Novel Chromatin Insulating Activities Uncovered upon Eliminating Known Insulators *in vivo*

Havva Ortabozkoyun[1,2*†], Pin-Yao Huang[1,2*], Hyunwoo Cho[3,4,5], Aristotelis Tsirigos[3,4,6], Esteban Mazzoni[7], Danny Reinberg[1,2†]

[1]Howard Hughes Medical Institute, NYU Grossman School of Medicine, New York, NY, USA.

[2]Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY, USA.

[3]Department of Pathology, NYU Grossman School of Medicine, New York, NY, USA.

[4]Applied Bioinformatics Laboratories, NYU Grossman School of Medicine, New York, NY, USA.

[5]Department of Medicine, Division of Precision Medicine, NYU Grossman School of Medicine, New York, NY, USA.

[6]Institute for Computational Medicine, NYU Grossman School of Medicine, New York, NY, USA.

[7]Department of Biology, New York University, New York, NY, USA.

*These authors contributed equally to this work.

[†] To whom correspondence should be addressed: Danny.Reinberg@nyulangone.org, Havva.Ortabozkoyun@nyulangone.org

Present Address (P.Y.H.): Ionis Pharmaceuticals, Carlsbad, CA, USA.
Present Address (H.C.): Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

Abstract

CCCTC-binding factor (CTCF) and MAZ are recognized insulators required for shielding repressed posterior genes from active anterior genes within the *Hox* clusters during motor neuron (MN) differentiation. CTCF and MAZ interact independently with cohesin and regulate three-dimensional genome organization. Here, we followed cohesin re-location upon CTCF and MAZ depletion in mouse embryonic stem cells (mESCs) to identify novel insulators. Cohesin relocated to DNA motifs for various transcription factors, including PATZ1 and other zinc finger proteins (ZNFs). Moreover, PATZ1 and ZNFs co-localized with CTCF, MAZ, and cohesin with apparent overlapping specificity as dictated by the site to be insulated. Similar to CTCF and MAZ, PATZ1 interacted with RAD21. *Patz1* KO mESCs exhibited altered global gene expression. While the absence of MAZ impacts anterior CTCF-boundaries as shown previously[1], *Patz1* KO led to de-repression of posterior *Hox* genes, resulting in cervicothoracic transformation of motor neuron (MN) fate during differentiation. These findings point to a varied, combinatorial binding of known and newly defined accessory factors as being critical for positional identity and cellular fate determination during differentiation.

Main Text

Introduction

In addition to the genetic information and chromatin structure, the three-dimensional (3D) spatial organization of the mammalian genome contributes to the temporal regulation of the transcriptome, which is key to the appropriate development of multicellular organisms[2,3]. Chromatin insulators have emerged as one of the main components of spatial genome organization[4,5]. Although multiple insulator proteins exist in *Drosophila*[5-7], CTCF has been the major insulator protein recognized in mammals for decades[8,9]. In vertebrate genomes, CTCF and the ring-shaped cohesin complex are essential for maintaining genome organization and their loss disrupts the structure of topologically associated domains (TAD)[10,11]. Chromatin loops are the building blocks of the 3D genome organization and their formation arises from the generally accepted model of cohesin-mediated loop extrusion[12,13]. This dynamic process is ultimately blocked and stabilized when cohesin encounters CTCF proteins bound to two convergently oriented CTCF DNA-binding sites, which form the loop anchors[11,14-16]. While CTCF is believed to be the main insulator in mammals[8], a number of additional factors, such as MAZ (and others, see below), exhibit functional roles in insulation, gene expression, and spatial genome organization[1,17].

Hox gene expression patterns emanating from the four *Hox* clusters in mice are regulated in a spatiotemporal manner during development and determine the anterior-posterior identity of cells[18]. During cervical motor neuron (MN) differentiation, MAZ functions similarly to CTCF in maintaining the integrity of antagonistic chromatin domain borders in the *Hox* clusters, a process that delimits active from repressed genes to ensure the positional identity of cells[1]. However, how

different chromatin boundaries are regulated and how specificity is attained at these boundaries during cellular fate determination is not yet clear. While CTCF and MAZ are required for the integrity of several chromatin boundaries at *Hox* clusters[1,19] and possibly other boundaries[20], it is not clear whether these factors are sufficient for cell fate specification. Furthermore, the molecular mechanisms by which additional putative insulation co-/factors contribute to gene regulation is still poorly understood. As the existence of other insulation factors in mammals are now becoming clearer, a systematic analysis of these candidates is instrumental to comprehending the appropriate formation and functioning of the 3D-genome.

Notably, when CTCF is degraded in embryonic stem cells (ESCs), cohesin, as represented by its RAD21 component, remains bound to chromatin but at new regions[16,21]. Interestingly, cohesin relocation in the absence of CTCF occurs at sites containing the MAZ motif[1], suggesting that MAZ functions as a barrier to cohesin. Approximately 20% of MAZ binding sites overlap with those of CTCF genome-wide in mouse ESCs (mESCs) and MNs[1], yet, whether MAZ functions alone or with other proteins in insulation is unclear. As CTCF and MAZ interact independently with cohesin on chromatin, we anticipated that other putative insulation factors would also exhibit chromatin-based cohesin interaction. We reasoned that the elimination of both CTCF- and MAZ-mediated barriers might expose cohesin relocation to the DNA-binding sites of other potential insulation factors.

To find additional insulators, we performed a genomics analysis to follow cohesin re-location in the absence of CTCF and MAZ and identified DNA motifs corresponding to transcription factors, including PATZ1 and various other zinc finger proteins (ZNFs). These proteins often co-localize

with the cohesin component RAD21, on chromatin. Our genomic analyses performed in several cell types coupled with a series of functional analyses of PATZ1 during differentiation led to the discovery that PATZ1 functions as a novel insulation factor and is critical to determining genomic boundaries and therefore, cellular fate. Our findings point to the necessity of different factors that function at distinct chromatin boundaries as well as to their potential, combinatorial action at a given boundary.

Results

Cohesin re-localized to MAZ binding sites in the absence of CTCF. While previous studies demonstrated cohesin redistribution upon the loss of CTCF[16,21,22], and that the DNA motifs of such re-localization resemble that of MAZ[1], the chromatin binding pattern of MAZ upon the loss of CTCF has not been reported. Thus, we further analyzed the chromatin binding patterns of the three recognized architectural proteins, CTCF, MAZ, and cohesin[1-3,10,11,17] by performing ChIP-seq in the presence (WT) and absence (ΔCTCF) of CTCF in the CTCF-degron mESCs (**Fig. 1**). As expected, cohesin re-localized in ΔCTCF mESCs (**Fig. 1a**), as reported earlier[16]. Moreover, the RAD21 chromatin-binding pattern co-localized with that of MAZ (**Fig. 1b**). In order to focus on the movement of cohesin as a function of the presence of CTCF, we categorized the ChIP-seq data of RAD21, CTCF, and MAZ into three clusters based on the cohesin signal in WT versus ΔCTCF mESCs: Cluster 1 (n=7,881), Cluster 2 (n=12,451), and Cluster 3 (n=10,018) (**Fig. 1c**). Notably, the RAD21 signal detected in ΔCTCF mESCs (Cluster 3) correlated with MAZ binding, which was also evident in the WT case (**Fig. 1c-e**). When all clusters were compared (**Fig. 1c)**, the co-localization of CTCF and RAD21 (Cluster 1 and 2) in WT mESCs resembled that of MAZ and

RAD21 in ΔCTCF mESCs (Cluster 3). These results suggest that MAZ may exhibit a barrier function at loci other than those defined in the *Hox* clusters[1].

MAZ functions as a barrier at sites of re-localized cohesin binding. As CTCF functions as a barrier to cohesin movement in loop-extrusion models[12,13,23,24] and MAZ binding overlaps that of cohesin in ΔCTCF mESCs (**Fig. 1**), we speculated that MAZ may act as a "secondary" barrier in the absence of CTCF. To test the functional impact of MAZ in relation to cohesin, we generated a *Maz* KO of both MAZ isoforms in the CTCF-degron mESCs (**Fig. 2a**). The *Maz* KO maintained its morphological integrity and showed comparable levels of CTCF, RAD21, and OCT4, a marker for ESCs, relative to the WT case (**Fig. 2a**). While the loss of CTCF in conjunction with that of *Maz* (ΔCTCF/ΔMAZ) was ineffectual with respect to RAD21 protein levels (**Fig. 2a**), RAD21 now exhibited a markedly reduced occupancy at its sites of re-localization in ΔCTCF, as shown in Cluster 3 (**Fig. 2b**, see **Extended Data Fig. S1** for all clusters). This phenomenon is exemplified by comparing the re-localized RAD21 at *Nat9* and *Dnajb12* gene loci in ΔCTCF versus ΔCTCF/ΔMAZ (**Fig. 2c-d**). Thus, MAZ appears to be a critical barrier for cohesin movement in the absence of the major insulator protein, CTCF.

Uncovering novel RAD21 re-localization to DNA-binding sites recognized by transcription factors including ZNFs, such as PATZ1. Given the common characteristics of CTCF[8,25-27] and MAZ[1,17] in insulation and the existence of multiple insulator proteins in *Drosophila*[5], we inspected cohesin re-localization in the absence of both CTCF and MAZ to identify novel insulation factors in addition to MAZ. Analysis of RAD21 re-localized regions in ΔCTCF/ΔMAZ mESCs revealed the DNA-motifs corresponding to various transcription factors, including the ZNFs (**Fig. 2e**, see

Extended Data Fig. S2 for a detailed list). Notably, amongst these ZNFs is VEZF1, which was shown previously to function as an insulator[28], thereby substantiating our strategy for identifying other novel insulation factors.

We next sought the possible co-localization of these ZNFs with RAD21 by analyzing the ChIP-seq data available from public databases such as ENCODE and NCBI (**Table S1**) in human HEK293 and HepG2 cell lines. Amongst the top candidates, PATZ1 (POZ/BTB and AT-Hook-Containing Zinc Finger Protein 1, see **Fig 4a**), also known as MAZ-related factor (MAZR), contains six zinc fingers that exhibit a strong homology to those of MAZ and binds *in vitro* to G-rich sequences that resemble the MAZ motif[1,17,29]. These characteristics of PATZ1 imply that it might function similarly to MAZ in cells and thus, PATZ1 was chosen for assessing the functional roles of ZNFs in the dynamics of cohesin chromatin binding.

PATZ1 and other ZNFs co-localize with RAD21 on chromatin and at loop anchors. To begin to ascertain if PATZ1 functions to anchor chromatin in a manner similar to that of CTCF and MAZ, we first cross-compared the ChIP-seq peaks of RAD21, CTCF, and PATZ1 in HEK293 and HepG2, and clustered the overlapping regions into 3 groups. Cluster 1 consists of 10,550 and 10,262 sites, which are bound by RAD21, CTCF, and PATZ1 in HEK293 and HepG2, respectively (**Fig. 3a, b** and **Extended Data Fig. S3a, b**). Cluster 2 represents RAD21 and CTCF co-localized regions, including 14,446 and 31,799 sites in HEK293 and HepG2, respectively (**Fig. 3a, b** and **Extended Data Fig. S3a, b**). Interestingly, in Cluster 3, we observed 4,336 and 8,823 sites that are bound by RAD21 and PATZ1 in the absence of CTCF signals, in HEK293 and HepG2, respectively (**Fig. 3a, b** and **Extended Data Fig. S3a, b**). These results suggested that indeed,

PATZ1 harbors the intrinsic potential of anchoring cohesin on chromatin in different cell types, independently of CTCF.

We next aligned the ChIP-seq signals of MAZ and other ZNFs, ZNF263, ZNF341, and ZNF467 (see **Extended Data Fig. S4** for motif alignment), based on the aforementioned clusters using RAD21, CTCF, and PATZ1. Depending on the cluster analyzed, we observed a similar distribution of the signals amongst the ZNFs compared, along with some noteworthy differences (**Fig. 3b**). While Cluster 1 indicates co-localization of ZNFs and MAZ at RAD21, CTCF, and PATZ1 co-occupied regions, Cluster 2 indicates varying signal intensities for ZNFs at RAD21 and CTCF co-occupied regions. Lastly, Cluster 3 shows co-localization of all ZNFs, but not CTCF, at RAD21 and PATZ1 co-occupied regions. For example, ZNF263 and ZNF467 exhibited higher signal intensities in Cluster 2 than PATZ1, MAZ, and ZNF341 and importantly, Cluster 3 displayed less signal intensity for CTCF (**Fig. 3b**), suggesting that each zinc finger protein has distinctive DNA-binding preferences, although their conserved motifs were first described as G-rich sequences (**Fig. 2e** and **Extended Data Fig. S2, S4**). At the level of individual loci, we observed that the cohesin peaks co-localized with the ZNF peaks in different combinations and as stated above, sometimes in the absence of CTCF binding (**Extended Data Fig. S3c-f**). These results altogether support the hypothesis that the ZNF proteins might serve as stable anchors for cohesin, apart from CTCF. Importantly, PATZ1 interacted with RAD21 in mESCs (**Fig. 3c**), similar to CTCF and MAZ[1,17,30], consistent with PATZ1 exerting a barrier role in stopping cohesin extrusion. In accordance, analysis of Hi-C data in HepG2 cells revealed that PATZ1 was found at either one or both sides of ~40% of the loop anchors, and stable loops were detected between PATZ1 binding peaks at regions devoid of CTCF binding (**Fig. 3d, e**).

To further evaluate the chromatin binding patterns of PATZ1 and the other ZNF proteins with respect to cohesin anchoring sites, we analyzed the ChIP-seq peaks of RAD21, PATZ1, ZNF263, ZNF341, and ZNF467 in HEK293 cells (**Extended Data Fig. S5a**). Overlapping regions were grouped into 4 clusters based on the number of ZNFs co-localized with RAD21 (**Fig. 3f**). Interestingly, the RAD21 peaks co-localized with various combinations of the ZNF protein peaks (**Fig. 3f**), suggesting that the ZNFs might cooperatively or uniquely anchor cohesin on chromatin at different loci and thereby establish locus-specific chromatin boundaries. Moreover, as the ZNFs are distinctively expressed across different tissues and cell types (**Extended Data Fig. S5b, c**), we speculate that specific cell types utilize unique or varying combinations of these ZNFs to establish cell-specific chromatin boundaries during differentiation and/or development[14,31]. Overall, these findings support a role for PATZ1 and other ZNF proteins in delimiting chromatin boundaries in specific loci and cell types.

