKBP} ES cells with dTAG followed by Western blotting of chromatin-associated proteins with an antibody to the v5 tag further confirmed the sensitivity of the isoforms to dTAG-mediated degradation, validating the presence of N-terminally truncated STAG1 protein variants (Figure 3K). Overall, our results indicate that complex transcriptional regulation gives rise to multiple Stag1 transcripts and protein isoforms with distinct regulatory regions and coding potential, the majority of which are expressed specifically in naïve ESCs. Our discovery of such naturally occurring STAG1 isoforms highlights the importance of STAG1 in ES cells and offers a unique opportunity to define the ES-specific functions of the divergent N- and C-terminal ends of STAG1 in the context of pluripotency gene expression and heterochromatin regulation.

Skewing the abundance of Stag1 isoforms promotes transitions between cell states.

To study the functional consequences of STAG1 isoform expression changes on ES cells, we took advantage of our detailed understanding of Stag1 transcript diversity to design custom siRNA pools to selectively target, or retain, specific isoforms (Figure 4A). Alongside the siRNAs from Figure 1 (SmartPool, SP), we designed siRNAs to specifically target the SATS 5'UTR (esiSATS), the 5' end (siSA1-5p) or the 3' end (siSA1-3p) of Stag1 mRNA (see Methods). We anticipated that the KD panels would not completely abolish all Stag1 transcript variants, but rather change the relative proportions, in effect skewing the levels of the N- and C-terminal ends of Stag1. 3p siRNAs were expected to downregulate full-length and N-term truncated isoforms and retain C-term truncated isoforms. Meanwhile 5p siRNAs would specifically retain N-term truncated isoforms.

siRNAs to the 5p and 3p ends of Stag1 reduce full-length Stag1 mRNA and protein with similar efficiency to SP KDs (Figures 4B, C, S4A), while esiSATS reduces Stag1 by

50-60%, indicating that the SATS TSS functions to enhance expression of Stag1 specifically in naïve ES cells. We confirmed that Stag1 isoform proportions were altered upon siRNA treatment using RNA-seq, RACE and immunoprecipitation. RNA-seq reads aligning to Stag1 in the different siRNA treatments were quantified to represent the residual N-terminal, middle and C-terminal read proportions (Figure 4D). Residual reads in the SP and 3p KDs aligned primarily to the N-terminus and were depleted from the C-term. While the 5p KD had the least read retention in the N-terminus, supporting the expectation that residual transcripts in the 5p KD have full C-terminal ends (Figure 4D). These results were further supported by quantifying specific splicing events from the KD RNA-seq datasets using VAST-tools (Figure S4B). In parallel, RACE was used to validate changes to the proportions of Stag1 isoforms. 5' RACE in ES cells treated with 5p siRNA revealed downregulation of full-length Stag1 transcript while several N-terminal truncated isoforms were upregulated compared to untreated cells (Figures 4E, S3A). Similarly, transcripts terminating at the canonical 3' end of Stag1 are strongly reduced in the SP and 3p siRNA KD samples and to a lesser extent in the 5p KD, while the transcript terminating in i25 is substantially enriched upon 3p KD (Figure 4E). Immunoprecipitation of STAG1 using an antibody which recognizes the N-terminus in Stag1^{NG_FKBP} ES cells treated with dTAG reveals the enrichment of STAG1-CtermΔ isoforms (Figure 4F, arrows). Thus, the siRNA panel used here provide us with a powerful tool to modulate the proportion of the divergent ends of STAG1 in ES cells and study their potential roles in cell fate regulation.

The C-terminus of STAG1 has a specific role in maintenance of the naïve pluripotent state.

We analyzed pluripotency gene expression in the different KDs using RT-qPCR and RNA-seq. RT-qPCR results suggested that 5p and 3p KDs may have differential effects on Nanog expression in serum-grown ES cells (Figure S4C). Namely, there was a consistent tendency towards Nanog downregulation in 3p KD while 5p KD had little effect on Nanog.

GSEA supported the differential effect of the SATS, 3p and 5p KDs on naïve and primed pluripotency signatures. In support of STAG1 playing a role in pluripotency, reducing Stag1 levels by targeting the ES-specific SATS promoter leads to downregulation of the naïve pluripotency gene signature and upregulation of the primed signature (Figure 4G), reminiscent of the phenotype from SP KD (Figure 1G). A similar, more prominent loss of the naïve signature was observed in 3p KD, while the opposite was true for 5p KD cells where the naïve signature was maintained (Figure 4G).

The distinct gene expression signatures of the 3p and 5p KDs are reflected in differences in cellular phenotypes. Cells treated with 3p siRNAs exhibited a further loss of self-renewal potential, consistent with the loss of the naïve pluripotency signature, with 20% of colonies exhibiting AP-staining compared to 30% of colonies in the SP KDs, and reduction of the area occupied by AP+ colonies by 40-60% (Figures 4H, S4D). This was not evident in the 5p KD, where the effect on self-renewal was similar to scrambled control treated cells (Figures 4H, S4D). Interestingly, unlike siRNA to Stag1, esiSATS results in a variable effect on self-renewal (ranging from between 5-35% reduction in AP+ area) (Figure 4H), likely because the SATS TSS is expressed in the most naïve cells of the population, the frequency of which varies significantly between FCS populations. Our results further confirm the importance of STAG1 in self-renewal and point to a specific role for the C-terminal of Stag1 in the maintenance of the naïve pluripotent state.

The N-terminus of Stag1 regulates heterochromatin and conversion to totipotent state.

Both the N-termΔ and C-termΔ STAG1 isoforms lost prominent IDRs (Figure 3H), thus we investigated the effect of these on heterochromatin structure. ES cells treated with 5p siRNAs exhibited variable H3K9me3 foci volumes, similar to the effect of SP KD (Figure 2G), implicating the N-terminus in H3K9me3 condensation (Figures S5A, B). On the other hand, H3K9me3 foci volumes in 3p KD cells were not significantly different from scr control cells. DNase I digestion of chromatin further confirmed global changes to chromatin accessibility upon 5p, but not 3p, KD whereby four of six experiments revealed chromatin decompaction (Figure S5C). We built up on these results by generating ES

cell lines expressing dox-inducible SA1$^{e5\Delta}$-GFP or SA1$^{i25\text{-}\Delta C}$-GFP, as representative of the N-terminal, AT-hookΔ and the C-termΔ groups respectively. H3K9me3 foci in SA1$^{i25\text{-}\Delta C}$-GFP-expressing cells revealed increased condensation compared to control cells, with a similar effect as SA1^{FL}-GFP. Meanwhile, the condensation of H3K9me3 foci in SA1$^{e5\Delta}$-GFP-expressing cells was significantly attenuated from SA1^{FL}-GFP cells (Figures 5A, B). Thus, while both the N- and C-termini of STAG1 play a role in heterochromatin structure, loss of the basic N-terminus (which interacts with DNA via the AT-hook region(Bisht et al., 2013)), significantly affects condensate structure, indicating its importance in heterochromatin compaction.

Since STAG1 was bound to repeats, which play significant roles in ES cell fate determination, and are involved in heterochromatin regulation, we also profiled the expression of LINE-1, IAPEz and MERVL elements in our KD panel. Surprisingly, MERVL elements were significantly derepressed in 5p KD, but not in SP or 3p KD conditions (Figures 5C and S5D). The effect was specific to MERVL, we saw no significant change to the expression of LINE1-T or IAPEz (Figure 5C), despite the enrichment of STAG1 at L1T elements (Figure 2D). Re-activation of MERVL is a hallmark of a rare subpopulation of totipotent cells (termed two-cell-like, 2C-L) which spontaneously arise in ES cell cultures and exhibit unique molecular and transcriptional features including chromatin decompaction, reminiscent of our 5p KD (Eckersley-Maslin et al., 2016; Ishiuchi et al., 2015; Macfarlan et al., 2012). Indeed, changes to MERVL expression in the 5p KD was accompanied by increased expression of additional hallmark 2C-L genes and chimeric transcripts (Dux, Gm6763, AW822073 and Gm4981) in serum-grown and 2i cells (Figures 5D, S5E). We note that the effect of 5p KD on 2C-L genes is modest in serum-grown ES cells and significantly stronger in naïve ES cells. Notably, all 2C-L genes analyzed remained unchanged in 3p KD conditions with a modest but insignificant upregulation in SP KD. Despite the modest but coordinated changes we observed by qRT-PCR of 2C-L genes (Figure 5D), GSEA using a published 2C gene set (Percharde et al., 2018) revealed a specific enrichment among the upregulated genes in 5p KDs that was not

observed in 3p KDs (Figure 5E, S5F). Together our results point to a specific role for the N-terminus of STAG1 in totipotency regulation.

To investigate this further, we asked whether the loss of the N-terminus of STAG1 drives conversion of ES cells towards a totipotent 2C-L state. We obtained ES cells expressing a dox-inducible *Dux-HA*-expression construct together with a MERVL-linked GFP reporter (Hendrickson et al., 2017) and used flow cytometry to measure the number of GFP-positive cells in our different Stag1 KD conditions (Figures 5F, G). Chaf1 is a chromatin accessibility factor previously shown to support conversion of ES cells towards totipotency (Ishiuchi et al., 2015). In support of the upregulation of the 2C gene set in 5p KD cells, we observed an 8-9% increase in the proportion of GFP-positive cells in 5p KD conditions compared to scramble treated controls, similar to the effect of Chaf1 KD (Figures 5F, G). There was a modest increase in GFP+ cells upon SP KD and no effect upon 3p KD. Chaf1 and 5p double KD had an additive effect on the proportion of GFP-positive cells, suggesting that the two proteins function in complementary pathways for conversion towards totipotency.

A 2C-specific Stag1 promoter.

Our results point to the N-terminus of Stag1 having a protective function in the conversion of ES cells towards totipotency. Given the known role for increased chromatin accessibility during conversion to 2C (Ishiuchi et al., 2015) and the importance of the N-terminus in heterochromatin organization reported here, we propose that the Stag1 N-terminus, and specifically the AT-hook within it, is involved in heterochromatin silencing by forming or maintaining a condensed heterochromatin structure. This predicts that the 2C-L cells which spontaneously arise within the ES population could express Stag1NΔ protein isoforms to support their preferred chromatin status and reinforce their cell state. To formally address this, we induced *DuxHA*-expression in the MERVL-GFP ES cells and performed 5' RACE on sorted GFP+ and GFP- cells (Figure 5H). We enriched several of the previously identified N-term truncated Stag1 transcripts in the GFP+ population

449 including e2/3Δ and e5Δ isoforms. Importantly, we also identified a transcript starting at
450 e7, similar to the one previously found in 5p KD ES cells (Figure 3B, S3C). Remarkably
451 however, the sequence preceding the TSS in e7 in *Dux*-induced cells was an MT2-
452 MERVL element, creating a chimeric, LTR-driven transcript akin to others specifically
453 expressed in the 2C-L state (Figures 5H, I). Thus, the 2C-L state selectively expresses
454 an N-term truncated Stag1 isoform which in turn supports the maintenance or emergence
455 of that state.