PATZ1 co-localizes with RAD21, CTCF, and MAZ on chromatin in mESCs. To further investigate the functional role of PATZ1 (**Fig. 4a**), we generated multiple *Patz1* KO lines in mESCs via CRISPR/Cas9 (**Extended Data Fig. S6a**). As expected, specific PATZ1 protein signals were lost in the *Patz1* KO lines, while the levels of RAD21, CTCF, and OCT4 were similar to those of WT (**Extended Data Fig. S6a**), as observed previously for the *Maz* KO[1]. Interestingly, the levels of MAZ were consistently decreased in the *Patz1* KO clones (**Extended Data Fig. S6a),** suggesting that PATZ1 regulates MAZ expression. To determine the genomic binding loci of PATZ1 in mESCs, we first performed several ChIP-seq experiments using commercially available antibodies, which unfortunately were ineffective for evaluating ChIP-seq data (data not shown). To address this issue, we introduced 3xFLAG-HA tagged PATZ1 (FH-PATZ1) into the *Patz1* KO

cells using the PiggyBac transposon system, generating the FH-PATZ1 cell line. The ectopic expression of 3xFLAG-HA-PATZ1 restored both PATZ1 and MAZ levels to the *Patz1* KO, while the levels of RAD21, CTCF, and OCT4 were not altered (**Fig. 4b**). We then determined the genomic binding sites of FH-PATZ1 in mESCs via ChIP-seq using an anti-HA antibody, which gave rise to FH-PATZ1-specific signals as evidenced by the absence of HA ChIP-seq signals in the *Patz1* KO background (**Extended Data Fig. S6b**). Similar to our observations in HEK293 and HepG2 cells, we found that approximately 25% of PATZ1 peaks co-localized with RAD21 either with or without CTCF in mESCs, and that MAZ ChIP-seq signals align comparatively with PATZ1 peaks (**Fig. 4c, d**). Similarly, PATZ1 and MAZ form combinatorial binding patterns near RAD21 peaks (**Fig. 4e, f**). While the binding patterns of PATZ1 and MAZ were similar at the example gene loci shown in **Fig. 4e**, other loci exhibited differing binding patterns (**Fig. 4f**). These data further support that PATZ1, MAZ, and perhaps other ZNFs contribute to cohesin anchoring on chromatin, and that these events are conserved in mouse and human.

Effects of PATZ1 depletion on cohesin binding and the transcriptome. To further evaluate the molecular function of PATZ1 with respect to cohesin chromatin-binding, we performed ChIP-seq with specific antibodies against RAD21, CTCF, and MAZ in WT and *Patz1* KO mESCs. 20% of RAD21 peaks and 45% of MAZ peaks showed reduced binding upon PATZ1 depletion, while less than 1% of the peaks showed an increase; the reduction in MAZ peaks may be attributable at least in part to the accompanying reduction in MAZ expression in the absence of PATZ1 (**Fig. 4b**). On the other hand, 2.6% and 4.5% of CTCF peaks were increased and decreased, respectively (**Extended Data Fig. S6c, d**). By comparing the ChIP-seq signals of these factors at the decreased RAD21 peaks, we found that the CTCF signals were minimally altered, whereas MAZ and PATZ1

signals were reduced concomitantly (**Fig. 5a**). Interestingly, inspection of individual loci revealed that the reduction in RAD21 peaks in the absence of PATZ1 could occur in the presence of CTCF binding and irrespective of MAZ binding (**Fig. 5b-c** and **Extended Data Fig. S6c-d**), suggesting that MAZ and/or PATZ1 facilitates cohesin anchoring independently of CTCF.

We next examined the effects of PATZ1 depletion on genome-wide gene expression in mESCs via RNA-seq. The loss of PATZ1 resulted in the differential expression of 1,146 genes (**Fig. 5d**), 315 of which were similarly impacted in the *Maz* KO (**Extended Data Fig. S6e**). Importantly, the genes encoding the cohesin loading factors, *Nipbl* and *Mau2*, were expressed at WT levels suggesting that the decreased RAD21 binding observed in the *Patz1* KO was not due to an inherent insufficiency in cohesin binding (**Extended Data Fig. S6f**). Similar to the *Maz* KO, Gene Ontology (GO) analysis of the *Patz1* KO mESCs revealed that several categories related to developmental processes were enriched compared to WT (**Fig. 5e**), suggesting that PATZ1 and MAZ might cooperatively regulate a subset of developmental genes.

Combinatorial binding of insulation factors at *Hox* gene borders is key to proper positional identity in motor neurons. Amongst those ZNFs identified at the sites of re-localized cohesin occupancy (**Fig. 2e**), we further analyzed PATZ1 and its chromatin binding profile in mESCs (**Fig. 4**). After genome-wide analysis of PATZ1 ChIP-seq (**Fig. 4**), we next examined *Hox* gene borders in the context of CTCF, RAD21, PATZ1, and MAZ (**Fig. 6a** and **Extended Data Fig. S7a, b**). While PATZ1 followed a binding pattern similar to that of MAZ at the *Hox* clusters (**Fig 6a**, see also **Extended Data Fig. S7a, b**), we noted a ChIP-seq peak of PATZ1 at *Hoxa7|9*. We speculated that PATZ1 may be determinant to the integrity of the *Hoxa7|9* border, as opposed to the

requirement of MAZ at the *Hoxa5|6*, as described previously[1]. Thus, we differentiated mESCs into cervical and thoracic MNs using established protocols[32,33] to foster the formation of a *Hoxa5|6* border and a caudal *Hoxa7|9* border, respectively (**Fig. 6b-c** and **Extended Data Fig. S7c**). Strikingly, the *Patz1* KO led to de-repression of *Hoxa9* and *Hoxa10* in cervical MNs, as compared to WT (**Fig. 6c-d, g**, see **Fig. 6e-g** for other *Hox* clusters). In accordance, the *Patz1* KO appeared to increase *Hoxa7-10* expression in the thoracic MNs, as compared to WT (**Fig. 6c**). The loss of PATZ1 in cervical MNs resulted in the differential regulation of ~3800 genes (padj < 0.01) with an enrichment of GO categories including cellular division, differentiation, and development processes (**Fig. 6h-i**).

Thus, when the *Patz1* KO and *Maz* KO ESCs were differentiated into cervical MNs, the *Patz1* KO impacted the *Hoxa7|9* border, while the *Maz* KO led to de-repression of earlier *Hox* genes, *i.e.* the *Hoxa5|6* border, as shown previously[1] (**Fig. 6g**). Moreover, our analysis of the *HoxC* and *HoxD* clusters pointed to PATZ1 binding preceding the *Hox9* genes (**Extended Data Fig. S7a-b**), and the loss of PATZ1 led to de-repression of *Hoxc9* and *Hoxd8-Hoxd9* (**Fig. 6e-f**), respectively. Importantly, the *Hox* genes retained basal levels of expression in the *Patz1* KO mESCs, similar to that of WT mESCs (**Extended Data Fig. S7d**), suggesting that PATZ1 does not directly modulate the expression of posterior *Hox* genes, but rather potentiates the establishment of the chromatin boundary upon differentiation. These results indicate that PATZ1 functions as a factor regulating the *Hoxa7|9* border, which is critical to the determination of rostro-caudal MN identity and notably, that PATZ1 and MAZ regulate distinct *Hox* gene borders. Thus, the differential roles of PATZ1 and MAZ at the *Hox* clusters underscore the existence of a combinatorial code of insulation factors that impact positional and/or cellular identity (**Fig. 7**).

Discussion

Cohesin adopts multi-step conformational changes to mediate DNA loop extrusion, which is primarily blocked by CTCF[23,34,35]. As loop-extruding cohesin can travel through chromatin roadblocks larger than its ring size[23,36], a predetermined process for stalling cohesin movement and blocking loop extrusion during development may entail the participation of a group of specific DNA binding factors that independently interact with cohesin in a regulated manner. To understand how cohesin regulates the formation of different boundaries, we analyzed its genomic re-localization upon the loss of both of the known insulators, CTCF and MAZ, with MAZ being a factor recently found to cooperate with CTCF in the establishment of specific boundaries within the *Hox* clusters[1]. The identification of these sites of cohesin re-localization pointed to the potential contribution of different sequence-specific DNA binding proteins (**Fig. 2e** and **Extended Data Fig. S2**), such as PATZ1, and other ZNFs that bind to these sites (**Fig. 3f**). While the acute depletion of CTCF led to an extensive redistribution of cohesin binding sites[16] (**Fig. 1**), the loss of PATZ1 resulted in a reduction in cohesin binding at a subset of sites, some of which exhibited a concomitant reduction in MAZ chromatin binding (**Fig. 5a**).

Upon the loss of CTCF[10] or MAZ[1,17] or their respective motifs[1,19,37], gene expression and genome architecture are impacted (shown schematically in **Fig. 7a** versus **7b**, left). Similarly, the loss of PATZ1 reduced RAD21 binding and impacted gene expression in mESCs, and its binding was found to correlate with loop anchors in HepG2 cells (**Fig. 7b**, left). Deletion of the CTCF motif or the MAZ motif, or the *Maz* KO has been shown to result in *Hoxa7* de-repression, impacting the *Hoxa5|6* CTCF boundary[1,19,37]. In a distinctive manner, the PATZ1 KO resulted in *Hoxa9* de-repression, impacting the *Hoxa7|9* CTCF boundary (**Fig. 7a** versus **7b**, right). The loss of PATZ1

in ESCs resulted in differential gene expression of mainly developmental processes (**Fig. 5e**), and revealed its role in CTCF-mediated boundary formation (*i.e. Hoxa7|9*) at the *Hox* clusters during differentiation into MNs (see **Fig. 6** and **Extended Data Fig. S7**). Our study highlights the specific contribution of PATZ1 at the *Hoxa7|9* boundary (**Fig. 6**), as opposed to MAZ that was shown to be critical at the *Hoxa5|6* boundary during cervical MN differentiation[1,19,37] (**Fig. 7**), suggesting that these accessory factors determine the specificity at a given chromatin border and/or CTCF-dependent boundary formation.

That the loss of PATZ1 in MNs gave rise to increased *Hox**a**9, Hox**c**9*, and *Hox**d**9* expression, highlights its regulatory role across multiple clusters (see **Fig. 6** and **Extended Data Fig. S7**). *Hoxc9* upregulation is particularly relevant to MN fate[38,39]. *Hoxc9* induces thoracic spinal MN fate during embryonic development (**Fig. 7b**). Furthermore, direct anterior *Hox* repression is critical to this function[38,39]. In agreement with strong *Hoxc9* de-repression, rostral *Hoxc6* expression was repressed in *Patz1* KO MNs (**Fig. 6e**). Thus, our results indicate that *Patz1* KO MNs adopt a more posterior fate than control MNs upon induction of cervical MN differentiation. In addition to the phenotype detailed in MNs, the study of PATZ1 in relation to mouse anterior-posterior patterning will be informative in the future, as early forelimb patterning requires *Hox9* paralogs[40].

Our studies herein underscore that different chromatin boundaries may contain various factors and that their combinatorial binding with/without CTCF and with cohesin could function to determine gene expression patterns during development. Indeed, given the various expression patterns observed for the ZNFs, MAZ, PATZ1, CTCF, and cohesin across different tissues (**Extended Data Fig. S5b**) or cell types (**Extended Data Fig. S5c**), we envision the possibility that lineage-specific

expression and/or binding of a subset of these novel accessory proteins/insulators sets the specificity of the boundaries. These observations led us to conceptualize a model that includes accessory factors (*i.e.* MAZ, PATZ1, and other ZNFs), in addition to CTCF and cohesin (**Fig. 7**). In this model, a subset of these factors would work in combination with cohesin to foster the specific regulation of chromatin interactions leading to the establishment of dynamic boundaries, critical to genome organization and thus, global gene expression, including those genes within the *HoxA* cluster during cervical MN differentiation (**Fig. 7a**). Consistent with our model, cell-type specific distal gene regulation has been shown to be dependent on the cohesin loading cycle through WAPL[41]. Cohesin turnover on chromatin during cellular differentiation fosters the potential for cell-specific factors to construct combinatorial binding patterns and establish differential chromatin boundaries[41].

Also of note are previous studies in *Drosophila*[4,5,42,43] and in mammals[44,45] that report the presence of multiple "insulators" at a given locus and/or their possible homo/hetero-dimerization properties that may increase the strength of insulation. In addition to these scenarios, the binding of such insulators alone or in a combinatorial fashion, as described in *Drosophila,* may provide unique features to local three-dimensional interactions that inherently impact enhancer-promoter functional contacts and consequently gene expression[4]. Indeed, in addition to the known multiple insulator binding proteins in *Drosophila* (*i.e.* BEAF32, dCTCF, Su(Hw), Zw5, GAF, CP190, MOD(MDG4))[5,6], later reports indicate additional insulators such as Pita and CG7928, the human ortholog of which is ZNF276[46,47]. Of note, several insulators such as dCTCF and Su(Hw) in *Drosophila* contain C2H2 zinc-finger domains implicated in DNA-binding, and GAF, MOD(MDG4) and CP190 in particular have BTB/POZ domains[5,48-50] involved in dimerization[51],

similar to PATZ1 in mammals[52]. Based on our analysis indicating the presence of PATZ1 at loops (**Fig. 3d-e**) and the multimerization properties of CTCF in mammals[53], the BTB/POZ domain of PATZ1 may be critical for homo/heterodimerization during loop formation. Furthermore, the DNA-binding motif of MAZ[1] (**Extended Data Fig. S4a**) and PATZ1 (**Extended Data Fig. S4b**) shows similarity to that of CLAMP, which is involved in dosage compensation by promoting chromatin insulation and long-range chromatin interactions in *Drosophila*[54-56]. Hence, we speculate that many of the candidates identified here through our analysis of the pattern of cohesin re-localization in the absence of currently known insulators, likely represent novel mammalian insulators/co-factors, which awaits further investigation (**Fig. 2e** and **Extended Data Fig. S2**).

While the present study focused on PATZ1 during differentiation of ESCs into MNs and clarified its role in determining MN identity based on *Hox* gene expression, we emphasize the importance of studying the other ZNFs and their combinatorial roles across various boundaries in a given cell type and during cellular fate determination (**Fig. 7**). As the general approaches described in this study provided the DNA motifs associated with the relocation of cohesin in the absence of CTCF and MAZ, we expect our findings to be broadly applicable to other borders, both *Hox* and non-*Hox*, and to other cellular fates.

References

1. Ortabozkoyun, H. *et al.* CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at Hox clusters. *Nat Genet* **54**, 202-212 (2022).
2. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).

3. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-5 (2012).

4. Phillips-Cremins, J.E. & Corces, V.G. Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**, 461-74 (2013).

5. Van Bortle, K. *et al.* Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res* **22**, 2176-87 (2012).

6. Moon, H. *et al.* CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep* **6**, 165-70 (2005).

7. Gurudatta, B.V. & Corces, V.G. Chromatin insulators: lessons from the fly. *Brief Funct Genomic Proteomic* **8**, 276-82 (2009).

8. Bell, A.C., West, A.G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396 (1999).

9. Lobanenkov, V.V. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. (1990).

10. Nora, E.P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e22 (2017).

11. Rao, S.S.P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e24 (2017).

12. Davidson, I.F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345 (2019).

13. Li, Y. *et al.* The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472-476 (2020).

14. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).

15. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**, 676-84 (2015).

16. Nora, E.P. *et al.* Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* **11**, 5612 (2020).

17. Xiao, T., Li, X. & Felsenfeld, G. The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proc Natl Acad Sci U S A* **118**(2021).

18. Wellik, D.M. Chapter 9 Hox Genes and Vertebrate Axial Pattern. in *Hox Genes* 257-278 (2009).

19. Narendra, V., Bulajic, M., Dekker, J., Mazzoni, E.O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev* **30**, 2657-2662 (2016).

20. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* **19**, 24-32 (2009).

21. Valton, A.L. *et al.* A cohesin traffic pattern genetically linked to gene regulation. *Nat Struct Mol Biol* **29**, 1239-1251 (2022).

22. Hsieh, T.S. *et al.* Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet* **54**, 1919-1932 (2022).

23. Davidson, I.F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated DNA loop extrusion. *bioRxiv*, 2022.09.08.507093 (2022).

24. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345-1349 (2019).

25. Bell, A.C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-5 (2000).

26. Chung, J.H., Whiteley, M. & Felsenfeld, G. A 5' Element of the Chicken Beta-Globin Domain Serves as an Insulator in Human Erythroid-Cells and Protects against Position Effect in Drosophila. *Cell* **74**, 505-514 (1993).

27. Recillas-Targa, F. *et al.* Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* **99**, 6883-8 (2002).

28. Dickson, J. *et al.* VEZF1 Elements Mediate Protection from DNA Methylation. *PLOS Genetics* **6**, e1000804 (2010).

29. Kobayashi, A. *et al.* A Combinatorial Code for Gene Expression Generated by Transcription Factor Bach2 and MAZR (MAZ-Related Factor) through the BTB/POZ Domain. **20**, 1733-1746 (2000).

30. Zlatanova, J. & Caiafa, P. CTCF and its protein partners: divide and rule? *Journal of Cell Science* **122**, 1275-1284 (2009).

31. Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737-743 (2020).

32. Dasen, J.S. & Jessell, T.M. Chapter Six Hox Networks and the Origins of Motor Neuron Diversity. in *Hox Genes* 169-200 (2009).

33. Mazzoni, E.O. *et al.* Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci* **16**, 1191-1198 (2013).

34. Bauer, B.W. *et al.* Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. *Cell* **184**, 5448-5464 e22 (2021).

35. Oldenkamp, R. & Rowland, B.D. A walk through the SMC cycle: From catching DNAs to shaping the genome. *Mol Cell* **82**, 1616-1630 (2022).

36. Pradhan, B. *et al.* SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Reports* **41**, 111491 (2022).

37. Narendra, V.R., P.P. ; An, D.; Raviram, R.; Skok, J.A.; Mazzoni,E.O.; Reinberg, D. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017-21 (2015).

38. Jung, H. *et al.* Global control of motor neuron topography mediated by the repressive actions of a single hox gene. *Neuron* **67**, 781-96 (2010).

39. Sweeney, L.B. *et al.* Origin and Segmental Diversity of Spinal Inhibitory Interneurons. *Neuron* **97**, 341-355 e3 (2018).

40. Xu, B. & Wellik, D.M. Axial Hox9 activity establishes the posterior field in the developing forelimb. *Proc Natl Acad Sci U S A* **108**, 4888-91 (2011).

41. Liu, N.Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat Genet* **53**, 100-109 (2021).

42. West, A.G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes Dev* **16**, 271-88 (2002).

43. Van Bortle, K. *et al.* Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* **15**, R82 (2014).

44. Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat Genet* **53**, 1064-1074 (2021).

45. Kang, J., Kim, Y.W., Park, S., Kang, Y. & Kim, A. Multiple CTCF sites cooperate with each other to maintain a TAD for enhancer-promoter interaction in the beta-globin locus. *FASEB J* **35**, e21768 (2021).

46. Kyrchanova, O. & Georgiev, P. Chromatin insulators and long-distance interactions in Drosophila. *FEBS Lett* **588**, 8-14 (2014).

47. Maksimenko, O. *et al.* Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res* **25**, 89-99 (2015).

48. Chetverina, D., Erokhin, M. & Schedl, P. GAGA factor: a multifunctional pioneering chromatin protein. *Cell Mol Life Sci* **78**, 4125-4141 (2021).

49. Melnikova, L. *et al.* Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proc Natl Acad Sci U S A* **101**, 14806-11 (2004).

50. Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O. & Corces, V.G. A Drosophila Protein That Imparts Directionality on a Chromatin Insulator Is an Enhancer of Position-Effect Variegation. *Cell* **82**, 587-597 (1995).

51. Bardwell, V.J. & Treisman, R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* **8**, 1664-77 (1994).

52. Aziati, I.D. *et al.* PATZ1 is required for efficient HIV-1 infection. *Biochem Biophys Res Commun* **514**, 538-544 (2019).

53. Saldana-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev* **28**, 723-34 (2014).

54. Kaye, E.G. *et al.* Differential Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage Compensation Complex to the Male X Chromosome. *Cell Rep* **22**, 3227-3239 (2018).

55. Bag, I., Dale, R.K., Palmer, C. & Lei, E.P. Correction: The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila (doi:10.1242/jcs.226092). *J Cell Sci* **132**(2019).

56. Jordan, W., 3rd & Larschan, E. The zinc finger protein CLAMP promotes long-range chromatin interactions that mediate dosage compensation of the Drosophila male X-chromosome. *Epigenetics Chromatin* **14**, 29 (2021).

57. Mazzoni, E.O. *et al.* Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci* **16**, 1191-8 (2013).

Acknowledgements: We thank L. Vales for reading and advice on the manuscript; Y. Grobler for providing *Drosophila* S2R$^+$ cells; D. Hernandez for technical assistance; past and present members of the Reinberg laboratory for discussions as the work was in progress. We also thank the NYU Grossman School of Medicine's Genome Technology Center, particularly A. Heguy, P. Zappile and P. Meyn, for help with sequencing; the Applied Bioinformatics Laboratories for providing bioinformatics support; and NYU Grossman School of Medicine's Cytometry and Cell Sorting Core, particularly P. Lopez and M. Gregory for help with FACS. This study utilized computing resources at the High-Performance Computing Facility of the Center for Health Informatics and Bioinformatics at the NYU Grossman School of Medicine. **Funding:** This work was supported by the National Institutes of Health (NIH) grant R01NS100897 and the Howard Hughes Medical Institute (D.R.); NIH grant R01NS100897 (E.O.M.); National Cancer Institute

(NCI)/NIH grants P01CA229086 and R01CA252239, and NCI/NIH Cancer Center Support Grant P30CA016087 (A.T.). The NYU Grossman School of Medicine's Genome Technology Center, the Applied Bioinformatics Laboratories, and the NYU Grossman School of Medicine's Cytometry and Cell Sorting Core are supported partially by the NIH/National Cancer Institute Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. **Author Contributions:** H.O., P.Y.H. and D.R. conceived the project, designed the experiments and wrote the paper. H.O. and P.Y.H performed the experiments and the bioinformatic analysis; H.C. performed the initial bioinformatic analysis of ChIP-seq data under the supervision of A.T.; and E.O.M. advised on the progression of this study. **Competing Interests:** D.R. was a cofounder of Constellation Pharmaceuticals and Fulcrum Therapeutics. Currently, D.R has no affiliation with either company. The authors declare that they have no other competing interests. **Data availability:** Sequencing data has been deposited at Gene Expression Omnibus and accession will be available immediately upon publication. The list of differentially expressed genes in *Maz* KO mESCs used in **Extended Data Fig. S6e** was obtained from the previous study[1]. Publicly available datasets used are listed in **Table S1**. All unique/stable materials generated in this study are available from the corresponding author upon reasonable request with a completed Materials Transfer Agreement. **Code Availability:** Analysis tools used in this study have been published before as described in the methods. The codes will be available upon reasonable request.

List of Supplementary Materials

Materials and Methods

Extended Data Figs. S1 to S7

Captions for Tables S1 to S7

References (Full Reference List, 58–78)

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S7

Supplementary Fig. 1

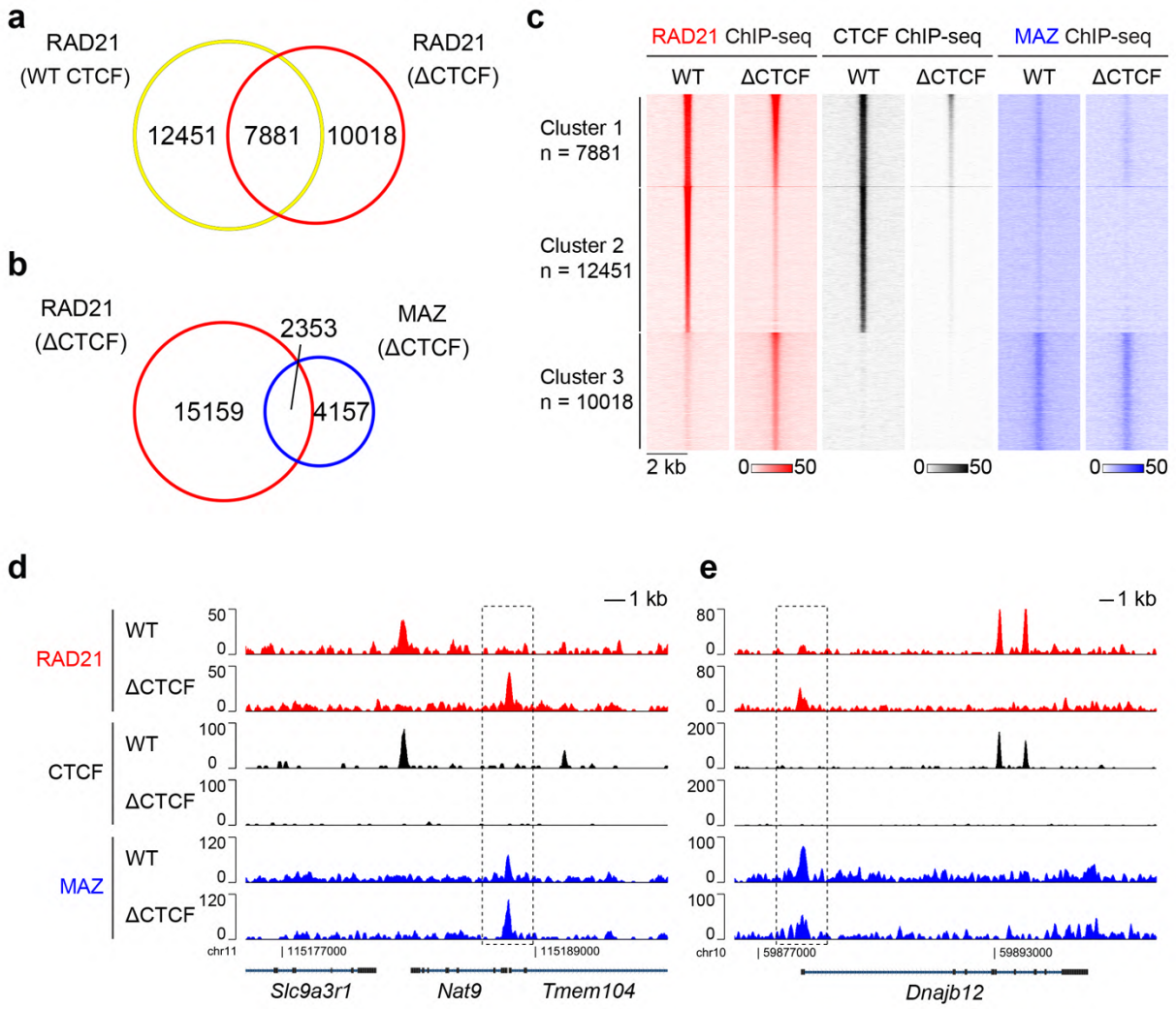


Figure 1 | RAD21 re-localizes to MAZ binding sites upon CTCF degradation in mESCs. a. Venn diagram showing RAD21 binding in mESCs with CTCF intact (WT, Untreated) versus CTCF degraded (ΔCTCF, Auxin treatment, 48 hr). **b.** Venn diagram showing MAZ and RAD21 binding in ΔCTCF mESCs. **c.** Heat maps of RAD21, CTCF, and MAZ ChIP-seq read density grouped as Cluster 1 (n=7881), Cluster 2 (n=12451), and Cluster 3 (n=10018) based on the alteration in RAD21 signal upon CTCF degradation within a 4 kb window. **d, e.** Normalized ChIP-seq densities for RAD21, CTCF, and MAZ at (**d**) *Nat9* and (**e**) *Dnajb12* loci where re-localized RAD21 overlaps with MAZ binding. ChIP-seq data is from one representative of two biological

replicates for CTCF and MAZ, and one biological replicate for RAD21 (see Nora *et. al.*[16] and Ortabozkoyun *et. al.*[1] for additional ChIP-seq datasets). RAD21-relocalized regions are indicated within dashed-lines.