456

457 **Stag1 regulates 2C fate via changes to nucleolar structure and function.**

458 As ES cells preferentially expressing STAG1NΔ isoforms led to conversion towards
459 totipotent cell states and STAG1 is localised to the nucleolus and interacting with
460 Nucleolin, we asked whether the STAG1-induced 2C-L state could be explained by the
461 association of STAG1 with the Nucleolin/Trim28 complex known to derepress Dux targets
462 (Percharde et al., 2018). While STAG1 proteins directly interact with Nucleolin as well as
463 Trim28 (Figures 5J), Dux itself is weakly or variably deregulated in Stag1 KDs (Figure
464 5D). Similarly, while MERVL is robustly derepressed upon 5p KD (Figures 5C, F), Stag1
465 protein is not enriched at MERVL elements (Figure 2D), suggesting that the observed
466 effects are unlikely to be via *Dux*-mediated derepression of targets and may be indirect.
467 Thus, we investigated whether the Stag1-induced 2C-L state could be explained by
468 changes to nucleolar structure or function. In this context, we examined the consensus
469 sequence of the rDNA locus and note that there are several SINE elements located within
470 the intergenic spacer (IGS). Analysis of Stag1 ChIP-seq alignments to this region is
471 complicated by the repetitive nature of the region, however there was evidence of
472 possible Stag1 binding to the B3 elements within the IGS, suggesting that Stag1 may
473 directly support nucleolar structure and function (Figure S6B).

474 We explored this hypothesis by investigating the effect of Stag1 loss on rRNA
475 transcription. ES cells were pulsed with 5-ethynyl uridine (EU) which becomes actively
476 incorporated into nascent RNA and enables detection of newly synthesized RNA either

spatially by immunofluorescence or globally by flow cytometry. Samples for immunofluorescence were co-stained with an antibody to Nucleolin to quantify changes in nascent rRNA transcription. Cells treated with scramble siRNA showed a distinct nucleolar structure and the EU signal could be seen throughout the nucleus, with a strong enrichment within the nucleolus as expected from rRNA expression (Figures 6A, B). A significant reduction in nascent rRNA signal was observed in all KD conditions compared to the si scr control. Interestingly, while the medians between the three siSA1 KDs were not dramatically different, the effect of the 5p KD on rRNA signal distribution was significantly different from the 3p KD, revealing different roles for the N and C-terminal ends of Stag1 in nucleolar function. In parallel, we used flow cytometry to validate the above and quantify the effect on global levels of transcription upon Stag1 loss. In support of the immunofluorescence results, we observed a reduction in nascent transcription in all siSA1 KDs compared to scrambled control (Figures 6C, D). SP and 5p KD reduced nascent transcription by 39% and 47% respectively, the bulk of which is likely from the observed change to rRNA expression. Meanwhile, cells treated with 3p siRNA exhibited a 16% reduction in global transcription compared to scrambled controls, further implying that the N- and C-terminal ends of Stag1 likely target different pools of newly synthesized transcripts.

Finally, given the changes to nascent rRNA transcription and the known role for translation in 2C-L state conversion (Eckersley-Maslin et al., 2016; Hung et al., 2013), we measured effects on translation. Indeed, previous results showed a stronger effect of Stag1 KD on Nanog protein compared to Nanog mRNA levels (Figures 1D, E). ES cells were pulsed with L-homopropargylglycine (HPG), an amino acid analog of methionine and flow cytometry was used to quantify the impact of Stag1 KD on global protein synthesis. The proportion of cells that had reduced incorporation of HPG increased significantly in the SP and 5p siRNA treated cells compared to scramble controls (32% and 35% of si Scr) (Figures 6E, F). We did observe a mild affect on global nascent translation in 3p KD treated cells (16% of si scr), although this was not significantly different from scramble controls (Figures 6E, F). Overall, our work has uncovered a novel

role for Stag1 in translation control and proposes that Stag1 may act directly at the rDNA locus or via Nucleolin to influence the regulation of rRNA expression. To our knowledge, a role for Stag1 in translation regulation has not been previously described and offers a completely new perspective on how Stag1 impacts cell fate, moving beyond its known roles in gene expression by regulating chromatin loops.

DISCUSSION

Cell fate transitions during early development are accompanied by extensive transcriptional and epigenetic reprogramming. In addition to changes in protein-coding gene expression, repetitive sequences play an integral role in developmental programmes (Jachowicz et al., 2017; Macfarlan et al., 2012; Percharde et al., 2018). Repetitive sequences are spatially clustered into heterochromatin domains to facilitate their regulation (Allshire and Madhani, 2018; Padeken and Heun, 2014). How cohesin function, or Stag proteins contribute to heterochromatin or repeat regulation has not been extensively studied in mammalian cells, despite the importance of both for cell identity.

While previous studies have implicated Stag proteins in cell fate decisions (Cuadrado et al., 2019; Viny et al., 2019), they have primarily focused on the impact to protein-coding gene regulation. By discovering a novel role for Stag1 in heterochromatin structure, repetitive elements and translation control, we significantly expand our understanding of Stag1 functions and thus deliver new insight into the mechanisms by which Stag proteins and cohesin regulation impact cell identity and lead to disease. Further, by comprehensively characterising naturally occurring Stag1 protein isoforms, we identify distinct roles for the N- and C-termini of Stag1 which shed mechanistic light on the reported distinct functions of the Stag paralogs (Cuadrado and Losada, 2020). For example, the N-terminal containing AT-hook of Stag1 represses heterochromatin condensation and its loss leads to global changes in chromatin compaction and reprogramming of ES cells towards totipotent 2C-L cells. Our observation that the AT-hook domain is required for heterochromatin compaction may explain the dominance of Stag1 in stem cell populations since Stag2 does not have an AT-hook and heterochromatin must be tightly regulated during early development.

We have characterized diverse transcription–regulatory events at the Stag1 locus in stem cell populations giving rise to protein isoforms with distinct regulatory domains. Importantly, this extensive Stag1 protein heterogeneity supports a continuum of cell states within the population since experimentally induced imbalances in isoform proportions skews cell fate probabilities. Using RACE in cells enriched for the 2C-L fate,

we find evidence for skewed expression of particular Stag1 isoforms, arguing that individual ES cells may naturally predominantly express particular Stag1 protein isoforms. Interestingly, the proportions of Stag1 isoforms likely originate in part from the stochastic process of splicing (Fiszbein and Kornblihtt, 2017; Gabut et al., 2011; Salomonis et al., 2010), which is itself further randomized by the fluctuating chromatin landscape (Mateo et al., 2019; Rodriguez et al., 2019), acting as a feedback mechanism for further Stag1 diversity. Thus, the naturally occurring Stag1 isoforms described here may act as intrinsic sources of noise that both directly and indirectly support cellular plasticity.

We propose that a balance of Stag1 protein isoforms is required for variation in gene and repeat expression and thus, a continuum of cellular states. The N-terminal containing AT-hook of Stag1 supports global chromatin compaction and heterochromatin condensation, suggesting that it plays an important role in clustering repetitive sequences into heterochromatin condensates for their regulation. Indeed, we observe dramatic changes to rRNA expression in 5p KD cells where isoforms lacking the N-terminus are specifically retained. Moreover, we note that our live cell microscopy also revealed condensate-like structures within the nucleoplasm. It is possible that the C-terminus of Stag1 regulates euchromatin condensate formation to support pluripotency gene networks. Several lines of evidence support this. First, many of the genes that are downregulated upon 3p KD are super-enhancer-associated genes which are known to form condensate structures (Sabari et al., 2018; Whyte et al., 2013). Further, global nascent transcription is reduced in 3p KD cells, although rRNA is not as affected as in the 5p KD, suggesting that the C-term of Stag1 may have a role in regulating a different subset of nascent transcripts. Thus, the Stag1 protein heterogeneity discovered here provides the necessary fluctuations in the levels and composition of the Stag1 disordered regions to impact the stability of both euchromatin and heterochromatin condensates, thereby supporting plasticity within the population.

Stag1 knockout (Stag1$^{\Delta/\Delta}$) ES cells give rise to mice which survive to E13.5 (Remeseiro et al., 2012b). At first this observation seems at odds with our report that Stag1 is required for pluripotency. However, our observations may in fact explain why the

Stag1$^{\Delta/\Delta}$ mouse model does not exhibit early embryonic lethality. In this model, only the 5' region of Stag1 was targeted, meaning that the Stag1 isoforms lacking the N-terminus may still be retained in the targeted ES cells. This is consistent with our results showing that 5p KD cells have not lost their ability to self-renew nor is their pluripotency gene signature affected. It further suggests that changes to heterochromatin may exist in these cells.

The role for Stag1 in heterochromatin structure described here may be quite general, and the fact that Stag1 binds to HP1a and repetitive sequences known to form constitutive heterochromatin suggests that this is true. Indeed, the AT-hook domain of Stag1 has been shown to be important for telomere cohesion (Bisht et al., 2013). Here we have specifically uncovered a role for Stag1 in nucleolar structure and function in stem cells. Transcriptionally inactive rDNA arrays are associated with the periphery of the nucleolus and accumulate marks of constitutive heterochromatin while actively transcribed rDNA loci are looped inside the nucleolus (Padeken and Heun, 2014). We observe Stag1 at both the interior and periphery, suggesting that it may be involved in multiple aspects of rDNA and nucleolar regulation. Stag1 is enriched at SINE elements and we note that the intergenic spacer region of the rDNA consensus contains multiple B3 and B2_Mm2 sequences, strongly suggesting that the changes in nascent rRNA transcription are direct effects of Stag1 loss and implicates Stag1 in the formation or the maintenance of the nucleolus. Indeed, CTCF has previously been implicated in nucleolar structure by tethering insulators to the periphery (Yusufzai et al., 2004). As well as impacting global translation, the change to nucleolar structure may also indirectly support the emergence of the 2C state by destabilizing complexes such as Nucleolin/Trim28 which are required for the derepression of repeat elements (Percharde et al., 2018).

Finally, a role for Stag proteins or cohesin in translation regulation has, to our knowledge, not been described before despite the fact that translation has important roles in development (Buszczak et al., 2014). How Stag1 contributes to translation in ES cells and whether it plays similar roles in other stem or progenitor populations are important future questions. Our results offer new perspectives on how Stag proteins and cohesin regulation contribute to cell identity during development and in disease.

Acknowledgments

This work would not be possible without the support of a Senior Research Fellowship from the Wellcome Trust awarded to S.H. (106985/Z/15/Z). We would like to thank Sally Lowell and Mattias Malaguti for advice throughout the project. We are grateful to the members of the Hadjur lab for critical discussions and reading of the manuscript. We thank M. Irima for help with VAST-tools pipelines; B. Cairns for the inducible Dux-HA, MERVL-GFP cell line; W.Reik for a second MERVL-reporter ES cell line; H. Rowe for advice on 2C-L cells and J. Vaquerizas for advice on repeat expression analysis. Thank you to Y. Guo and J. Manji in the Cancer Institute CRUK Centre FACS and Imaging core facilities for their invaluable assistance.

Author Contributions

D.P. and S.H. conceived the project. D.P. designed and performed all the experiments on ES cells with assistance from S.W. S.W. performed all protein analysis, generated the SA1-NG-FKBP ES cell line, performed the Spinning Disk microscopy and helped with the siRNA knockdown experiments. W.V. performed all bioinformatic analyses with the exception of the Stag1 enrichments at repeat elements, which was done by M.B. P.D. and S.P. provided advice on CRISPR targeting. D.P. and S.H. formatted all figures and wrote the manuscript with input from all authors.

Declaration of Interests

The authors declare no competing interests.

Figure legends

Figure 1. STAG1 is required for pluripotency.

(A) Log2 fold change of pluripotency factors and cohesin subunit gene expression assessed by qRT-PCR during embryonic stem cell (ES) differentiation towards Epiblast-like (EpiLC) cells. Multiple primer pairs were used for SA1 and SA2 mRNA. Data are derived from two biological replicates.

(B) Whole cell protein extracts (WCL) from serum-grown (FCS) or naïve (2i) ES cells, EpiLC and mouse embryonic fibroblasts (MEF) and analyzed by western blot (WB) for levels of SA1, SA2 and Smc3.

(C) WB analysis of SA1 levels in WCL, cytoplasmic and chromatin fractions upon treatment with scrambled siRNAs (siScr) or SmartPool SA1 siRNAs (siSA1) for 24hrs in 2i ES cells. Tubulin (Tub) and H3 serve as fractionation and loading controls.

(D) Relative expression of Nanog mRNA by qRT-PCR in 2i- (left panel, n=19) or FCS-grown (right panel, n=20) ES cells after no (UT), si scr or si SA1 treatment. Whiskers and boxes indicate all and 50% of values respectively. Central line represents the median. Asterisks indicate a statistically significant difference as assessed using 2-tailed T-test. * $p<0.05$, ** $p<0.005$, n.s., not significant.

(E) Mean Fluorescence Intensity (MFI) of Nanog protein assessed by Immunofluorescence in 2i ES cells treated with si scr, a second siRNA control to Luciferase (esi Luc) and si SA1. Cells were stained for Nanog, SA1 and counterstained with DAPI. Quantifications and statistical analysis were done as above. Each treatment was repeated a minimum of two times (n>100/condition). **** $p<0.0005$.

(F) Volcano plot displaying the statistical significance (-log2 p-value) versus magnitude of change (log2 fold change) from RNA-sequencing data produced in FCS ES cells treated with si scr or siSA1 for 24hrs (n=3). Vertical blue dashed lines represent changes of 2-fold. Selected genes associated with pluripotency and differentiation have been highlighted.

(G) Enrichment score (ES) plots from Gene Set Enrichment analysis (GSEA) using curated Naïve or Primed gene sets (see Methods). Negative and positive normalized (NES) values point to the gene set being over-represented in the top most down- or up-regulated genes in SA1 KD ES cells, respectively. Vertical bars refer to individual genes in the gene set and their position reflects the contribution of each gene to the NES.

(H) Area occupied by AP+ colonies in ES cells treated with si scr and si SA1 from three independent biological replicates where n>50 colonies/condition were counted.

(I) CRISPR/Cas9 was used to knock-in a NeonGreen-v5-FKBP tag on both alleles of endogenous Stag1 at the C-terminus (SA1$^{NG\text{-}FKBP}$). The resultant Stag1 protein is 42kDa larger. Shown are known features of SA1 including the N-terminal AT-hook (AT) and the stromalin conserved domain (SCD). WB analysis of SA1 levels in a targeted ES clone after treatment with DMSO-only or dTAG for 16 hrs. Tubulin (Tub) serves as a loading control.

(J) Analysis of the area of AP+ colonies as above in wildtype ES and in a SA1$^{NG\text{-}FKBP}$ ES clone treated with DMSO-only or dTAG for 16 hrs. Data is represented as mean ± SEM of three independent experiments and relative to the DMSO control.

See also Figure S1.

Figure 2. STAG1 associates with AT-rich heterochromatin.

(A) Live-cell Spinning Disk confocal images of three SA1$^{NG\text{-}FKBP}$ cells counterstained with Hoechst. Arrows indicate notable regions of overlap of SA1 and Hoechst, including inside the nucleolus and at Hoechst-dense foci (top cell), and at the nucleolar periphery (bottom cells).

(B) Mean Intensity of SA1-NeonGreen signal from (A) within the whole nucleus or Hoechst-dense foci as identified using Imaris in the SA1$^{NG\text{-}FKBP}$ clone treated with DMSO or dTAG. n=270 cells/condition.

(C) co-Immunoprecipitation of endogenous SA1 with Smc3, HP1a and Nucleolin in ES cells.

(D) Percentage of SA1 peaks (unique reads) at promoters, exons, transposable elements introns and intergenic sequences.

(E) STAG1 ChIP-seq data (unique and multimapping reads) aligned to full-length Repeat elements of the indicated SINE, LINE and LTR families.

(F) Confocal images of Immunofluorescence (IF) to SA1 and H3K9me3 in siscr or siSA1 treated ES cells, counterstained with DAPI.

682 (G) Imaris quantification of the volume of H3K9me3 foci from (F). Box plots and statistical
683 analysis were done as above. Each treatment was repeated twice (n>100/condition).
684 $p<0.05$.

685 (H) Global chromatin accessibility as detected by DNase I digestion of genomic DNA at
686 different concentrations from 200,000 cells/condition in siScr or siSA1 ES cells.

687 (I) Confocal images of IF to GFP and H3K9me3 in ES cells expressing full-length SA1
688 tagged with GFP (SA1-FL), counterstained with DAPI.

689 (J) Imaris quantification of the volume of H3K9me3 foci from (I). Box plots and statistical
690 analysis were done as above. n>50/condition. $p<0.0005$.

691 (K) PONDR (Predictor of Natural Disordered Regions) analysis of SA1 protein using VSL2
692 predictor at http://www.pondr.com showing consecutive stretches of disordered regions
693 corresponding to the N- and C-terminus of Stag1.

694 See also Figure S2.

695

696 **Figure 3. Diverse Transcription-Regulatory Control of Stag1 in ES cells.**

697 (A, B) 5' Rapid Amplification of cDNA ends (RACE) for SA1 in ES and EpiLC cells. Arrows
698 indicate bands which were cloned and sequenced. In A, red star indicates SATS TSS and
699 red arrow indicates canonical TSS. In B: red indicates full length Stag1 with both SATS
700 and can TSSs; dark blue indicates alternatively spliced variants, skipping of various exons
701 in 5' region; light blue indicates the TSSs at e6, e7.

702 (C) 3' RACE for SA1 in ES cells. Arrows indicate bands which were cloned and
703 sequenced. Red indicates canonical full-length end; green indicates end in i25.

704 (D) Top, schematic of the Stag1 gene annotation in mm10 and the identified TSS and
705 TTSs from RACE indicated. Bottom, aligned sequence clones from the PCR mini-screen
706 and their predicted impact on the SA1 protein, right. Green arrows and red bars within the
707 transcripts indicate start of the coding sequence and the TTS respectively. Shown also
708 are the regions which code for the AT hook and SCD domains.

709 (E) Percent Spliced In (PSI) calculations based on VAST-Tools analysis of RNA-seq from
710 multiple 2i (blue) and FCS (red) datasets (see Methods for details of libraries). Data are
711 shown relative to Neural Stem cell (NSC) frequencies to highlight the events that are ES-
712 specific.

(F) Genome topology at the Stag1 locus. Hi-C contact maps in ES (2i) and NS cells of the 900kb region on chromosome 9 containing the Stag1 topologically associated domain (TAD). TADs are denoted with a vertical line and as repressed (orange) or active (blue). Shown also are tracks for Genes, Nanog and CTCF ChIP-seq as well as a track indicating the directionality of CTCF binding sites (red, forward; blue, reverse). Aligned to the Gene track are also the SA1 transcripts discovered above where red represents the untranslated regions and blue the coding body. UMI-4C-seq viewpoints (asterisks on the ChIP tracks) are positioned to the leftmost CTCF site ('CTCF bait') and to the Nanog site 40 kb upstream of the Stag1 canonical TSS ('Nanog bait'). For each bait, UMI information for each cell type is shown as well as the comparative plots where red represents an enrichment of contacts in ES compared to NS.

(G) Top, cartoon depicting functional domains within Stag1 protein, including the AT-hook (aa 3-58); Stromalin conserved domain (SCD, aa 296-381) and the C-terminus. Bottom, the predicted Stag1 protein isoforms based on transcript analysis with estimated sizes for each isoform. Purple boxes in the 105kDa and 90kDa isoforms represent retained introns.

(H) PONDR tracks as before shown for the N-terminal truncated (top) and C-terminal truncated (bottom) transcripts.

(I) Chromatin immunoprecipitation of SA1 from ES cells.

(J) WB analysis of SA1 isoforms in chromatin fractions from ES cells treated with siscr and siSA1. H3 serves as a fraction and loading control.

(K) Chromatin immunoprecipitation for the v5 tag in $SA1^{NG\text{-}FKBP}$ ES cells treated with DMSO-only or dTAG. Note, SA1 bands now run 42kDa higher due to the addition of the tag.

See also Figure S3, Table S1 and Table S2.

Figure 4. Fluctuations in the levels of the Stag1 isoforms skews cell fates.

(A) Schematic of the siRNA pools used in this study. esiRNA SATS represents 'enzymatically-prepared' siRNAs (see Methods).

(B) Relative expression of SA1 mRNA by qRT-PCR in FCS ES cells after no treatment (UT), or upon si scr, si SA1 SP, si 3p, si 5p or esi SATS treatment. Data is represented as mean ± SEM, n=3. * $p<0.05$

744 (C) WB analysis of SA1 levels in FCS ES WCL upon treatment with the same siRNA
745 panel from (B). Tubulin serves as a loading control.

746 (D) RNA-seq reads (as transcripts per million, TPM) aligning to sectioned Stag1 in
747 datasets from the various siRNA pools and presented as relative to untreated ES RNA-
748 seq. N-terminal reads include SATS and exons 1-8, Mid reads are from exons 12-19 and
749 C-terminal reads include exons 20-25 and exons 26-34.

750 (E) Left, 5' and Right, 3' RACE for SA1 in ES cells treated with the siRNA panels. Arrows
751 indicate bands which were cloned and sequenced and colour-coded as before.

752 (F) Chromatin immunoprecipitation using an N-terminal Stag1 antibody in $SA1^{NG\text{-}FKBP}$ ES
753 cells treated with DMSO-only or dTAG for 16 hrs. Green arrow indicates residual C-
754 terminal truncated Stag1 isoforms.

755 (G) Enrichment score (ES) plots from GSEA using Naïve or Primed gene sets as in Figure
756 1G and RNA-seq data from the different siRNA treated ES cell samples.

757 (H) Area occupied by AP+ colonies in ES cells treated with the siRNA panel from three
758 independent biological replicates. n>50 colonies/condition were counted.

759 See also Figure S4.

760

761 **Figure 5. Stag1 N-terminus represses heterochromatin and promotes transition**
762 **towards totipotency.**

763 (A) Confocal images of IF to GFP and H3K9me3 in ES cells expressing SA1-FL-GFP,
764 SA1-exon5Δ-GFP or SA1-CtermΔ-GFP. Nuclei were counterstained with DAPI.

765 (B) Imaris quantification of the volume of H3K9me3 foci per cell from (A). Box plots and
766 statistical analysis were done as before. n>50/condition. ** $p<0.005$, **** $p<0.0005$.

767 (C, D) Relative expression of (C) repeat elements; MERVL, LINE1-T and IAPEz RNA
768 (n=8 biological replicates each) and (D) 2C-L genes; Gm6763 (n=6) and Dux (n=5) by
769 qRT-PCR in FCS ES cells after treatment with the siRNA panel. Boxes and whiskers as
770 before. * $p<0.05$, n.s, not significant as per 2-tailed T-test. NB. The p-value for Gm6763
771 expression in 5p KD vs UT = 0.07.

772 (E) Enrichment score (ES) plots from GSEA using a published 2C-L gene set (Methods)
773 and RNA-seq data from the 3p and 5p siRNA treated ES cell samples.

(F) Representative FACS analysis of the proportion of ES cells expressing a MERVL-GFP reporter in the different siRNA treated cells and including siRNA to Chaf1 as a positive control. Percentage of MERVL-GFP+ cells based on Flo-Jo analysis is shown in red.