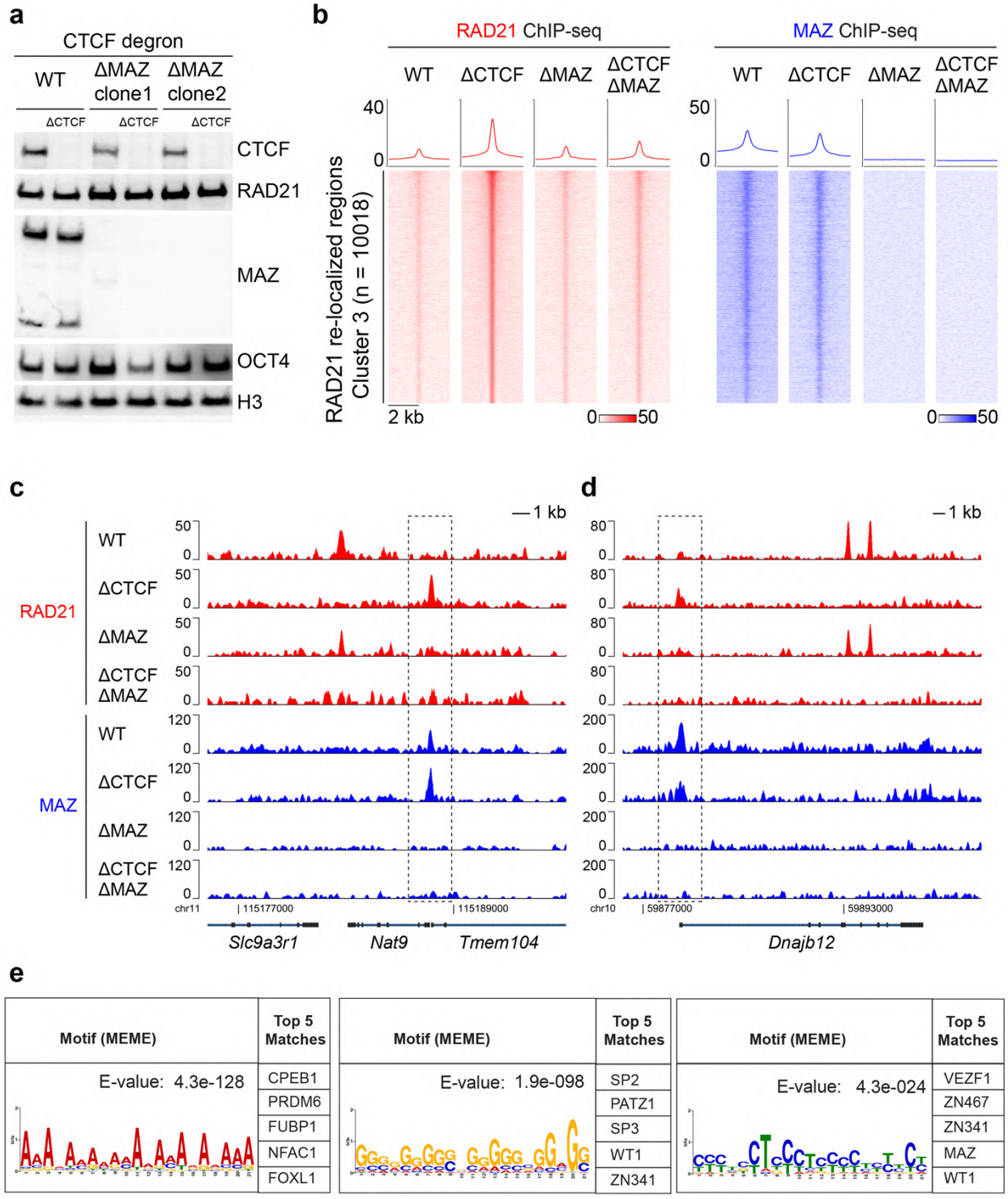


Figure 2 | Loss of MAZ in ΔCTCF mESCs reduced RAD21 re-localization, indicating a possible barrier function for MAZ. a. Western blot analysis of CTCF, MAZ, RAD21, OCT4,

and Histone H3 under WT, ΔCTCF, ΔMAZ, and ΔCTCF/ΔMAZ conditions in ESCs. ΔMAZ represents two independent *Maz* KO clones. **b.** Heat maps of RAD21 and MAZ ChIP-seq read density in RAD21 re-localized sites (Cluster 3, n=10018) within a 4 kb window under WT, ΔCTCF, ΔMAZ, and ΔCTCF/ΔMAZ conditions in mESCs. Average density profiles for ChIP-seq under each condition is indicated above the heat map. ChIP-seq data is from one representative of two biological replicates for CTCF and MAZ, and one biological replicate for RAD21. **c, d.** Normalized ChIP-seq densities for RAD21 and MAZ at (**c**) *Nat9* and (**d**) *Dnajb12* wherein re-localized RAD21 is reduced upon MAZ KO. RAD21-relocalized regions are indicated within dashed-lines. **e.** Motif analysis of RAD21 re-localized regions in the absence of CTCF and MAZ. Motifs associated with re-localized RAD21 peaks were identified by *de novo* MEME motif analysis, along with the corresponding top matches by Tomtom motif comparison tool. Motif search in MEME was performed *de novo* until 1000 sites were reached and corresponding e-values are depicted in the table (see **Extended Data Fig. S2** for the detailed list).

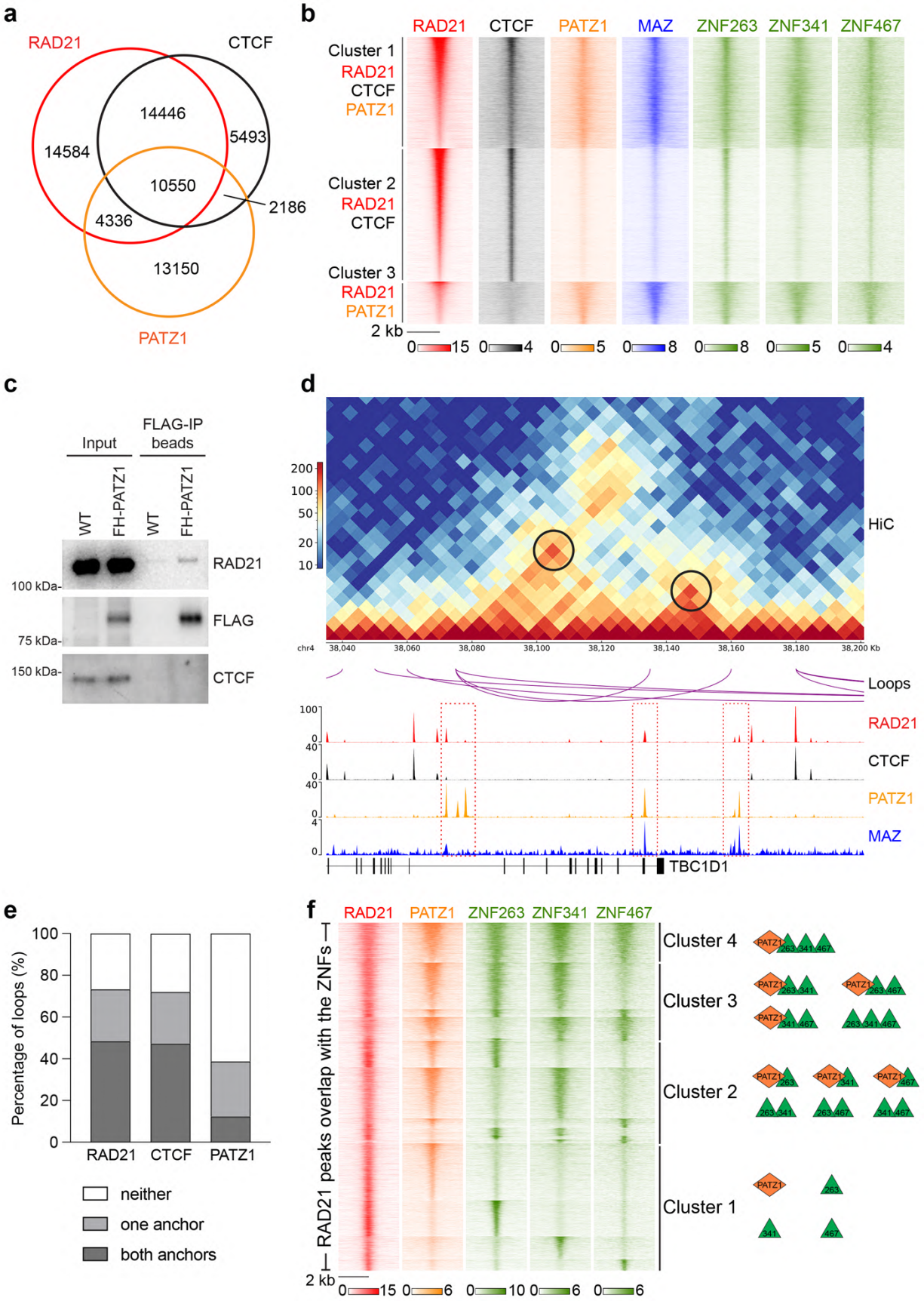

a
RAD21
CTCF
PATZ1
14446
14584
5493
10550
2186
4336
13150
b
RAD21
CTCF
PATZ1
MAZ
ZNF263
ZNF341
ZNF467
Cluster 1
RAD21
CTCF
PATZ1
Cluster 2
RAD21
CTCF
Cluster 3
RAD21
PATZ1
2 kb
0 15
0 4
0 5
0 8
0 8
0 5
0 4
c
Input
FLAG-IP beads
WT
FH-PATZ1
WT
FH-PATZ1
RAD21
100 kDa-
FLAG
75 kDa-
CTCF
150 kDa-
d
HiC
200
100
50
20
10
chr4
38,040
38,060
38,080
38,100
38,120
38,140
38,160
38,180
38,200 Kb
Loops
RAD21
CTCF
PATZ1
MAZ
TBC1D1
e
Percentage of loops (%)
100
80
60
40
20
0
RAD21
CTCF
PATZ1
neither
one anchor
both anchors
f
RAD21
PATZ1
ZNF263
ZNF341
ZNF467
RAD21 peaks overlap with the ZNFs
Cluster 4
Cluster 3
Cluster 2
Cluster 1
2 kb
0 15
0 6
0 10
0 6
0 6

Figure 3 | PATZ1 and other zinc finger proteins, ZNF263, ZNF341, and ZNF467, co-localize with RAD21 on chromatin and at loop anchors in HEK293 and HepG2 cells. a. Venn diagram showing RAD21, CTCF, and PATZ1 binding in HEK293 cells. **b.** Heat maps of RAD21, CTCF, PATZ1, MAZ, and other zinc finger proteins, ZNF263, ZNF341 and ZNF467 in HEK293 cells. ChIP-seq read density was grouped as Cluster 1, Cluster 2, and Cluster 3 based on the indicated overlaps with RAD21 signal within a 4 kb window. **c.** Western blot analysis of RAD21, FLAG, and CTCF upon FLAG-PATZ1 immunoprecipitation from mESCs (n=2). **d.** Visualization of Hi-C contact matrices for a zoomed-in region around the *TBC1D1* locus in HepG2 cells. Shown below are loops with PATZ1 at both anchors in HepG2 cells, ChIP-seq read densities for RAD21, CTCF, PATZ1, and MAZ, and gene annotations. ChIP-seq data in HepG2 cells is from two combined biological replicates. **e.** Percentage of Hi-C loops in HepG2 cells overlapping with RAD21, CTCF, and PATZ1 ChIP-seq peaks. **f.** Heat maps of RAD21, PATZ1, ZNF263, ZNF341, and ZNF467 in HEK293 cells. ChIP-seq read density was grouped as Cluster 1, Cluster 2, Cluster 3, and Cluster 4 based on the combinatorial overlaps of zinc finger proteins with RAD21 within a 4 kb window in HEK293 cells. The model on the right side indicates combinatorial binding of the indicated factors in each cluster (see **Extended Data Fig. S5a**). ChIP-seq data in HEK293 cells is from one replicate for RAD21 and one representative of two biological replicates for others. The source of genomics data used is listed in **Table S1**.

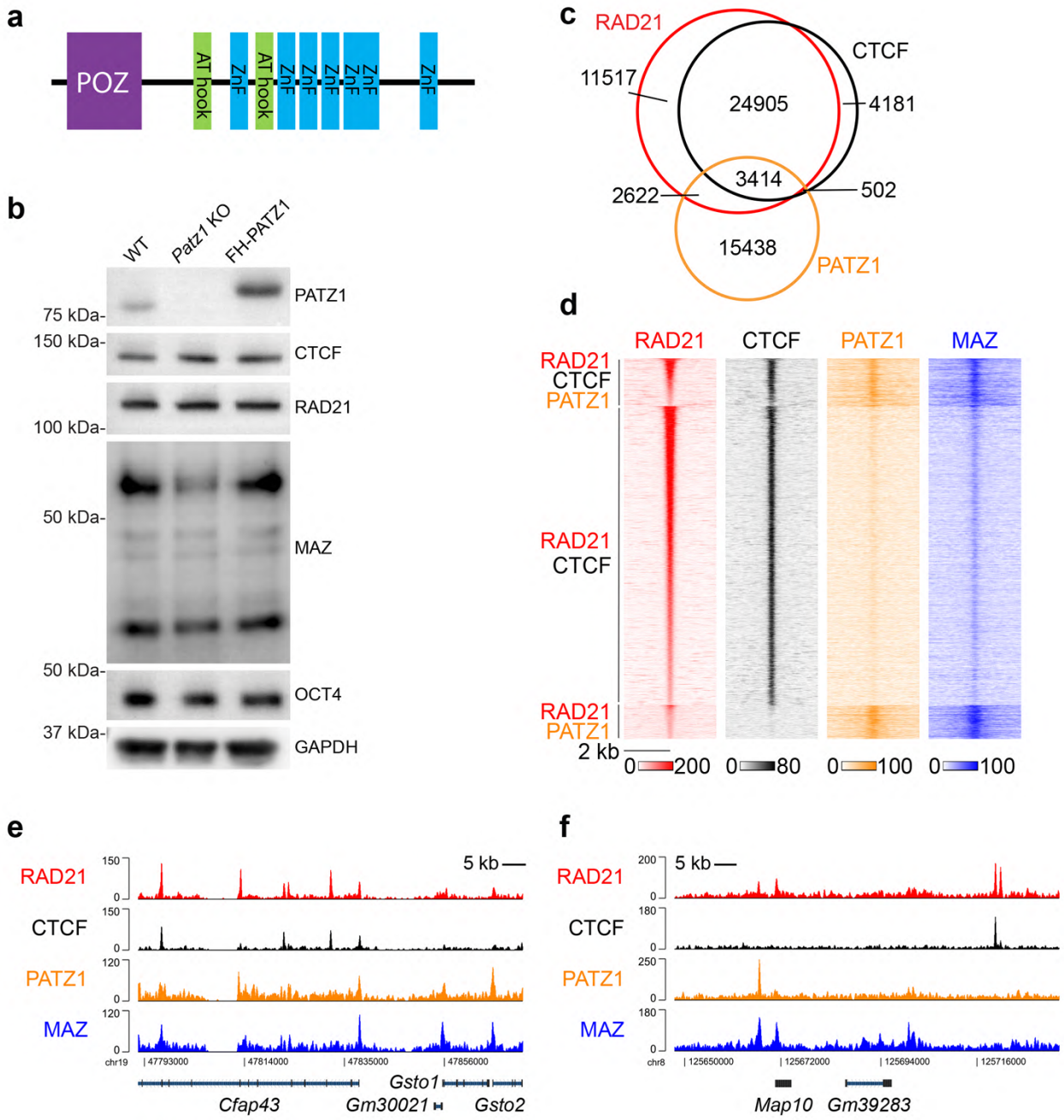


Figure 4 | PATZ1 co-localizes with RAD21 on chromatin in mESCs. a. Schematic of PATZ1 protein domains including zinc fingers (ZnFs), AT hook domains, and the BTB/POZ domain. **b.** Western blot analysis of PATZ1, CTCF, RAD21, MAZ, OCT4, and GAPDH in WT, *Patz1* KO, and FH-PATZ1 mESCs (see **Extended Data Fig. S6a** for four independent *Patz1* KO clones). **c.**

Venn diagram showing RAD21, CTCF, and PATZ1 binding in mESCs. **d.** Heat maps of RAD21, CTCF, PATZ1, and MAZ ChIP-seq read densities grouped as Cluster 1, Cluster 2, and Cluster 3 based on the indicated overlaps with RAD21 signal within a 4 kb window in mESCs. **e, f.** Normalized ChIP-seq densities for RAD21, CTCF, PATZ1, and MAZ, wherein co-localizing peaks were visualized. ChIP-seq data is from one representative of two biological replicates.

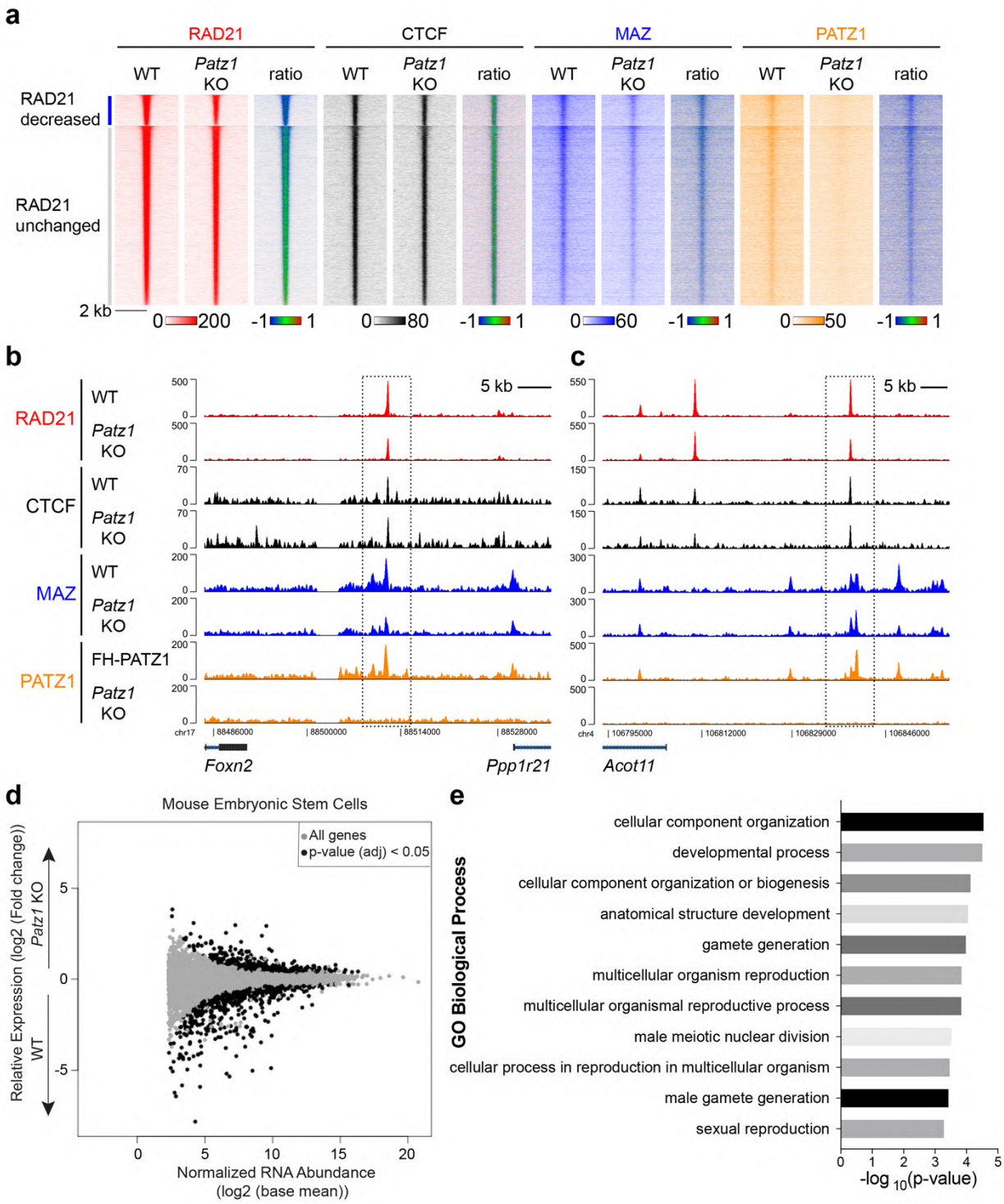


Figure 5 | Loss of PATZ1 impacts RAD21 and MAZ chromatin binding in mESCs and gene expression of developmental processes. a. Heat maps of RAD21, CTCF, MAZ, and PATZ1

ChIP-seq read densities in WT and *Patz1* KO ESCs at RAD21 peaks, clustered into RAD21-decreased and -unchanged sites. Each ratiometric heat map plotted to the right shows the $\log_2$ fold change (*Patz1* KO/WT) of the signals. **b, c.** Normalized ChIP-seq densities for RAD21, CTCF, MAZ, and PATZ1 at the indicated loci wherein MAZ and RAD21 signal was altered. ChIP-seq data is from one representative of two biological replicates. **d.** Differentially expressed genes by RNA-seq upon *Patz1* KO in ESCs from three biological replicates (see all in **Table S2**). **e.** Gene Ontology (GO) analysis shows the categories of biological processes that were enriched in the differentially expressed genes in *Patz1* KO versus WT mESCs. PANTHER overrepresentation test tools were used for GO analysis (see all in **Table S3**).

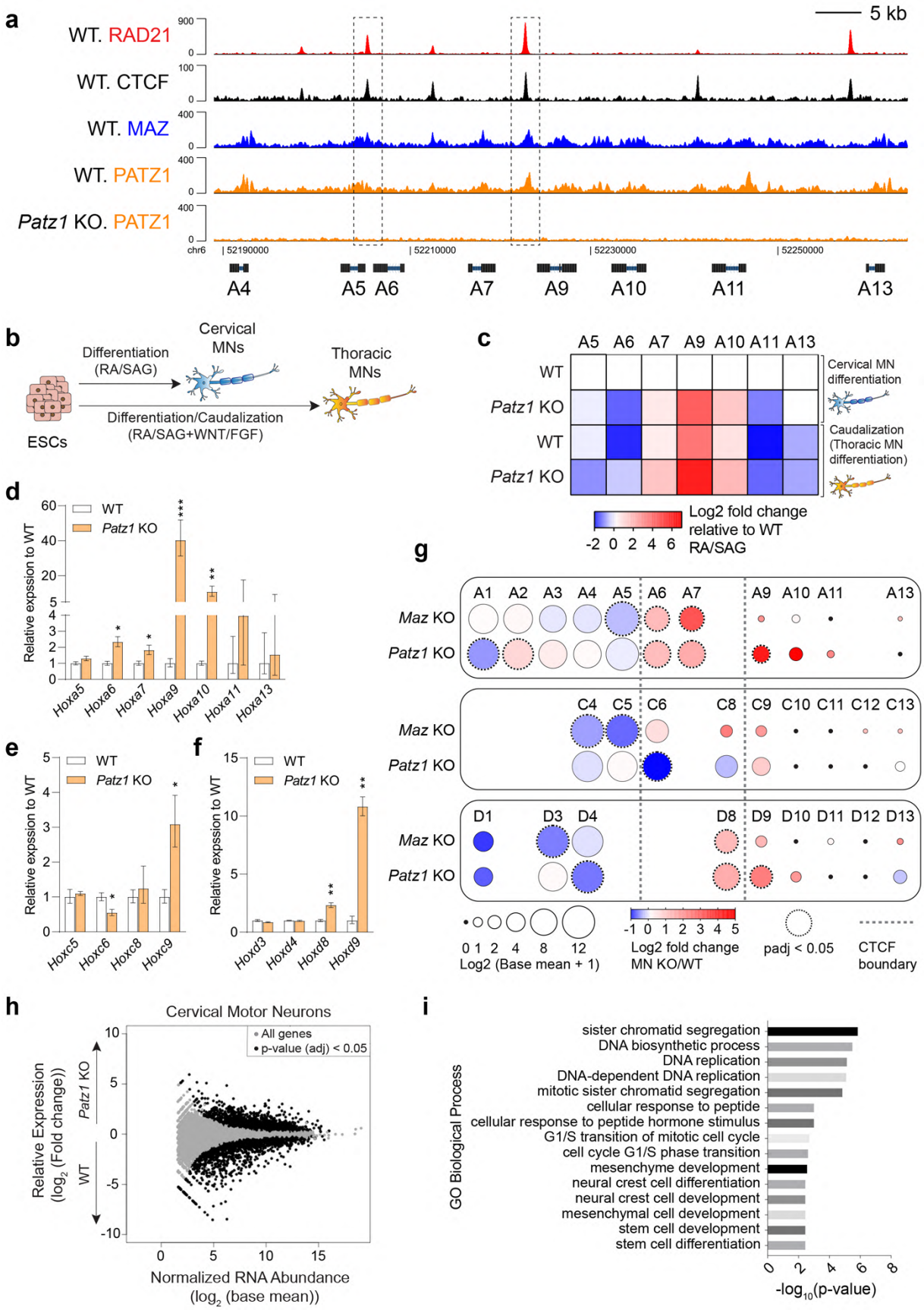

a
5 kb
WT. RAD21
WT. CTCF
WT. MAZ
WT. PATZ1
Patz1 KO. PATZ1
chr6
52190000
52210000
52230000
52250000
A4
A5 A6
A7
A9
A10
A11
A13
b
Differentiation (RA/SAG)
Cervical MNs
Thoracic MNs
ESCs
Differentiation/Caudalization (RA/SAG+WNT/FGF)
c
A5 A6 A7 A9 A10 A11 A13
WT
Patz1 KO
WT
Patz1 KO
Cervical MN differentiation
Caudalization (Thoracic MN differentiation)
Log2 fold change relative to WT RA/SAG
d
WT
Patz1 KO
Relative expssion to WT
Hoxa5
Hoxa6
Hoxa7
Hoxa9
Hoxa10
Hoxa11
Hoxa13
e
Hoxc5
Hoxc6
Hoxc8
Hoxc9
f
Hoxd3
Hoxd4
Hoxd8
Hoxd9
g
Maz KO
Patz1 KO
A1 A2 A3 A4 A5 A6 A7 A9 A10 A11 A13
C4 C5 C6 C8 C9 C10 C11 C12 C13
D1 D3 D4 D8 D9 D10 D11 D12 D13
Log2 (Base mean + 1)
Log2 fold change MN KO/WT
padj < 0.05
CTCF boundary
h
Cervical Motor Neurons
All genes
p-value (adj) < 0.05
Relative Expression (log2 (Fold change))
Patz1 KO
WT
Normalized RNA Abundance (log2 (base mean))
i
GO Biological Process
sister chromatid segregation
DNA biosynthetic process
DNA replication
DNA-dependent DNA replication
mitotic sister chromatid segregation
cellular response to peptide
cellular response to peptide hormone stimulus
G1/S transition of mitotic cell cycle
cell cycle G1/S phase transition
mesenchyme development
neural crest cell differentiation
neural crest cell development
mesenchymal cell development
stem cell development
stem cell differentiation
-log10(p-value)

Figure 6 | Combinatorial binding of insulation factors at distinct *Hox* gene borders determines rostral-caudal patterning in MNs. a. Normalized ChIP-seq densities for RAD21, CTCF, MAZ, and PATZ1 in WT and *Patz1* KO mESCs at the indicated regions in the *HoxA* cluster. **b.** Differentiation of mESCs into cervical and thoracic MNs *in vitro* (see Methods for details). **c.** Heat map of relative gene expression in WT versus *Patz1* KO MNs at the *HoxA* cluster in cervical MNs and thoracic MNs from one biological replicate (see **Extended Data Fig. S7c** for three biological conditions). WNT/FGF signaling was utilized for MN caudalization, as described previously[32,57] (see Methods for details). **d-f.** RT-qPCR analysis for the indicated *Hox* genes in (**d**) the *HoxA* cluster, (**e**) the *HoxC* cluster, and (**f**) the *HoxD* cluster in WT, MAZ KO, and *Patz1* KO cervical MNs. RT-qPCR signal is normalized to *Atp5f1* and *ActB* levels. The fold-change in expression was calculated relative to WT MNs. All RT-qPCR results are represented as mean values and error bars indicating $\log_2$(SE) across three biological replicates (two-sided Student's *t*-test without multiple testing correction; ***$P \leq 0.001$, **$P \leq 0.01$, *$P < 0.05$). **g.** Heat maps of relative gene expression in WT versus *Patz1* KO MNs at the *HoxA* cluster from three biological replicates. WT and *Maz* KO mESCs in E14 *Hoxa5:a7* dual-reporter background were used[1]. **h.** Differentially expressed genes by RNA-seq upon *Patz1* KO in MNs from two biological replicates (see all in **Table S4**). **i.** GO analysis showing the top biological processes enriched in the differentially expressed genes in *Patz1* KO versus WT MNs. PANTHER overrepresentation test tools were used for GO analysis and top 15 categories having a fold enrichment > 2.5 were plotted (see all in **Table S5**).