(G) Proportion of MERVL-GFP+ cells in the different siRNA conditions relative to the siChaf1 positive control. Data is represented as mean ± SEM of four independent experiments. * $p<0.05$, ** $p<0.005$, n.s, not significant as per 2-tailed T-test.

(H) 5' RACE for SA1 in Dux-HA, MERVL-GFP ES cells sorted for GFP+ cells. Arrows indicate bands which were cloned and sequenced and colour-coded as before.

(I) Sequence of the 5'RACE product identifying a novel Stag1 TSS with direct splicing of exon7 to MT2_MERVL element.

(J) Co-Immunoprecipitation of SA1 with Nucleolin and Trim28 in ES cells.

See also Figure S5.

Figure 6. rRNA expression and nascent translation upon Stag1 loss.

(A) Confocal images of IF to Nucleolin (Ncl) and Nascent RNA in ES cells treated with the siRNA panel. Nuclei were counterstained with DAPI. Note the accumulation of nascent RNA within the nucleolus.

(B) Imaris quantification from (A) of the mean intensity of nascent RNA (EU-488) within the nucleoli, as defined by Nucleolin signal. Box plots and statistical analysis were done as before. Data are from two biological replicates, n>50/condition, except for siSA1 5p where n>35. ** $p<0.005$, **** $p<0.0005$ as per 2-tailed T-test.

(C, D) Global analysis of Nascent transcription by measuring EU-488 incorporation using Flow cytometry. (C) representative Flo-Jo analysis of EU incorporation in ES cells treated with the siRNA KD panel. (D) Quantification of the change in EU incorporation relative to si Scr treated cells. Data are from two biological replicates *$p<0.05$, **$p<0.005$ as per 2-tailed T-test.

(E, F) Global analysis of Nascent translation by measuring HPG-597 incorporation using Flow cytometry. (E) Representative Flo-Jo analysis of HPG incorporation in ES cells treated with the siRNA KD panel. (F) Quantification of the change in EU incorporation

804 relative to si Scr treated cells. Data are from four biological replicates *p<0.05, **p<0.005
805 as per 2-tailed T-test.

806 See also Figure S6.

807

METHODS

Embryonic stem cell culture and siRNA-mediated knockdown.

Male mouse E14 embryonic stem (ES) cells were cultured in serum (FCS) or naïve (2i) conditions. Serum-cultured cells were grown on 0.1% gelatin-coated plates in GMEM, 10% FCS (Sigma), NEAA, Na Pyruvate, 0.1 mM ßMercaptoethanol (BMe), Glutamax, and freshly added LIF (1:10,000). 2i-cultured cells were grown on plates coated with Fibronectin, in DMEM:F12/Neurobasal 1:1, KnockOut Serum Replacement, N2, B27, Glutamax, 1μM PD0325901, 3μM CHIR9902, 0.1 mM BMe, and freshly added LIF as above. DuxHA/MERVL-GFP cells were cultured in 2i conditions. siRNAs were purchased from Horizon Discovery (previously Dharmacon) or Sigma (for 'enzymatically-derived' esiRNAs). siRNA knockdowns (KDs) were performed for 24hr with the exception of those in Figure 5 which were performed for 72hr. Knockdowns were performed in 6-well plates where 200,000 cells were seeded for 72 hr KDs, and 400,000 for 24 hr KD. 50pmol siRNAs were transfected using RNAiMax Lipofectamine at the time of seeding, and after 48 hrs for 72hr timepoints. Two siRNA controls were used, scrambled (scr) was D-001810-10 and Luciferase (esiLuc) control purchased from Sigma. siSA1 'SmartPool' (SP) was derived from equimolar ratios of commercial siRNAs (D-041989-02, -04, -05, -06, -07, -08). siSA1 5p was a custom Duplex siRNA sequence (AGGAGCAGGUCGUGGAAGAUU). siSA1 3p was derived from equimolar ratios of commercial siRNAs J-041989-05, -07, -08. esiRNA to SATS was purchased from Sigma as a custom-made product to the entire SATS 5'UTR (mm10 chr9:100,597,794-100,598,109).

qRT-PCR analysis

Total RNA was isolated using Monarch RNA prep kit (NEB). Reverse transcription was performed on 0.5 μg DNase-treated total RNA using Lunascript RT (NEB) in 20μl reactions. qPCR was performed using 2x SensiFAST SYBR No-ROX kit (Bioline) in 20 μl reactions using 1μl of RT reaction as input and 0.4μM each primer.

Alkaline Phosphatase (AP) assay and quantification

Cells were seeded in 6 well plates and transfected with siRNAs at the time of plating as above. After 24 hrs, cells were collected for RNA isolation and KD efficiency analyzed by qRT-PCR. Cells from each condition were counted and 1,000 cells per well seeded into a new 6-well plate. Cells were re-transfected after 48 hrs using 5 pmol of siRNAs. Cells were fed every day. Four days

after seeding cells at clonal density, the cells were assayed for alkaline phosphatase (AP) expression using StemTAG Alkaline Phosphatase staining kit (Cell Biolabs CBA-300). AP stained cells were imaged in 6-well plates using a M7000 Imaging System (Zeiss) with a 4X objective and a Trans-illumination brightfield light source. For quantification, AP-high and AP-low colonies from each condition were counted. Area occupied by AP-high colonies was also measured using ImageJ, and plotted as fraction of total area of all colonies.

RACE (Rapid Amplification of cDNA Ends) and PCR mini screen

RACE was performed using GeneRacer kit (RLM RACE, Invitrogen L1500). 2μg of total RNA was used as input. Final products were amplified by nested PCR, using Kapa 2x MasterMix. First PCR was done in a 50μl reaction using 1μl RT as input, 25 cycles. DNA was purified using Qiagen PCR Purification kit, and nested PCR was performed on a tenth of the first PCR for 30 cycles. Viewpoint for 5'RACE was in exon 2 (Fig 3A) or exon 8 (Fig 3B) of Stag1. Viewpoint for 3'RACE was in exon 23 (Fig 3C). RACE primer details can be found in Table S3. PCR products were excised from the gel, A-tailed using Klenow exo- (NEB) and cloned into pCR4-TOPO vector (Invitrogen). At least three clones were sequenced per PCR product. For the PCR Mini-Screen, forward primers at either SATS or canonical 5' UTR were used with reverse primers either at the end of Stag1 canonical coding sequence, or at the end of coding sequence in intron 25 (see Table S3). PCR was performed using Kapa 2x MasterMix. DNA was excised from the gel, A tailed, and cloned into pCR4-TOPO. At least six clones per PCR product were Sanger-sequenced. Sequences from the PCR Mini-screen were aligned using Minimap2 (2.14-r884) in 'splice' mode to ensure long read splice alignment (Fig 3D and S3A).

PONDR Predictions

Internally disordered regions were predicted using VSL2 predictor at http://www.pondr.com.

CRISPR-Mediated Stag1 Knock-in Cell Line Generation

The guide RNA targeting Stag1 3' terminal coding region was designed using Tagin Software (http://tagin.stembio.org) and purchased from IDT. Lyophilised gRNA was rehydrated in RNA duplex buffer (100μM). The single stranded oligodeoxynucleotides (ssODN) encoding mNeonGreen (mNG)-V5-FKBP12^{F36V} and the left and right homology arms was designed using the software tool ChopChop (https://chopchop.cbu.uib.no) and purchased as a High-Copy Amp-resistant plasmid from Twist Bioscience. 2.2μl gRNA (100μM) was mixed with 2.2μl tracrRNA

873 ATTO 550nm (IDT) and annealed together. The RNA duplex was then incubated with 20μg S.p
874 Cas9 Nuclease V3 (IDT) for 10min at room temperature and stored on ice prior to transfection.
875 Linearised KI sequence was mixed with 100% DMSO and denatured at 95°C for 5min. The
876 ssODN was plunged immediately into ice. The RNP complex was mixed with confluent 2i-grown
877 ES cells re-suspended in P3 transfection buffer (Lonza) before being transferred to an
878 electroporation microcuvette well (Lonza). Transfection was performed using a 4D Amaxa
879 electroporator. Post-nucleofection, the cells were seeded into a fibronectin-coated 6 well plate
880 with fresh ES media. The media was changed daily for four days before being expanded into a
881 T75 flask. Confluent ES cells were FACS sorted for GFP+ population (BD FACS Aria Fusion Cell
882 Sorter) and sparsely seeded into 10 cm plates. Clones were manually picked into 96 well plates
883 and expanded for selection by v5 IF, genotyping and Sanger sequencing.
884
885 **Dox-inducible Stag1-GFP isoform cell lines**
886 Stag1 isoforms were cloned into pCW57.1 vector (Addgene 41393), modified using Gibson
887 assembly to include an EGFP tag at the 3'end of the Gateway cassette, using Gateway
888 recombination by LR clonase. For primers used to clone the isoforms see Supplementary Table
889 S3. Plasmids were transfected into 2i-grown ES cells using Lipofectamine 3000 and cells grown
890 in Puromycin-supplemented media (1μg/ml) for ten days to make stable lines. Isoform expression
891 was induced using 2μg/ml Doxycycline for 24 hrs, and the population enriched for GFP-positive
892 cells using FACS. For IF experiments, isoforms were induced by adding Dox for 48 hours.
893
894 **Protein Lysates, Fractionations and Western blotting.**
895 Whole cell lysates (WCL) were collected by lysis in RIPA buffer (150mM NaCl, 1% NP-40
896 detergent, 0.5% Sodium Deoxycholate, 0.1% SDS, 25mM Tris-HCl pH 7.4, 1mM DTT) and
897 sonicated at 4°C for x5 30 second cycles using Diagenode Bioruptor. Insoluble material was
898 pelleted and the supernatant lysate was quantified using BSA Assay (Thermo Scientific). For
899 cellular fractionations, a cellular ratio of $5x10^6$ cells/80μl buffer was maintained throughout the
900 protocol. Cells were re-suspended in Cell Membrane Lysis Buffer (0.1% Triton X, 10mM HEPES
901 pH 7.9, 10mM KCl, 1.5mM MgCl2, 0.34M sucrose, 10% glycerol, 1mM DTT), incubated on ice for
902 5min and centrifuged for 5min at 3700rpm to collect the cytoplasmic sample. The pellet was
903 washed and then re-suspended in Nuclear Lysis Buffer (3mM EDTA, 0.2mM EGTA, 1mM DTT)
904 and incubated on ice for 1 hr. Nuclear lysis was aided by sonication with a handheld homogeniser
905 (VWR) for 10sec at 10min intervals. The nucleoplasmic supernatant and chromatin pellet were

separated by centrifugation at 9000rpm for 10min at 4°C. The chromatin pellet was re-suspended in 160μl 2X Laemmli Buffer (Bio-Rad). Equal volumes of each fraction were used for Western Blotting (WB). Cytoplasmic and nucleoplasmic protein samples were diluted in 2X Laemmli Buffer and boiled for 5min at 95°C, then loaded on a 4-20% SDS-PAGE gel (Bio-rad) or a 3-8% Tris Acetate gel (Invitrogen). Proteins were wet transferred onto a PDVF membrane (Millipore) and assessed for successful transfer with Ponceau Red (Sigma). The membrane was blocked with 10% milk and incubated with primary antibodies in 1% milk, 0.1% Tween-PBS overnight at 4°C. Membranes were imaged with SuperSignal West Femto Maximum Sensitivity (Thermo) on an ImageQuant.