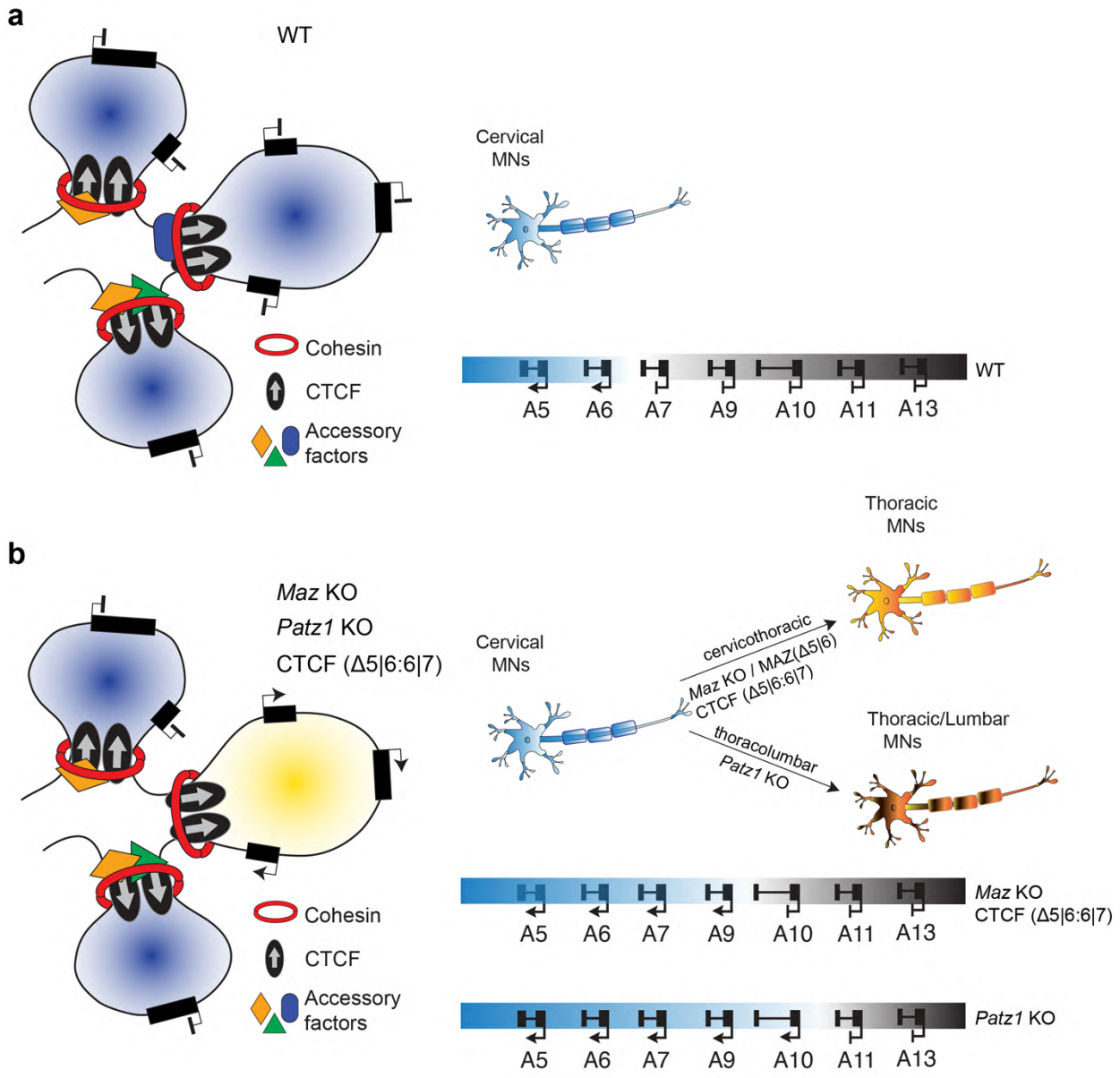


Figure 7 | Combinatorial binding of insulation factors at boundaries are critical to the regulation of cellular identity. a. Model indicating the anterior-posterior MN identity in the WT setting in the presence of different accessory/insulation factors. **b.** Model indicating the transition of anterior-posterior MN identities impacted upon the loss of different accessory/insulation factors. The illustrations on the left sides demonstrate the regulation of gene expression through the combinatorial binding of accessory factors in addition to cohesin and CTCF at loop anchors.

Supplementary Materials for

Novel Chromatin Insulating Activities Uncovered upon Eliminating Known Insulators *in vivo*

Havva Ortabozkoyun[1,2,*†], Pin-Yao Huang[1,2,*], Hyunwoo Cho[3,4,5], Aristotelis Tsirigos[3,4,6], Esteban Mazzoni[7], Danny Reinberg[1,2†]

[†] Correspondence to: Danny.Reinberg@nyulangone.org, Havva.Ortabozkoyun@nyulangone.org

This PDF file includes:

Materials and Methods
Extended Data Figs. S1 to S7
Captions for Tables S1 to S7
References (Full Reference List, 58–78)

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S7
Supplementary Fig. 1

Methods

This study has been performed under compliance with ethical regulations and approved by NYU/NYULMC Institutional Biosafety Committee.

Cell culture and motor neuron differentiation

E14 mESCs (ES-E14TG2a, ATCC, CRL-1821) were cultured in standard medium supplemented with LIF, and 2i conditions (1 mM MEK1/2 inhibitor (PD0325901, Stemgent) and 3 mM GSK3 inhibitor (CHIR99021, Stemgent)). Cervical MN differentiation was performed as described previously[1,37]. Briefly, mESCs were differentiated into embryoid bodies (EBs) in 2 days, and further patterning was induced with addition of 1 μM all-trans-retinoic acid (RA, MilliporeSigma) and 0.5 μM smoothened agonist (SAG, Calbiochem). For caudalization of MNs, previously described methods were applied[32,57]. mESCs were differentiated into EBs in 2 days, and patterning was induced with 1 μM RA, 0.5 μM SAG, 75 ng/ml mouse WNT3A (R&D systems, Cat. 1324-WN-010/CF) and murine FGF-basic (Peprotech, Cat. 450-33) in a gradient of the described concentrations (50 ng/ml, 100 ng/ml, and 150 ng/ml) for 4 days. Biological replicates stand for independent differentiation experiments performed. 293FT cells (Thermo Fisher Scientific, R70007) were cultured in standard medium as described in the manufacturer's protocol.

CRISPR genome editing

Single-guide RNAs (sgRNAs) were designed using CRISPR design tools in Benchling (https://benchling.com). All sgRNAs were cloned into pSpCas9(BB)-2A-GFP (PX458) (Addgene, #48138) or pSpCas9(BB)-2A-Puro (PX459) V2.0 vector (Addgene, #62988). The sgRNAs were transfected into mESCs using Lipofectamine 2000 (Invitrogen), as described previously[1,37]. Single clones from GFP-positive FACS sorted cells or puromycin (InvivoGen)-resistant cells were genotyped and confirmed by sequencing. When necessary, PCR products were further assessed by TOPO cloning and sequencing to distinguish the amplified products of different alleles. The

sequencing chromatograms were aligned in Benchling. All sgRNAs and genotyping primers are shown in **Table S6**.

Cell line generation

***Maz* KO cells in CTCF-AID background**

To generate *Maz* KO cell lines in CTCF-AID background, CTCF-AID mESCs[10] were transfected with sgRNA in PX458 vector targeting the *Maz* locus. Knock-out of *Maz* was confirmed by genotyping, sequencing, and western blot. sgRNA and genotyping primers were described previously[1].

***Patz1* KO cells**

To generate *Patz1* KO cell lines, WT E14TG2a mESCs were transfected with sgRNA in PX459 targeting the *Patz1* locus. Knock-out of *Patz1* was confirmed by genotyping, sequencing, and western blot.

FH-PATZ1 cells

The mouse *Patz1* coding sequence (NM_019574) with an HA-tag sequence fused to its 5' end was cloned into a PiggyBac vector (pPB-CAG-3xFLAG-empty-pgk-hph, Addgene, #48754) to create pPB-CAG-3xFLAG-HA-PATZ1-pgk-hph plasmid via Gibson assembly (NEB, #E2611). *Patz1* KO cells were transfected with 50 ng of pPB-CAG-3xFLAG-HA-PATZ1-pgk-hph plasmid and 100 ng of Super PiggyBac Transposase plasmid (System Biosciences, #PB210PA-1) using Lipofectamine 2000 (Thermo Fisher Scientific), and selected with 800 ug/mL of Hygromycin B for stable cells. Expression of 3xFLAG-HA-PATZ1 was evaluated by western blot.

Cloning of CβF-PATZ1

To generate CβF-PATZ1, mouse *Patz1* coding sequence was inserted into PCR-linearized CβF vector[1] via Gibson assembly.

Immunoprecipitation

For immunoprecipitation from mESCs, nuclear extracts of E14 WT and FH-PATZ1 cells were prepared as described previously[58]. Briefly, buffer A (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KC1 and 0.5 mM DTT) was used to remove the cytoplasmic fraction, and buffer C (20 mM HEPES, pH 7.9, 25% (v/v) glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, and 0.2 mM EDTA) was used to extract nuclear proteins. The NaCl concentration in the nuclear extracts was then diluted to 250 mM using BC50 buffer (20 mM HEPES, pH 7.9 and 50 mM NaCl), and a 10% NP40 solution was added to the diluted nuclear extracts yielding a final 0.1% concentration. Anti-FLAG® M2 Magnetic Beads (MilliporeSigma, #M8823) were used for immunoprecipitation, and proteins bound to the beads were released by the addition of SDS loading buffer, resolved on a 10% Bis-Tris gel, and analyzed using the specific antibodies listed in **Table S7**.

For immunoprecipitation from 293FT cells, CβF-PATZ1 expression plasmids were transfected into 293FT cells using PEI, and nuclei was prepared using TMSD and BA450 buffers, as described previously[59,60]. Briefly, TMSD buffer (20 mM HEPES, 5 mM $MgCl_2$, 85.5 g/L sucrose, 25 mM NaCl, and 1 mM DTT) was used for cytosol removal, and nuclear extraction was done in BA450 buffer (20 mM HEPES, 450 mM NaCl, 5% glycerol, and 0.2 mM EDTA). FLAG affinity immunoprecipitation and FLAG peptide elution were performed as previously described[1].

Expression analysis

Total RNA was purified from cells with RNAeasy Plus Mini kit (Qiagen) and reverse transcription was performed on 1 μg RNA by using Superscript III (Life Technologies) and random hexamers (Thermo Fisher Scientific). RT-qPCRs were performed in replicates on 100 ng cDNA using PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) on a CFX384 Touch Real-Time PCR detection system (Bio-Rad). The primers are listed in **Table S6**. For RNA-seq analysis, 1 μg

RNA was used to prepare ribominus RNA-seq libraries according to the manufacturer's protocols by the NYU Genome Technology Center.

Chromatin immunoprecipitation sequencing

ChIP-seq experiments were performed as described[1,61]. Briefly, cells were fixed with 1% formaldehyde, nuclei were isolated, and chromatin was fragmented to ~250 bp in size using a Diagenode Bioruptor. ChIP was performed using antibodies listed in **Table S7**. Chromatin from *Drosophila* (1:100 ratio to mESC-derived chromatin), and *Drosophila*-specific H2Av antibody were used as a spike-in control in each sample. For ChIP-seq, libraries were prepared as described[1,37] using 1-30 ng of immunoprecipitated DNA.

Library construction

ChIP-seq libraries were prepared as described[37]. RNA-seq libraries were prepared using KAPA library preparation kits by the NYU Genome Technology Center.

Data analysis

RNA-seq analysis

RNA-seq data was analyzed as described[37]. Briefly, sequence reads were mapped to mm10 reference genome with Bowtie 2 (version 2.3.4.1)[62] and normalized differential gene expression was obtained with DEseq2 (version 1.26.0)[63,64]. Differential gene expression analysis was performed using the Wald test built into DEseq2 with an FDR cutoff of 0.05. Relevant expression and p-values are listed for differentially expressed genes in **Table S2** and **Table S4**. PANTHER database was used for Gene Ontology (GO) analysis[65]. Relevant GO terms are listed in **Table S3** and **Table S5**. Venn diagrams were drawn by using online tools: GeneVenn (https://www.bioinformatics.org/gvenn/) and Venn Diagram Plotter (https://pnnl-comp-mass-spec.github.io/Venn-Diagram-Plotter/).

ChIP-seq analysis

ChIP-seq experiments were analyzed as described previously[1,61]. In brief, sequence reads were mapped to mm10 or dm6 reference genome with Bowtie 2 (version 2.3.4.1) using default parameters[62]. Quality filtering and removal of PCR duplicates were performed by using SAMtools (version 1.9)[66]. MACS (version 1.4.2) was used for narrow peak calling using default parameters of 'macs2'[67]. After normalization with the spike-in *Drosophila* read counts or total read counts (for RAD21, to avoid normalization artifacts[16]); ChIP-seq heat maps, density plots, and tracks were generated using Easeq[68] (http://easeq.net, version v 1.111). ChIP-seq reads were extended to 150 bp using default parameters in Easeq. 'ChIPpeakAnno' package (version 3.20.1) from Bioconductor[69] was used to draw Venn diagrams to visualize the overlap among ChIP-seq samples. In addition, BEDTools (version 2.27.1) were also used for the assessment of overlaps[70]. ChIP-seq "bed" file coordinates were converted into "fasta" by using fetch sequences tool within Regulatory Sequence Analysis Tools (RSAT)[71]; MEME (version 5.5.1) was used for motif analysis of PATZ1 in mESCs[72], and Tomtom (version 5.5.1) was used as a motif comparison tool[73].

For public data, filtered alignment files from ENCODE (https://www.encodeproject.org/) were directly used for the analysis. When filtered alignment files were not available, raw sequence reads from NCBI were analyzed via the Seq-N-Slide pipeline (https://doi.org/10.5281/zenodo.5550459). The data used are listed in **Table S1**. Human reference genome hg38 were used for comparative analysis. ChIP-seq read densities were normalized to fragments per million reads per kbp. Easeq[68] or pyGenomeTracks[74,75] was used to generate the heat maps and representative tracks. The UpSet plot was generated using Intervene[76] (https://asntech.shinyapps.io/intervene/).

Analysis of RAD21, CTCF, and PATZ1 occupancy in loop anchors of HepG2 cells

Significantly enriched chromatin loops were called using HiCCUPS from the Juicer package[77] with default parameters. Overlaps between RAD21, CTCF, and PATZ1 ChIP-seq peaks and chromatin loops were analyzed with pairToBed from the BEDTools suite (version 2.27.1)[70] and pgltools (version 1.2.1)[78]. Visualization of Hi-C and associated ChIP-seq data was made with pyGenomeTracks[74,75].

Statistical Analysis

Statistical analysis related to experiments have been described above in each section. Statistical analyses of RT-qPCR were analyzed with CFX Maestro Software 2.3 (Bio-Rad).

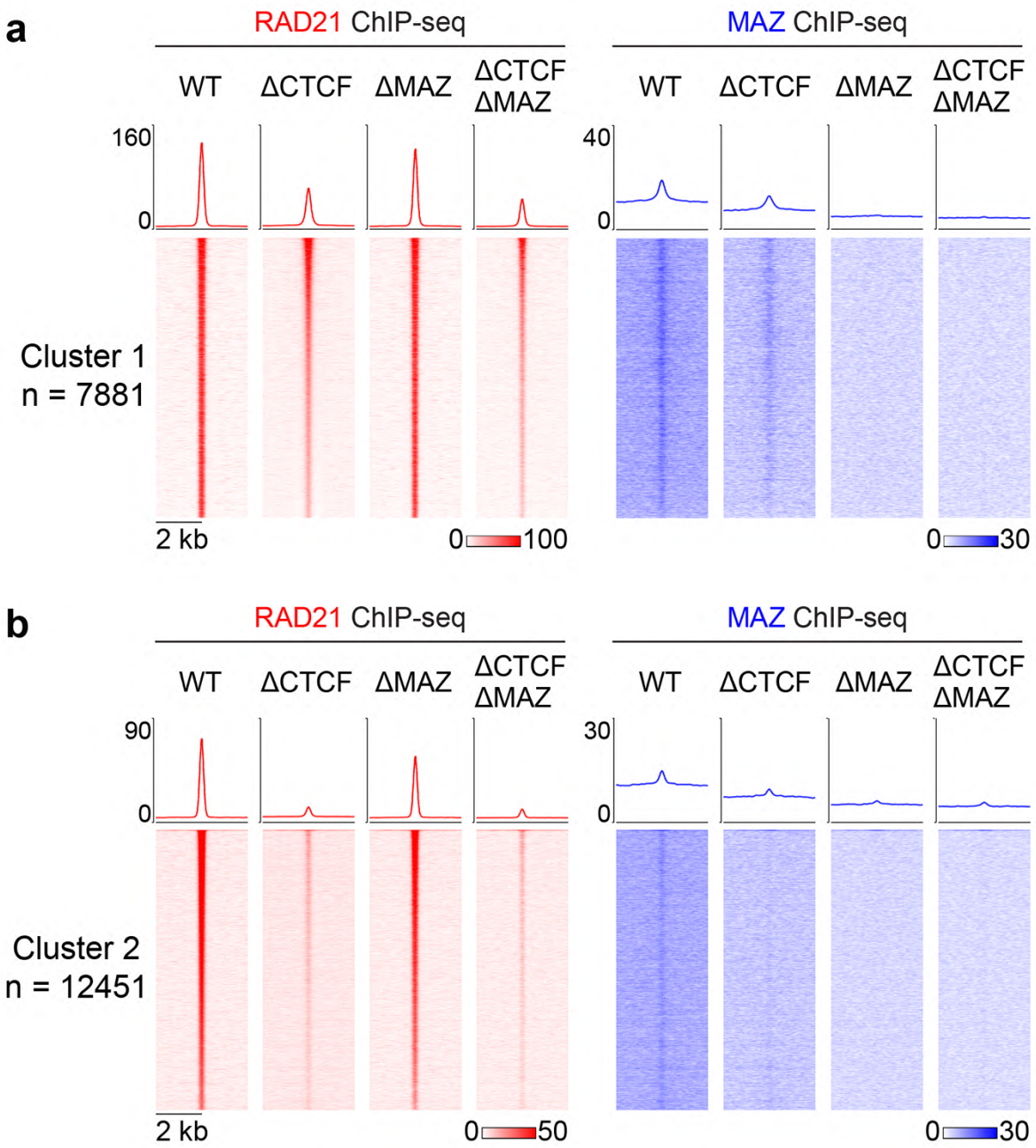

a
RAD21 ChIP-seq
MAZ ChIP-seq
WT
ΔCTCF
ΔMAZ
ΔCTCF ΔMAZ
160
0
40
0
Cluster 1
n = 7881
2 kb
0
100
0
30
b
RAD21 ChIP-seq
MAZ ChIP-seq
WT
ΔCTCF
ΔMAZ
ΔCTCF ΔMAZ
90
0
30
0
Cluster 2
n = 12451
2 kb
0
50
0
30

Figure S1 | Loss of MAZ in CTCF-degraded mESCs results in the reduction of re-localized RAD21 signal at regions co-occupied by MAZ and RAD21. a, b. Heat maps of RAD21 and MAZ ChIP-seq read density in (**a**) Cluster 1 (n=7,881) and (**b**) Cluster 2 (n=12,451) within a 4 kb window in WT, ΔCTCF, ΔMAZ, and ΔCTCF/ΔMAZ conditions in ESCs. Average density profiles for ChIP-seq under each condition is indicated above the heat map. ChIP-seq data is from one representative biological replicate for CTCF degron ESCs and one representative of two biological replicates for MAZ KO clones in CTCF degron background.

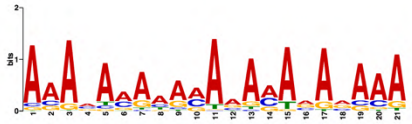
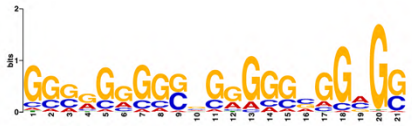
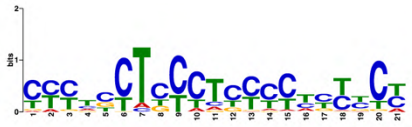
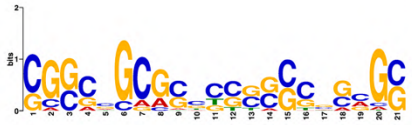
Motif (MEME)	E-value	Sites	Top 10 Matches (Tomtom)	
			Mouse	Human
	4.3e-128	1000	STAT1	CPEB1
			FUBP1	PRDM6
			FOXJ3	FUBP1
			IRF3	NFAC1
			FOXJ3	FOXL1
			ONEC2	FOXG1
			IRF5	STAT1
			STAT2	FOXJ3
			TCF7	IRF1
			IRF1	ANDR
	1.9e-098	1000	SP2	SP2
			SP5	PATZ1
			SP3	SP3
			WT1	WT1
			SP1	ZN341
			MAZ	ZN467
			ZBT17	MAZ
			KLF15	ZN263
			KLF3	VEZF1
			ZN281	TBX15
	4.3e-024	879	SP5	VEZF1
			MAZ	ZN467
			WT1	ZN341
			FLI1	MAZ
			SP5	WT1
			KLF15	PATZ1
			SP4	FLI1
			ELF5	ZN263
			SP3	KLF15
			ETS2	RXRA
	5.8e-004	440	SP2	SP2
			ZFX	SP1
			SP3	SP3
			SP1	AP2D
			AP2D	KLF3
			KLF3	SP1
			SP4	KLF9
			SP1	THAP1
			ZBT17	SP4
			MXI1	KLF12

Figure S2 | Motif analysis of RAD21 re-localized regions in the absence of CTCF and MAZ. Motif identification for re-localized RAD21 peaks in the absence of CTCF and MAZ by *de novo* MEME motif analysis, along with the corresponding top matches by Tomtom motif comparison tool. Motif search in MEME has been performed *de novo* until 1000 sites were reached and corresponding e-values are depicted in the table.

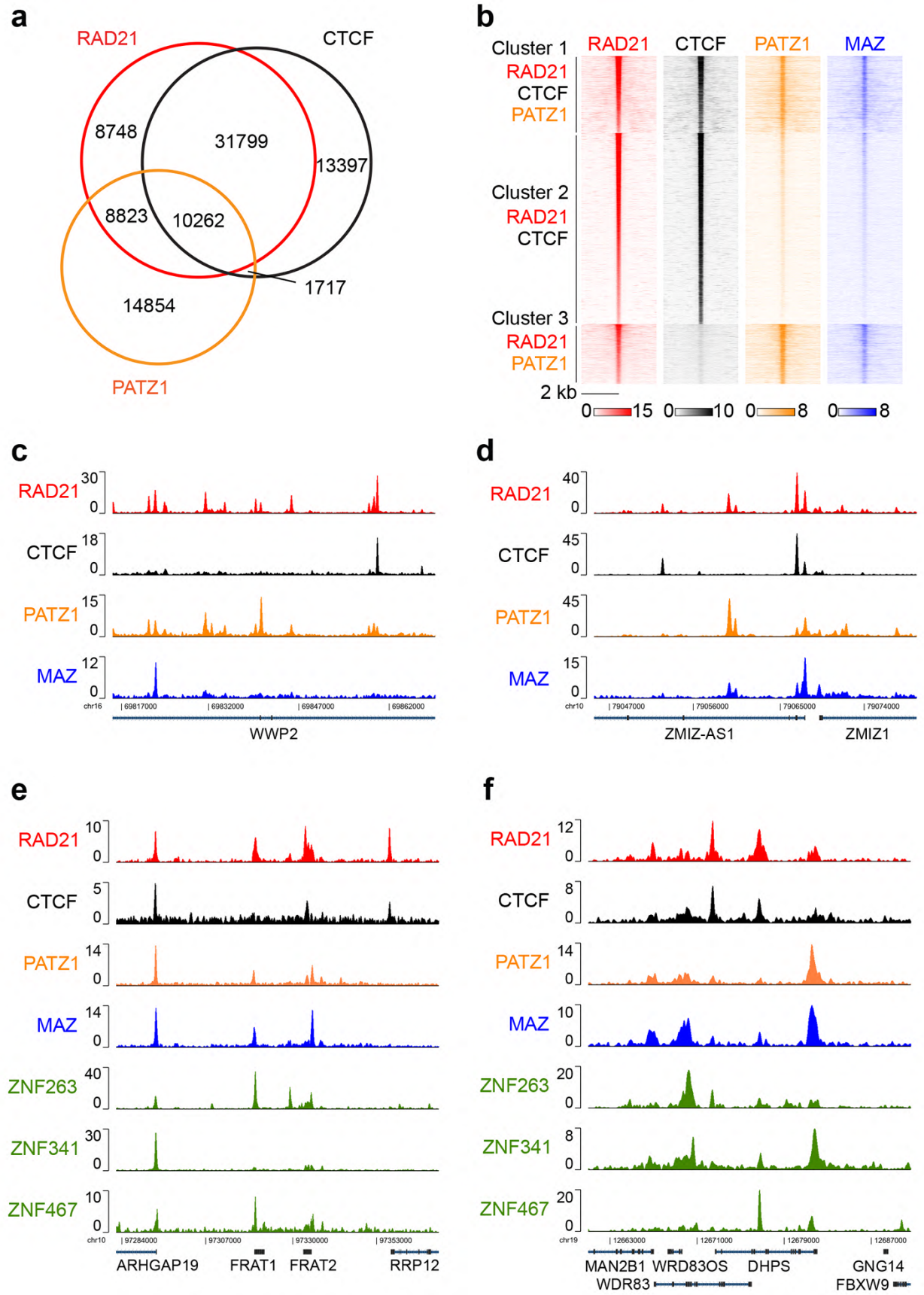

a
RAD21
CTCF
PATZ1
8748
31799
13397
8823
10262
1717
14854
b
Cluster 1
RAD21
CTCF
PATZ1
MAZ
RAD21
CTCF
PATZ1
Cluster 2
RAD21
CTCF
Cluster 3
RAD21
PATZ1
2 kb
0 15
0 10
0 8
0 8
c
RAD21
CTCF
PATZ1
MAZ
WWP2
d
RAD21
CTCF
PATZ1
MAZ
ZMIZ-AS1
ZMIZ1
e
RAD21
CTCF
PATZ1
MAZ
ZNF263
ZNF341
ZNF467
ARHGAP19
FRAT1
FRAT2
RRP12
f
RAD21
CTCF
PATZ1
MAZ
ZNF263
ZNF341
ZNF467
MAN2B1
WRD83OS
DHPS
GNG14
WDR83
FBXW9

Figure S3 | PATZ1 co-localizes with RAD21 on chromatin in HepG2 and HEK293 cells. a. Venn diagram showing RAD21, CTCF, and PATZ1 binding in HepG2 cells. **b.** Heat maps of RAD21, CTCF, PATZ1, and MAZ ChIP-seq read density in HepG2 cells clustered as Cluster 1, Cluster 2, and Cluster 3 based on the indicated overlaps with RAD21 signal within a 4 kb window. **c, d.** Normalized ChIP-seq densities for RAD21, CTCF, PATZ1, and MAZ, wherein peaks co-localizing with RAD21 and/or CTCF were visualized in HepG2 cells. ChIP-seq data in HepG2 cells is from two combined biological replicates. **e, f.** Normalized ChIP-seq densities for RAD21, CTCF, PATZ1, MAZ, and other zinc finger proteins, ZNF263, ZNF341, and ZNF467, wherein RAD21 and/or CTCF co-localizing peaks were observed in HEK293 cells. ChIP-seq data in HEK293 cells is from one replicate for RAD21 and one representative of two biological replicates for others. The source of genomics data used is listed in **Table S1**.