Chromatin Co-Immunoprecipitation (co-IP)

Cells were re-suspended in 0.1% NP-40-PBS (1ml/1x10^7 cells) with 1X Protease Inhibitors (Roche) and 1mM DTT, and centrifuged at 1500rpm for 2min at 4°C. The pellet was re-suspended in Nuclear Lysis Buffer (3mM EDTA, 0.2mM EGTA, 1X Protease Inhibitors, 1mM DTT), vortexed for 30sec before being incubated on a rotator for 30min at 4°C and centrifuged at 6500g for 5min at 4°C to isolate the glassy chromatin pellet. This was re-suspended in High Salt Chromatin Solubilisation Buffer (50mM Tris-HCl pH 7.5, 1.5mM MgCl2, 300mM KCl, 20% glycerol, 1mM EDTA, 0.1% NP-40, 1mM Pefabloc, 1X Protease Inhibitors, 1mM DTT) with Benzonase (Sigma) (6U/1x10^7) and incubated on rotator for 30min at 4°C. Chromatin was digested with 3x 10sec sonication at 30% intensity with a Vibra-Cell probe. The supernatant was collected by centrifugation at 1300rpm for 30min at 4°C, and then diluted to 200mM KCl concentration with no KCL buffer. 30μl of Dynabeads (Invitrogen) were used per co-IP. Beads were washed 2x in 200mM KCl IP Buffer, re-suspended in IP Buffer with 10μg of the IP antibody, or an IgG-containing serum to match the species of the IP antibody and placed on rotator for 5h at 4°C. Beads were washed 3x in IP buffer and then incubated in 1mg chromatin lysate on a rotator overnight at 4°C. The beads were washed, re-suspended in 2X Laemmli Buffer (Bio-Rad), boiled for 10min at 95°C and used for WB as above.

Immunofluorescence and Microscopy

ES cells were cultured on fibronectin or gelatin-coated cover glass in 6-well plates. Cells were fixed in 4% Paraformaldehyde for 5min and incubated in 0.1% Triton X-PBS for 10min before being washed and blocked in 10% FCS-PBS for 20min. Primary antibodies were diluted in 10% FCS, 0.1% Saponin (Sigma) and incubated overnight at 4°C. The next day, the cells were

incubated with an Alexa fluorophore-conjugated secondary antibody diluted in 10% FCS, 0.1% Saponin for 1 hr at room temperature, washed and mounted on cover slides with ProLong Diamond Antifade Mountant with DAPI (Invitrogen). Z-stacks imaging of fixed cells was done using a LSM 880 confocal microscope (Zeiss) with a 63X oil objective. Analysis was performed using Imaris 9.6 (Oxford instruments). Live cell imaging was performed using a 3i Spinning Disc confocal microscope (Zeiss). Stag1-mNG-V5-FKBP12^{F36V} cells were seeded in an 8-chambered coverglass (Lab-Tek II) and DMSO or dTAG (500nM) were added for 24hr before imaging. Directly prior to imaging, cells were incubated with Hoechst 33342 (BD Pharmingen) for 45min, and then replaced with fresh 2i ES media. Cells were imaged as confocal Z-stacks using DAPI and GFP lasers with a 63X objective and 1.4 Numerical Aperture.

Antibodies used in this study

Protein	Catalogue No.	Company	Figure references
Stag1/SA1, N-term epitope	ab4455	Abcam	1B, C, I, S1C, K, 2C, S2C, E, 3J, S3G, 4C, F, 5J
Stag1/SA1, C-term epitope	ab4457	Abcam	2F, S5A, 3I
Stag2/SA2	A300-158A	Bethyl	1B, S1C
Smc3	ab9263	Abcam	1B, 2C
Nanog	ab70482	Abcam	1E, S1F
Tubulin (Tub)	T5168	Sigma	1C, 1I, S2E, 4C, S6A
Actin	Mab8929	Novus	S1C
H3	ab1791	Abcam	1C
v5	14-6796-82	Invitrogen	3K
HP1a	2616	Cell Signalling	2C, S2B, C
Nucleolin (Ncl)	ab22758	Abcam	2C, 5J, 6A, S6A
H3K9me3	ab8898	Abcam	2F, I, S2E, 5A, S5A
H3K4me3	ab8580	Abcam	S2E
Alexa488-anti-GFP (GFP)	A-21311	ThermoFisher	2I, S2A, B, 5A
Trim28	MA1-2023	ThermoFisher	5J

Chromatin accessibility analysis by DNase I treatment.

DNase I digestion was performed as in Huo et al (Mol Cell, 2020), with modifications. 200,000 cells per condition were resuspended in DNaseI digestion buffer (50mM Tris-HCl pH 7.5, 5mM

$MgCl_2$, 0.1 mM $CaCl_2$, 0.2% Triton X, 5mM Na butyrate, protease inhibitor) and incubated for 10 min at room temp. DNase I (50U/μl, ThermoFisher ES0523) was diluted in 1X DNase buffer and added to cells to have the following Units/μl: 1.25, 2.5, 5, 7.5 and 10. Cells were incubated at 37C for 10 min in the thermoblock, shaking at 1,000 rpm. To stop the digestion, 10μl 0.5M EDTA and 10ul 10% SDS was added and incubated 10-15min at room temperature. 400μl TE buffer was added, followed by 10μl PureLink RNaseA (ThermoFisher, 20mg/ml, 12091021), and incubated at 37C for 1hr. Proteinase K digestion was then performed by adding 100μg of Proteinase K and incubating at 55C for 3hr to overnight. To isolate DNA, 30μl of 5M NaCl and 525 μl Isopropanol was added, DNA precipitated at room temperature for 15 min, pelleted by high speed centrifugation at 4C for 20min, dried, resuspended in 20μl of TE buffer and loaded on 1% agarose gel. The gel was stained using SYBR Green.

Nascent transcription and translation analysis

For nascent transcription analysis, we used the Click-iT® RNA Alexa Fluor® 488 HCS Assay (Invitrogen C10327). ES cells were labelled with 1mM EU for 45min at 37C in fresh ES media. Cells were fixed in solution or onto coverslips with 3.7% paraformaldehyde and permeabilised with 0.5% Triton-X solution. Cells were incubated with the Click-iT reaction cocktail for 30min. Cells were then either processed further for Immunofluorescence as per methods described above (directly to the blocking step) or analysed by flow cytometry on a BD Fortessa X20. For the Nascent translation analysis, Click-iT™ HPG Alexa Fluor™ 594 Protein Synthesis Assay Kit (Invitrogen C10429) was used. Cells were pre-incubated in Methionine-free media for 30 min in the 37C incubator before addition of L-homopropargylglycine (HPG) at 50μM. Cells were incubated with HPG for 30 min, then collected, fixed, permeabilized, and stained using Click-It reaction in low retention tubes. HPG incorporation was measured by Flow Cytometry. FACS analysis (in Figures 5,6) was done with FloJo software (version 10.7.1).

Next generation Sequencing and Analysis

Genomic data generated in this study (RNA-seq, PacBio-seq and UMI4C-seq) was submitted to GEO with the Accession GSE160390.

RNA sequencing (RNA-seq) library preparation and sequencing

ES cells were treated for 24hrs with siRNA pools to Stag1 (SA1) and two sets of control siRNAs, scrambled (SCR) and Luciferase (Luc). There are three replicate sets for SP KD and two for the

siRNA pools (SATS, 3p, 5p). Total RNA was isolated using NEB Monarch RNA prep kit. 1μg of total RNA was rRNA-depleted using NEBNext rRNA depletion kit (Human/Mouse/Rat). Libraries were prepared from 10-50ng rRNA-depleted total RNA, depending on availability of material, using NEBNext Ultra II directional RNAseq kit according to manufacturer's instructions using 8 cycles of PCR. All ESC FCS libraries were rRNA depleted and only the ESC 2i libraries were PolyA-enriched before library prep. Two rounds of PolyA+ enrichment were performed. RNA-seq libraries were sequenced on the Illumina HiSeq3000 platform, 75bp paired-end or single-end reads. Reads were quality controlled using FASTQC. RNA-seq data was processed using the RNA-seq Nextflow pipeline (v19.01.0), with the following parameters –aligner hisat2 –genome mm10, with –reverse_stranded specified for paired-end samples. FeatureCounts output was parsed through edgeR (v3.16.5) and DESeq2 (v1.14.1) to generate normalised expression counts. The normalised counts for RNAseq (Figure 1) were calculated in edgeR. Low expressed genes were removed (rowSum cpm <2 across SCR and SA1SP replicates), normalisation factors were calculated using calcNormFactors and dispersions estimated using estimateDisp. The edgeR volcano plot statistics were calculated using the exactTest and topTags functions. To generate the normalised counts for RNAseq experiments required to calculate the log2FC GSEA ranked lists, the FeatureCounts output for all experiments was combined into a single table and read into DESeq2. A DESeq2 object was built using the function DESeqDataSetFromMatrix and estimation of size factors and dispersions were calculated using the DEseq function. Normalised counts were calculated using the 'counts' function. Low expressed genes (rowSum normalised count <10 across all samples) were removed.

GSEA

Broad Institute GSEAPreranked (v4.0.3) was used to determine the enrichment of curated genesets within our RNA-seq data. For each sample a ranked list was generated with genes ranked in descending order by their log2FC value using normalised expression scores from DEseq2. Log2FC per gene was calculated between the KD and its respective SCR using the following calculation: Log2(normalised_counts KD +1) - log2(normalised_counts SCR +1). In the case of experiments with multiple KD replicates, the average log2 normalised count was used. Three gene sets were assayed in this study, 'naïve pluripotency', 'primed pluripotency' and '2C signatures'. The naïve and primed pluripotency gene sets were curated in-house from Fidalgo M et al. (CSC, 2016) where genes were selected if they had $\geq$2 fold change. The naïve and primed gene sets contained 661 and 580 genes respectively. The 2C signatures gene set (147 genes)

was obtained from Percharde M et al. (Cell, 2018). Gene sets were classed as having significant enrichment if the p-value was ≤0.05 and the normalised enrichment score (NES) exceeded +/- 1.

VAST-TOOLS

VAST-TOOLS was used to generate Percent Spliced In (PSI) scores, a statistic which represents how often a particular exon is spliced into a transcript using the ratio between reads which include and exclude said exon. Paired-end RNA-seq datasets were submitted to VAST-TOOLS (v2.1.3) using the Mmu genome (Tapial J et al, Gen Res 2017). Briefly, reads are split into 50nt words with a 25nt sliding window. The 50nt words are aligned to a reference genome using Bowtie to obtain unmapped reads. These unmapped reads are then aligned to a set of predefined exon-exon junction (EJJ) libraries allowing for the quantification of alternative exon events. The output was further interrogated using a script which searches all hypothetical EEJ combinations between potential donors and acceptors within Stag1. PSI scores could be obtained providing there was at least a single read within our RNAseq data that supported one of these potential events. Some datasets were combined to have enough reads for the analysis. See Table S1 for PSI values and names of RNA-seq libraries used for analysis in Figure 3E and S4B.

Quantifying sectioned Stag1

Stag1 was split into 5 sections; SATS, e1-e8, e12-e19, e20-e25, e26-e34. Using Kallisto (v0.46.1), raw RNAseq reads were used to quantify each section of Stag1. Kallisto was run in quant mode, using the –rf-stranded parameter, outputting a TPM per Stag1 section. A line plot was generated showing TPM in relative to UT.

PacBio library, sequencing and analysis

ES cells were cultured in naïve 2i conditions and PolyA-enriched mRNAs were hybridized to a custom Biotinylated oligonucleoltide probe set. Post-capture, mRNAs were amplified using the Clontech SMARTer PCR cDNA Synthesis Kit with 9 cycles and used in the SMRTbell library prep according to manufacturers instructions. The library was sequenced on the SMRTseq 2000 platform. PacBio reads were processed through the SMRTLINK v8.0.0 IsoSeq3 pipeline. 403,995 Circular consensus sequences (CCS) were generated using default parameters (--minPasses = 1, --min-rq = 0.8, CCS Polish = No). Further refining through lima (removal of adapters and correct orientation of sequences), poly-A trimming and concatemer removal resulted in 265,106 full length

non-chimeric (FLNC) reads. FLNC reads were aligned to the mm10 genome using Minimap2 with the following parameters (-ax splice, -uf, -k14).

ChIP-seq Analysis

Previously published Stag1 Chromatin Immunoprecipitation-sequencing (ChIP-seq) datasets from ES 2i cells (GSE126659, only Replicate 1 and 2 libraries) were trimmed using trim_galore and aligned to mm10 using bowtie2. Peak detection was performed with MACS2 using uniquely reads (MAPQ≥2). Peaks were overlapped with genomic features in a hierarchical manner (promoters > exons > repeats > introns > intergenic), and overlap frequency was compared with a randomly shuffled version of the peaks. To identify repeat families enriched for STAG1 peaks, a previously described pipeline was used (Deniz O et al. Nat Comm, 2020) that compares family-levels overlap frequency with that observed in 1,000 permutations of random peak shuffling. Coverage profiles across specific TE families were generated using HOMER and including multi-mapping reads (MAPQ<2).

UMI-4C library preparation.

1x10^7 cells were fixed at RT for 10min in 1% formaldehyde and fixation was quenched with 0.125M Glycine for 5min. Cells were then lysed on ice in 10ml Lysis Buffer (10mM NaCl, 10mM Tris-HCl pH 8.0, 0.25% NP40, protease inhibitor) for 30min, followed by 10 strokes of douncing using a tight pestle. Nuclei were pelleted, 8min 700 rcf, washed in 1ml 1.2X DpnII buffer in Protein LoBind tubes (Eppendorf) and resuspended in 500 μl 1.2X DpnII buffer. **15ul of 10% SDS** was added and incubated for 1hr at 37°C shaking at 650 rcf. **50ul of 20% TritonX** was added to quench the SDS and incubated for 15 min at 37°C with shaking. **750U of DpnII was added and incubated overnight at 37**°C **with interval shaking. The next morning, nuclei were** pelleted at 4°C by 650 rcf for 5 min and resuspended in 500μl 1X DpnII buffer. 500U DpnII was added and incubated for an additional four hours. The nuclei were washed twice in 100 μl of 1X T4 Ligase Buffer and resuspended in 200 μl Ligase Buffer. **6ul of T4 DNA Ligase was added and incubated for 3hr at 16°**C**.** Nuclei were then pelleted, resuspended in 200 μl 1x fresh Ligase Buffer, 6μl of T4 DNA Ligase added, and incubated overnight at 16°C. Samples were treated with 20μl of ProtK (NEB Molecular Biology Grade), incubated for 3 hrs at 55°C and 5 hrs at 65°C to reverse crosslinks. Samples were treated with **RNase A** (PureLink, Invitrogen) for 1 hr at 37°C and DNA was extracted and precipitated overnight. For library preparation, 3x5μg of ligated DNA was sonicated using Covaris (10% duty cycle, intensity 5, cycle burst 200, 70sec). Samples were end-repaired

using DNA PolII Klenow Large Fragment (NEB), A-tailed using Klenow (exo-) (NEB), and Illumina indexed adapters ligated using Quick DNA Ligase (NEB). Reactions were denatured at 95°C for 3 min, placed on ice, and purified using 1.2X SizeSelect AmpPure beads to recover ssDNA. Libraries were amplified using GoTaq (Promega), with 20 cycles for PCR1 and 15 cycles for nested PCR2 on 50% material from 1st PCR. For custom UMI bait sequences, see Table S3.

Hi-C and UMI-4C-seq analysis

Hi-C libraries were analysed as previously described (Barrington 2019). UMI-4C tracks were processed using the 'umi4cPackage' pipeline (v0.0.0.9000) (Schwartzman, O et al. Nat Meth 2017). Briefly, raw reads are parsed through the UMI-4C pipeline, those reads containing the bait and padding sequence are retained and de-multiplexed. Reads lacking the padding sequence are considered non-specific and are removed from further analysis. Retained reads are split based on a match to the restriction enzyme sequence to create a segmented fastq file. The first 10 bases of read 2 are extracted and attached to the segments derived from each read pair. Mapping to mm10 is done with Bowtie2. Read pairs that have reverse complement segments are mapped to a restriction fragment ID, with the fragment ID, strand and distance from each end represented within a fragment-chain table. UMI filtering is used to determine the number of molecules supporting each ligation event. The resulting UMI-4C tracks are then imported into R, and data from multiple bait replicates can be merged by summing the molecule counts per ligated fragment, at which point contact intensity profiles and domainograms around the viewpoint can be generated (see Figure 3). The contact intensity profile represents the mean number of ligations within a genomic window, with the resolution of the contact intensity profile being determined by the window size (set to 15 here). The domainogram reports the mean contact per fend at a series of window sizes, a stacked representation of contact intensity values in increasing window sizes from 10 to 300 fragment ends, their colour can be used to identify peak locations. ES and NSC contact profiles were compared after normalisation to correct for bias (see Schwartzman et al for further details). For the compared profiles, the total molecule count for restriction fragment ends for each are calculated at three ranges around the viewpoint. One profile is selected as a reference and the second is scaled to the first using the ratio in total molecule counts between the two profiles as the scaling factor. Below the contact profile is the profile resolution indicator, which shows the number of fends required to include at least 15 UMI molecules. The darker the colour, the larger the window size required. The domainogram at the bottom represents the log2

1118 ratio between the domainogram values of the compared profiles and highlights locations where
1119 ES has more contacts than NSC or vice versa.
1120

Supplementary titles and legends.

Figure S1. Stag1 is required for pluripotency, Related to Figure 1.

(A) Cartoon of the cohesin complex including the core trimer subunits of Smc1a, Smc3 and Rad21 complexed with either Stag1 or Stag2.

(B) Relative expression of Stag1 and Stag2 mRNA by qRT-PCR in 2i- or FCS-grown ES cells, EpiLCs and MEFs. Data is represented as mean ± SEM of two independent experiments and relative to Actin control expression.

(C) WCL from naïve (2i) ES and EpiLCs, sorted for cells in the G1 phase and analyzed by WB for levels of SA1 and SA2. Actin serves as a loading control.

(D) Relative expression of Stag1 mRNA by qRT-PCR in FCS- (left panel, n=20) or 2i-grown (right panel, n=19) ES cells upon treatment with si scr or si SA1. Whiskers and boxes indicate all and 50% of values respectively. Central line represents the median.

(E) Cell cycle analysis of Hoechst-stained 2i ES cells after treatment with siScr or siSA1 siRNAs for 24hrs. Shown are the percentages of cells in G1 or G2 phases. These are the same cells that were used for the RNA-sequencing experiments shown in Figure 1F.

(F) MFI of Stag1 protein assessed by IF in 2i ES cells treated with si scr and si SA1. Cells were stained for SA1 and counterstained with DAPI. Data is from the second biological replicate (replicate 1 is in Fig 1E), n>100 cells/condition **** $p<0.0005$.

(G) Enrichment score (ES) plots from GSEA using Naïve or Primed gene sets as in Figure 1G and RNA-seq data from the other two siSA1 treated ES cell samples.

(H) AP+ colonies in ES cells (purple), as a percentage of all colonies (pink and purple) treated with siscr and si SA1. Data are from three independent biological replicates and merged. See Fig 1H for the individual experiments.

(I-K) CRISPR/Cas9 targeting of the C-terminus of endogenous Stag1. (I) Schematic of the targeted locus and location of primers used for genotyping. (J) Shown are four ES cell clones representing integration into both alleles, one allele and a wildtype clone. (K) IF using the v5 tag in a homozygote clonal line treated with dTAG for 16 hrs shows complete loss of the NeonGreen signal.

Figure S2. Stag1 associates with heterochromatin, Related to Figure 2.

(A, B) Confocal IF images in dox-inducible full-length SA1-GFP ES cells (SA1-FL-GFP) stained with (A) GFP and (B) HP1a. Nuclei were counterstained with DAPI. NB, colocalization of SA1-GFP with DAPI-dense foci and HP1a.

(C) Replicate experiment for co-Immunoprecipitation of endogenous SA1 with Smc3, HP1a and Nucleolin in ES cells.

(D) Enrichment of Stag1 ChIP at additional LTR and DNA transposon elements.

(E) WCL from ES treated with the siRNA panel and analyzed by WB for levels of SA1, H3K9me3 and K3K4me3. Tubulin serves as a loading control.

(F) Global chromatin accessibility as detected by DNase I digestion of genomic DNA in siScr or siSA1 ES cells. NB, In two of six biological replicates, we observed increased compaction upon SA1 loss, as is shown here.

Figure S3. Transcription-Regulatory Control of Stag1 in ESC, Related to Figure 3.

(A) Aligned Stag1 transcript variants identified from 5'RACE in Figure 3A, B. Arrows refer to the bands on the RACE gels which were cloned and sequenced. NB, the diversity of skipping events that all result in a functional loss of the 5' end of Stag1.

(B) Over-exposure of the 5'RACE gel shown in Figure 3B to better show the small RACE products.

(C) Close-up of the 5' RACE sequence that identified a new TSS at exon 7 spliced directly to a sequence in trans carrying regulatory elements.

(D) Close-up of the 3' RACE sequence that identified a new alternative TTS in intron 25 (sequence shown in dark blue).

(E) Initial PCR screen in ES 2i and MEFs using various combinations of forward (5') primers (SATS, canonical TSS, Alt exon 1 TSS) and reverse (3') primers (canonical TTS, Alt intron 25 TTS). NB. SATS is only expressed in ES cells; canonical, full-length Stag1 is more expressed in ES compared to MEFs; and the alternative intron 25 TTS is most often expressed with a canonical TSS.

(F) Top, Strategy for mRNA capture of cohesin genes for Pac-Bio long read RNA-sequencing. Bottom, full-length transcripts sequenced on the PacBio platform includes many isoforms already discovered using RACE and PCR cloning methods.

(G) Chromatin Immunoprecipitation of SA1 in ES and EpiLCs.

Figure S4. Fluctuations in Stag1 isoforms skews cell fates, Related to Figure 4.

(A) Relative expression of Stag1 mRNA by qRT-PCR in FCS- (leftmost and rightmost panels, n=7) or 2i-grown (middle panel, n=6) ES cells upon si scr or the si SA1 panel. Whiskers and boxes as before. NB, all siRNAs knockdown Stag1 levels to a similar extent with the exception of esiRNA SATS which reduced Stag1 by ~40-50%.

(B) Percent Spliced In (PSI) calculations based on VAST-Tools analysis of RNA-seq from ES cells treated with the siRNA panel. Data are shown relative to untreated ES cells.

(C) Relative expression of Nanog mRNA by qRT-PCR in FCS-grown ES cells upon si scr or the si SA1 panel (n=13). Whiskers and boxes as before. NB, the modest, but different influence of the 5p and the 3p KDs on Nanog levels.

(D) AP+ colonies in ES cells (purple), as a percentage of all colonies (pink and purple) treated with the siRNA panel. Data are from two independent biological replicates and merged. See Fig 4H for the individual experiments.

Figure S5. Stag1 N-terminus represses heterochromatin and reprograms cells, Related to Figure 5.

(A) Confocal IF images of GFP and H3K9me3 in ES cells treated with siRNA panel. Nuclei were counterstained with DAPI.

(B) Imaris quantification of the volume of H3K9me3 foci per cell from (A). Box plots and statistical analysis were done as before. Data are from two biological replicates, n>50/condition. ****p<0.0005.

(C) Global chromatin accessibility as detected by DNase I digestion of genomic DNA in siScr or siSA1 3p or 5p ES cells. NB. There was no effect on accessibility in the 3p KD but a consistent decompaction observed in the 5p KD.

(D, E) Relative expression of (C) MERVL and (D) genes associated with the 2C-L state by qRT-PCR in 2i-grown ES cells after 72hr of treatment with si scr or the si SA1 panel. Data are represented as mean +/- SEM from n=3 biological replicates. NB, the effect of the 5p KD on the expression of these hallmark 2C-L genes/repeats is much more significant than in the FCS-grown ES cells shown in Figure 5C, D.

(F) Enrichment score (ES) plots from GSEA using 2C gene sets as in Figure 5E and RNA-seq data from the different siRNA treated ES cell samples.

Figure S6. Stag1 deregulates nascent rRNA expression, Related to Figure 6.

(A) WCL from ES cells treated with the siRNA panel and analyzed by WB for global levels of Nucleolin. Tubulin serves as a loading control.

(B) Top, cartoon of the consensus Mus musculus ribosomal DNA (rDNA) (GenBank: BK000964.3), showing the ribosomal genes and the intergenic spacer (IGS) region which contains several SINE elements (Red, B2_Mm2; Green, B3). Bottom, Stag1 ChIP replicates aligned to this region. NB, possible Stag1 binding to the B3 elements in the IGS.

REFERENCES

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS ONE *5*, e10531.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol. *19*, 229–244.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Barrington, C., Georgopoulou, D., Pezic, D., et al. (2019) Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. Nature Communications.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270–1284.

Bisht, K.K., Daniloski, Z., and Smith, S. (2013). SA1 binds directly to DNA through its unique AT-hook to promote sister chromatid cohesion at telomeres. J. Cell. Sci. *126*, 3493–3503.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Bonev, B., Cohen, N.M., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557.e1–557.e24.

Borsos, M., and Torres-Padilla, M.-E. (2016). Building up the nucleus: nuclear organization in the establishment of totipotency and pluripotency during mammalian development. Genes Dev. *30*, 611–621.

Buszczak, M., Signer, R.A.J., and Morrison, S.J. (2014). Cellular Differences in Protein Synthesis Regulate Tissue Homeostasis. Cell *159*, 242–251.

Canudas, S., and Smith, S. (2009). Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. J Cell Biol *187*, 165–173.

Cardozo Gizzi, A.M., Cattoni, D.I., Fiche, J.-B., Espinola, S.M., Gurgo, J., Messina, O., Houbron, C., Ogiyama, Y., Papadopoulos, G.L., Cavalli, G., et al. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. Molecular Cell *74*, 212–222.e215.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Cuadrado, A., and Losada, A. (2020). Specialized functions of cohesins STAG1 and STAG2 in

3D genome architecture. Curr. Opin. Genet. Dev. *61*, 9–16.

Cuadrado, A., Giménez-Llorente, D., Kojic, A., Rodriguez-Corsino, M., Cuartero, Y., Martín-Serrano, G., Gomez-Lopez, G., Marti-Renom, M.A., and Losada, A. (2019). Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and Polycomb Domains in Embryonic Stem Cells. Cell Rep *27*, 3500–3510.e3504.

Cuartero, S., Weiss, F.D., Dharmalingam, G., Guo, Y., Ing-Simmons, E., Masella, S., Robles-Rebollo, I., Xiao, X., Wang, Y.-F., Barozzi, I., et al. (2018). Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. Nat Immunol *19*, 932–941.

Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell *164*, 1110–1121.

Deniz, Ö., Ahmed, M., Todd, C.D., Rio-Machin, A., Dawson, M.A., and Branco, M.R. (2020). Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. Nature Communications *11*, 3506–3514.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R.V., Milagre, I., et al. (2016). MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. Cell Rep *17*, 179–192.

Feng, G., Tong, M., Xia, B., Luo, G.Z., Wang, M., Xie, D., Wan, H., Zhang, Y., Zhou, Q., and Wang, X.J. (2016). Ubiquitously expressed genes participate in cell-specific functions via alternative promoter usage. EMBO Rep. *17*, 1304–1313.

Finn, E.H., Pegoraro, G., Brandão, H.B., Valton, A.-L., Oomen, M.E., Dekker, J., Mirny, L., and Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. Cell *176*, 1502–1515.e1510.

Fiszbein, A., and Kornblihtt, A.R. (2017). Alternative splicing switches: Important players in cell differentiation. Bioessays *39*, 1600157.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell *147*, 132–146.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., et al. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693–707.e14.

1298 Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and
1299 Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally
1300 regulated IFNG locus. Nature *460*, 410–413.

1301 Hara, K., Zheng, G., Qu, Q., Liu, H., Ouyang, Z., Chen, Z., Tomchick, D.R., and Yu, H. (2014).
1302 Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric
1303 cohesion. Nature Publishing Group *21*, 864–870.

1304 Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W., Wike, C.L., Weaver,
1305 B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and
1306 human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat.
1307 Genet. *49*, 925–934.

1308 Horsfield, J.A., Anagnostou, S.H., Hu, J.K.-H., Cho, K.H.Y., Geisler, R., Lieschke, G., Crosier,
1309 K.E., and Crosier, P.S. (2007). Cohesin-dependent regulation of Runx genes. Development
1310 *134*, 2639–2649.

1311 Hung, S.S.C., Wong, R.C.B., Sharov, A.A., Nakatake, Y., Yu, H., and Ko, M.S.H. (2013).
1312 Repression of global protein synthesis by Eif1a-like genes that are expressed specifically in the
1313 two-cell embryos and the transient Zscan4-positive state of embryonic stem cells. DNA Res *20*,
1314 391–402.

1315 Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-
1316 Terrones, D., Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.-E. (2015). Early
1317 embryonic-like cells are induced by downregulating replication-dependent chromatin assembly.
1318 Nature Publishing Group *22*, 662–671.

1319 Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.-E.
1320 (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early
1321 mouse embryo. Nat. Genet. *49*, 1502–1510.

1322 Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier,
1323 C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene
1324 expression and chromatin architecture. Nature *467*, 430–435.

1325 Kline, A.D., Moss, J.F., Selicorni, A., Bisgaard, A.-M., Deardorff, M.A., Gillett, P.M., Ishman,
1326 S.L., Kerr, L.M., Levin, A.V., Mulder, P.A., et al. (2018). Diagnosis and management of Cornelia
1327 de Lange syndrome: first international consensus statement. Nat. Rev. Genet. *19*, 649–666.

1328 Kojic, A., Cuadrado, A., De Koninck, M., Giménez-Llorente, D., Rodriguez-Corsino, M., Gomez-
1329 Lopez, G., Le Dily, F., Marti-Renom, M.A., and Losada, A. (2018). Distinct roles of cohesin-SA1
1330 and cohesin-SA2 in 3D chromosome organization. Nature Publishing Group *25*, 496–504.

1331 Kresoja-Rakic, J., and Santoro, R. (2019). Nucleolus and rRNA Gene Chromatin in Early
1332 Embryo Development. Trends Genet. *35*, 868–879.

1333 Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard,
1334 D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role
1335 for phase separation in heterochromatin. Nature *547*, 236–240.

1336 Lehalle, D., Mosca-Boidron, A.-L., Begtrup, A., Boute-Benejean, O., Charles, P., Cho, M.T.,
1337 Clarkson, A., Devinsky, O., Duffourd, Y., Duplomb-Jego, L., et al. (2017). STAG1 mutations
1338 cause a novel cohesinopathy characterised by unspecific syndromic intellectual disability. J Med
1339 Genet *54*, 479–488.

1340 Leiserson, M.D.M., Vandin, F., Wu, H.-T., Dobson, J.R., Eldridge, J.V., Thomas, J.L.,
1341 Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., et al. (2015). Pan-cancer network analysis
1342 identifies combinations of rare somatic mutations across pathways and protein complexes. Nat.
1343 Genet. *47*, 106–114.

1344 Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L.,
1345 Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., et al. (2020). The structural basis for
1346 cohesin-CTCF-anchored loops. Nature *578*, 472–476.

1347 Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A.,
1348 Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with
1349 endogenous retrovirus activity. Nature *487*, 57–63.

1350 Martin, C., Beaujean, N., Brochard, V., Audouard, C., Zink, D., and Debey, P. (2006). Genome
1351 restructuring in mouse embryos during reprogramming and early development. Dev. Biol. *292*,
1352 317–332.

1353 Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A., and Boettiger, A.N. (2019).
1354 Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature *568*, 1–24.

1355 Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006).
1356 Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell *10*,
1357 105–116.

1358 Mishiro, T., and Tsutsumi, S. (2009). Architectural roles of multiple chromatin insulators at the
1359 human apolipoprotein gene cluster. Embo J. *28*, 1234–1245.

1360 Misulovin, Z., Schwartz, Y.B., Li, X.-Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen,
1361 M.B., Pirrotta, V., Biggin, M.D., et al. (2007). Association of cohesin and Nipped-B with
1362 transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma *117*, 89–
1363 102.

1364 Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J.,
1365 Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive
1366 genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.
1367 Nat. Genet. *34*, 267–273.

1368 Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb,
1369 M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-
1370 specific protein degradation. Nature Chemical Biology *14*, 1–16.

1371 Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N.,
1372 Wingett, S., Fraser, P., and Tanay, A. (2017). Cell-cycle dynamics of chromosomal organization
1373 at single-cell resolution. Nature *547*, 61–67.

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Längst, G. (2010). Initial genomics of the human nucleolus. PLoS Genet. *6*, e1000889.

Novo, C.L., Tang, C., Ahmed, K., Djuric, U., Fussner, E., Mullin, N.P., Morgan, N.P., Hayre, J., Sienerth, A.R., Elderkin, S., et al. (2016). The pluripotency factor Nanog regulates pericentromeric heterochromatin organization in mouse embryonic stem cells. Genes Dev. *30*, 1101–1115.

Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., Joti, Y., Tomita, M., Hibino, K., Kanemaki, M.T., et al. (2017). Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. Molecular Cell *67*, 282–293.e287.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., and Dunker, A.K. (2003). Predicting intrinsic disorder from amino acid sequence. Proteins *53 Suppl 6*, 566–572.

Orgil, O., Matityahu, A., Eng, T., Guacci, V., Koshland, D., and Onn, I. (2015). A conserved domain in the scc3 subunit of cohesin mediates the interaction with both mcd1 and the cohesin loader complex. PLoS Genet. *11*, e1005036.

Padeken, J., and Heun, P. (2014). Nucleolus and nuclear periphery: velcro for heterochromatin. Curr. Opin. Cell Biol. *28*, 54–60.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422–433.

Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell *174*, 391–405.e19.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744–757.e24.

Rao, S.S.P., Huang, S.-C., Hilaire, B.G.S., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305–309.e324.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Remeseiro, S., Cuadrado, A., Carretero, M., Martínez, P., Drosopoulos, W.C., Cañamero, M., Schildkraut, C.L., Blasco, M.A., and Losada, A. (2012a). Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. Embo J. *31*,

2076–2089.

Remeseiro, S., Cuadrado, A., López, G.G., Pisano, D.G., and Losada, A. (2012b). A unique role of cohesin-SA1 in gene regulation and development. Embo J. *31*, 2090–2102.

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell *160*, 1145–1158.

Rodriguez, J., Ren, G., Day, C.R., Zhao, K., Chow, C.C., and Larson, D.R. (2019). Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. Cell *176*, 213–226.e218.

Romero-Pérez, L., Surdez, D., Brunet, E., Delattre, O., and Grünewald, T.G.P. (2019). STAG Mutations in Cancer. Trends Cancer *5*, 506–520.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*, eaar3958–13.

Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. Proc. Natl. Acad. Sci. U.S.a. *107*, 10514–10519.

Schlesinger, S., and Meshorer, E. (2019). Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. Dev. Cell *48*, 135–150.

Schwalie, P.C., Ward, M.C., Cain, C.E., Faure, A.J., Gilad, Y., Odom, D.T., and Flicek, P. (2013). Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. Genome Biol. *14*, R148–15.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., H Haering, C., Mirny, L., and Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51–56.

Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. *23*, 2066–2077.

Soardi, F.C., Machado-Silva, A., Linhares, N.D., Zheng, G., Qu, Q., Pena, H.B., Martins, T.M.M., Vieira, H.G.S., Pereira, N.B., Melo-Minardi, R.C., et al. (2017). Familial STAG2 germline mutation defines a new human cohesinopathy. NPJ Genom Med *2*, 7–11.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Rudan, M.V., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. Embo J. *32*, 3119–3129.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017).

1449 Phase separation drives heterochromatin domain formation. Nature *547*, 241–245.

1450 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,
1451 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment
1452 analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc
1453 Natl Acad Sci USA *102*, 15545–15550.

1454 Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A.,
1455 Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., et al. (2017). An atlas of
1456 alternative splicing profiles and functional associations reveals new regulatory programs and
1457 genes that simultaneously express multiple major isoforms. Genome Res. *27*, 1759–1768.

1458 Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur,
1459 S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain
1460 architecture. CellReports *10*, 1297–1309.

1461 Viny, A.D., Bowman, R.L., Liu, Y., Lavallée, V.-P., Eisman, S.E., Xiao, W., Durham, B.H.,
1462 Navitski, A., Park, J., Braunstein, S., et al. (2019). Cohesin Members Stag1 and Stag2 Display
1463 Distinct Roles in Chromatin Accessibility and Topological Control of HSC Self-Renewal and
1464 Differentiation. Cell Stem Cell *25*, 682–696.e688.

1465 Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with
1466 single-cell genomics. Nature Publishing Group *34*, 1145–1160.

1467 Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae,
1468 G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by
1469 CCCTC-binding factor. Nature *451*, 796–801.

1470 Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee,
1471 T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-
1472 enhancers at key cell identity genes. Cell *153*, 307–319.

1473 Winters, T., McNicoll, F., and Jessberger, R. (2014). Meiotic cohesin STAG3 is required for
1474 chromosome axis formation and sister chromatid cohesion. Embo J. *33*, 1256–1270.

1475 Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S.,
1476 Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and
1477 chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins.
1478 Embo J. *36*, 3573–3599.

1479 Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF
1480 interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent
1481 insulation activity. Mol. Cell. Biol. *31*, 2174–2183.

1482 Yuan, B., Neira, J., Pehlivan, D., Santiago-Sim, T., Song, X., Rosenfeld, J., Posey, J.E., Patel,
1483 V., Jin, W., Adam, M.P., et al. (2019). Clinical exome sequencing reveals locus heterogeneity
1484 and phenotypic variability of cohesinopathies. Genet Med *21*, 663–675.

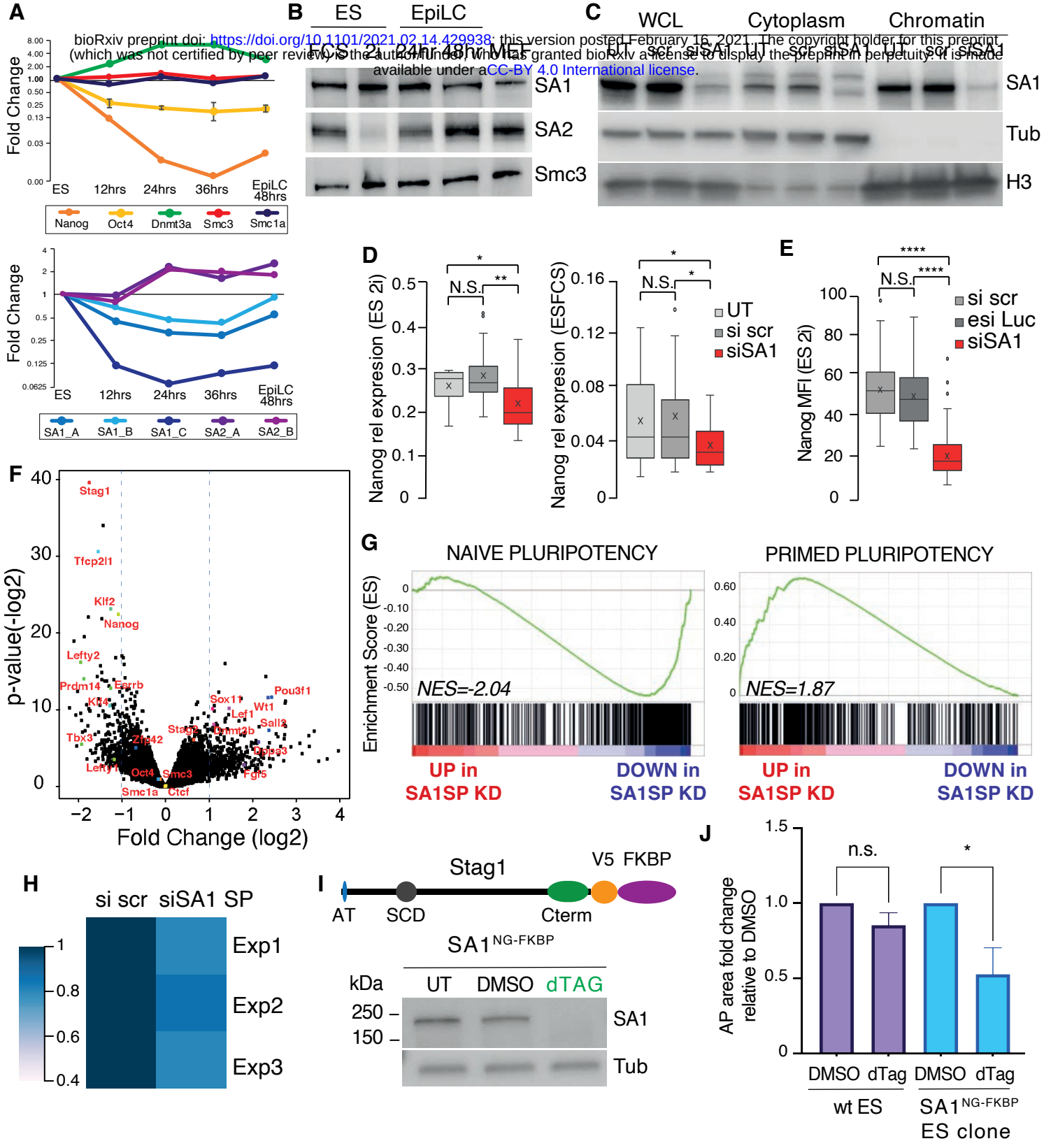
1485 Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator
1486 to subnuclear sites, suggesting shared insulator mechanisms across species. Molecular Cell *13*,

291–298.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U.S.a. *111*, 996–1001.

Figure 1.

Pezic et al.



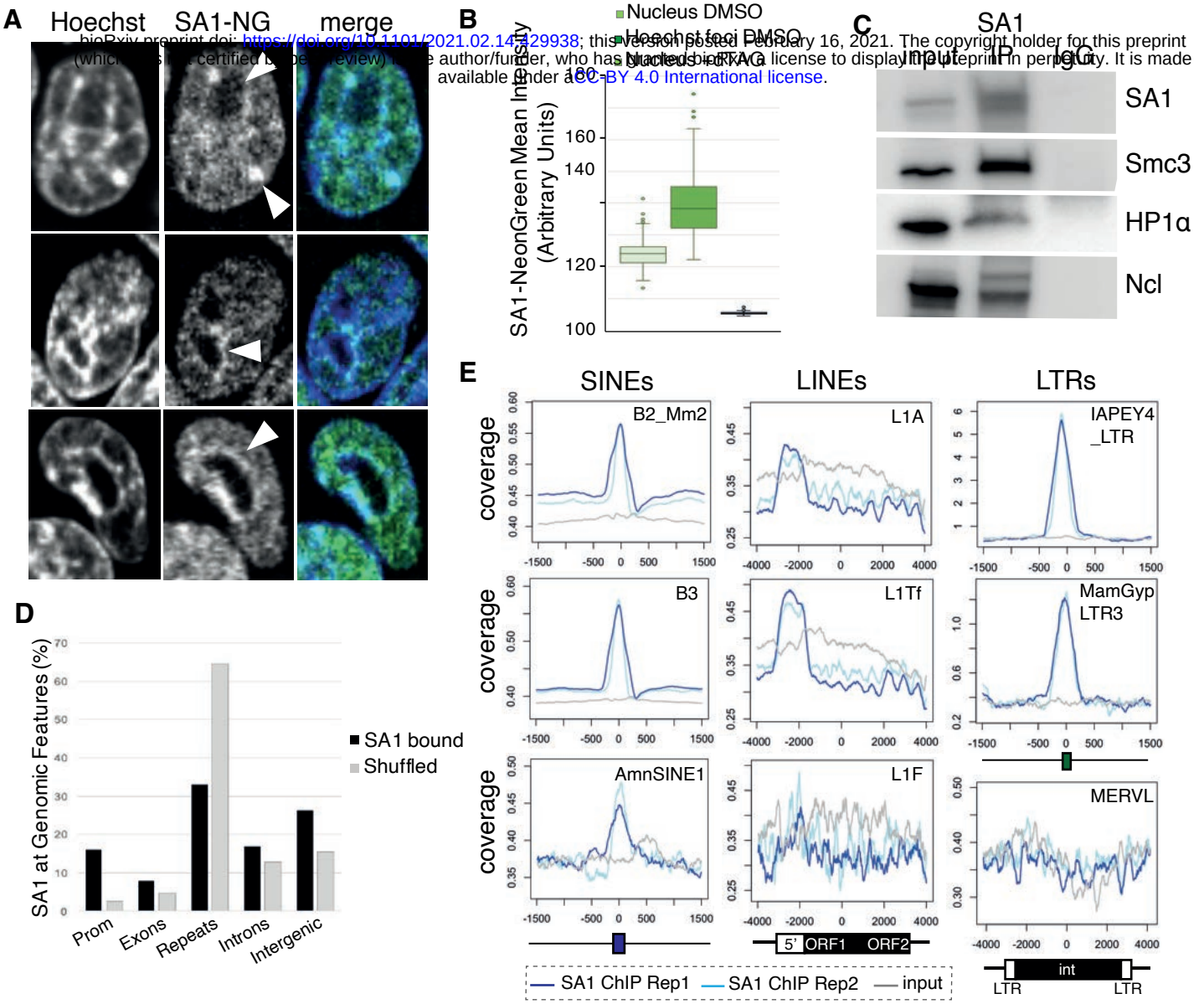


Figure 3. *Pezic et al.*