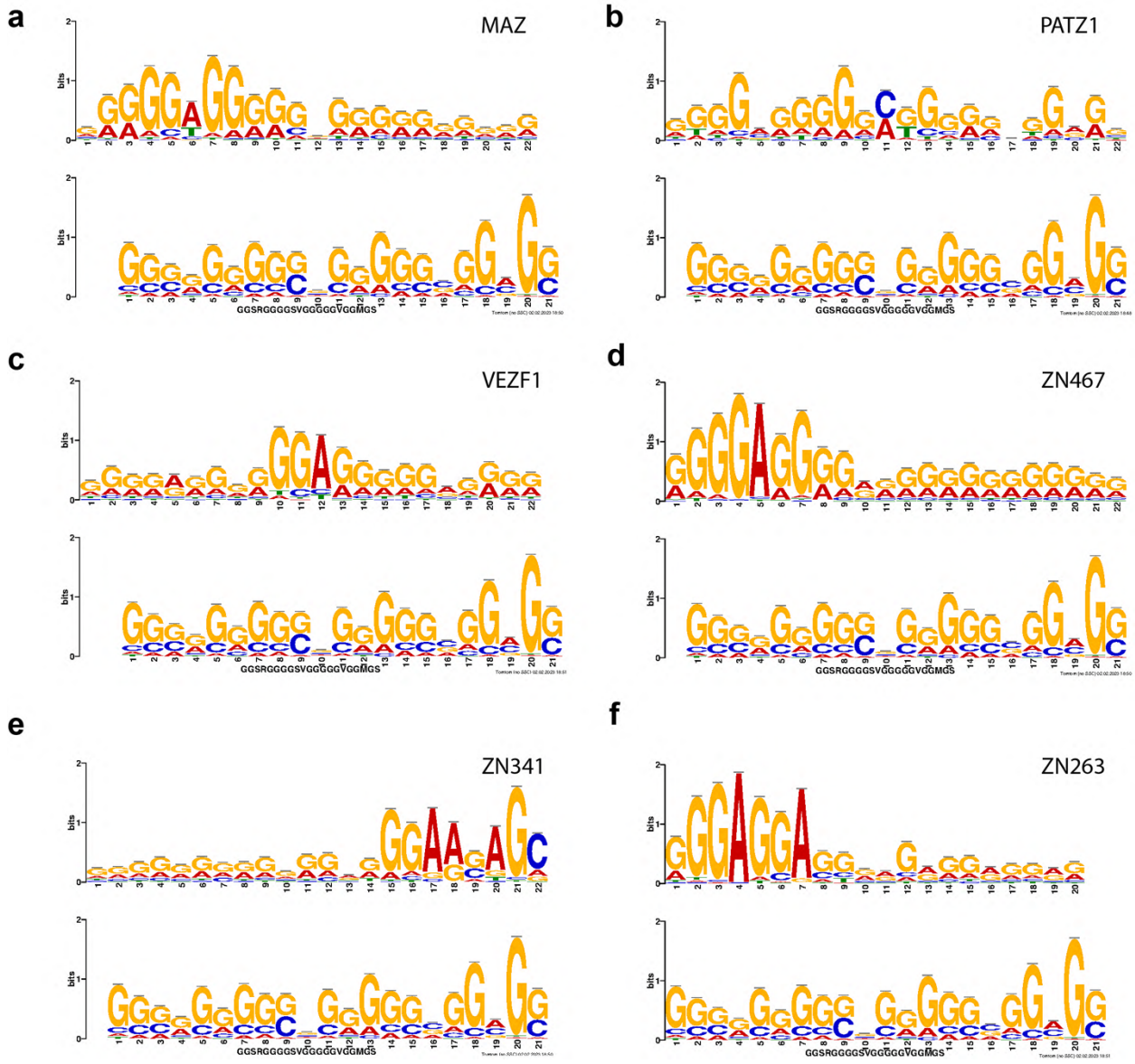

a
MAZ
b
PATZ1
c
VEZF1
d
ZN467
e
ZN341
f
ZN263
bits
GGSRGGGGSVGGGGVGGMGS

Figure S4 | Motif alignment and comparison for the top candidates at RAD21 re-localized regions in the absence of CTCF and MAZ. Representative alignment of top motif matches to the candidates: MAZ (**a**), PATZ1 (**b**), VEZF1 (**c**), ZN467 (**d**), ZN341 (**e**), and ZN263 (**f**). *De-novo* motif analysis was performed as depicted in **Extended Data Fig. S2** for re-localized RAD21 peaks in the absence of CTCF and MAZ. Motif alignments were generated through Tomtom motif comparison tool.

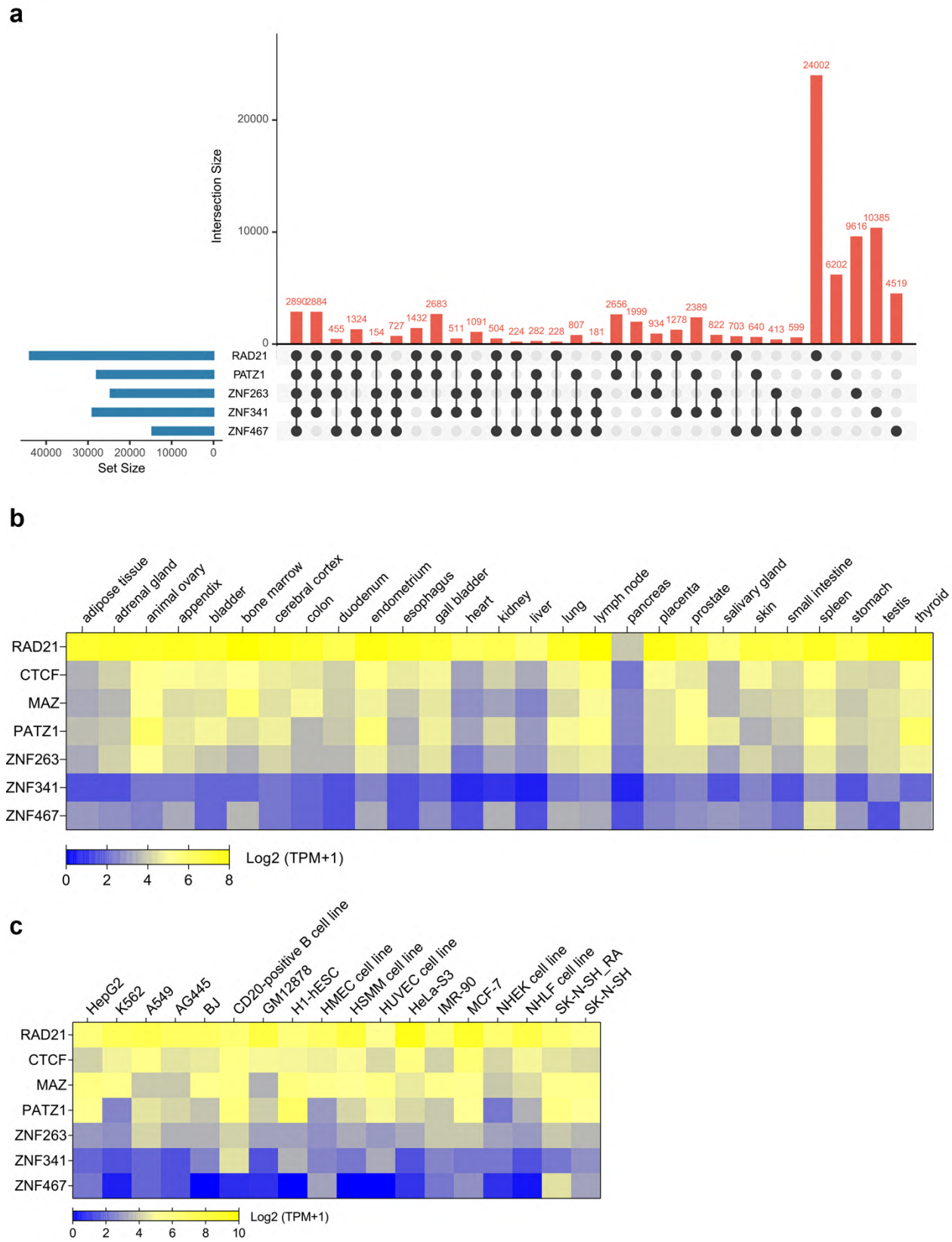

a
Intersection Size
20000
10000
0
2890
2884
455
1324
154
727
1432
2683
511
1091
504
224
282
228
807
181
2656
1999
934
1278
2389
822
703
640
413
599
24002
6202
9616
10385
4519
RAD21
PATZ1
ZNF263
ZNF341
ZNF467
40000
30000
20000
10000
0
Set Size
b
adipose tissue
adrenal gland
animal ovary
appendix
bladder
bone marrow
cerebral cortex
colon
duodenum
endometrium
esophagus
gall bladder
heart
kidney
liver
lung
lymph node
pancreas
placenta
prostate
salivary gland
skin
small intestine
spleen
stomach
testis
thyroid
RAD21
CTCF
MAZ
PATZ1
ZNF263
ZNF341
ZNF467
Log2 (TPM+1)
0
2
4
6
8
c
HepG2
K562
A549
AG445
BJ
CD20-positive B cell line
GM12878
H1-hESC
HMEC cell line
HSMM cell line
HUVEC cell line
HeLa-S3
IMR-90
MCF-7
NHEK cell line
NHLF cell line
SK-N-SH_RA
SK-N-SH
RAD21
CTCF
MAZ
PATZ1
ZNF263
ZNF341
ZNF467
Log2 (TPM+1)
0
2
4
6
8
10

Figure S5 | PATZ1 and other ZNFs are expressed at varying levels across different tissues and cell types and co-localize with RAD21 on chromatin in HEK293 cells. a. Overlap of RAD21, PATZ1, ZNF263, ZNF341, and ZNF467 binding in HEK293 cells. **b, c.** Heat map of RNA-seq expression [$\log_2$ (TPM+1)] of RAD21 and ZNFs across different human tissues (**b**) and ENCODE cell lines (**c**).

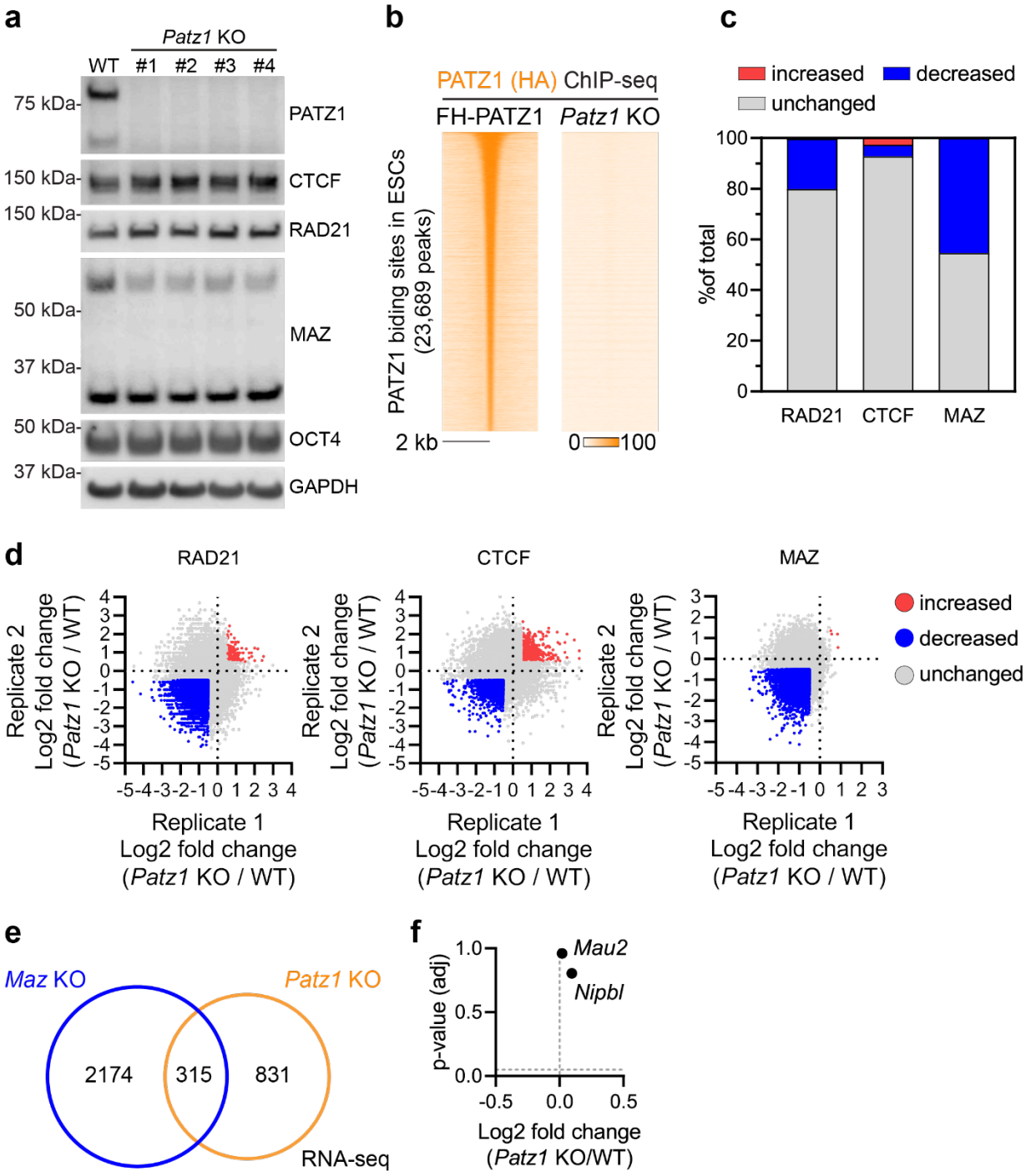

a
Patz1 KO
WT #1 #2 #3 #4
75 kDa-
PATZ1
150 kDa-
CTCF
150 kDa-
RAD21
50 kDa-
MAZ
37 kDa-
50 kDa-
OCT4
37 kDa-
GAPDH
b
PATZ1 (HA) ChIP-seq
FH-PATZ1 Patz1 KO
PATZ1 biding sites in ESCs (23,689 peaks)
2 kb
0 100
c
increased
decreased
unchanged
%of total
RAD21 CTCF MAZ
d
RAD21
CTCF
MAZ
Replicate 2 Log2 fold change (Patz1 KO / WT)
Replicate 1 Log2 fold change (Patz1 KO / WT)
increased
decreased
unchanged
e
Maz KO
Patz1 KO
2174
315
831
RNA-seq
f
Mau2
Nipbl
p-value (adj)
Log2 fold change (Patz1 KO/WT)

Figure S6 | The effects of *Patz1* depletion in mESCs. a. Western blot analysis of PATZ1, CTCF, RAD21, MAZ, OCT4, and GAPDH in WT and *Patz1* KO clones. **b.** Heat maps of PATZ1 (HA) ChIP-seq read densities in FH-PATZ1 and *Patz1* KO mESCs at PATZ1 peaks, showing the specificity of PATZ1 ChIP-seq signals. **c.** Changes in ChIP-seq densities of RAD21, CTCF, and MAZ in *Patz1* KO with the cutoff of ±0.5 $\log_2$(fold change) observed in two independent biological replicates. **d.** Ratio of RAD21, CTCF, and MAZ ChIP-seq densities between *Patz1* KO and WT from 2 independent biological replicates. **e.** Venn diagram indicating the overlap of differentially expressed genes upon *Patz1* KO and *Maz* KO in mESCs from three biological replicates. **f.** The fold change in expression of the cohesin loader genes, *Nipbl* and *Mau2*, in mESCs was assessed by RNA-seq from three biological replicates.

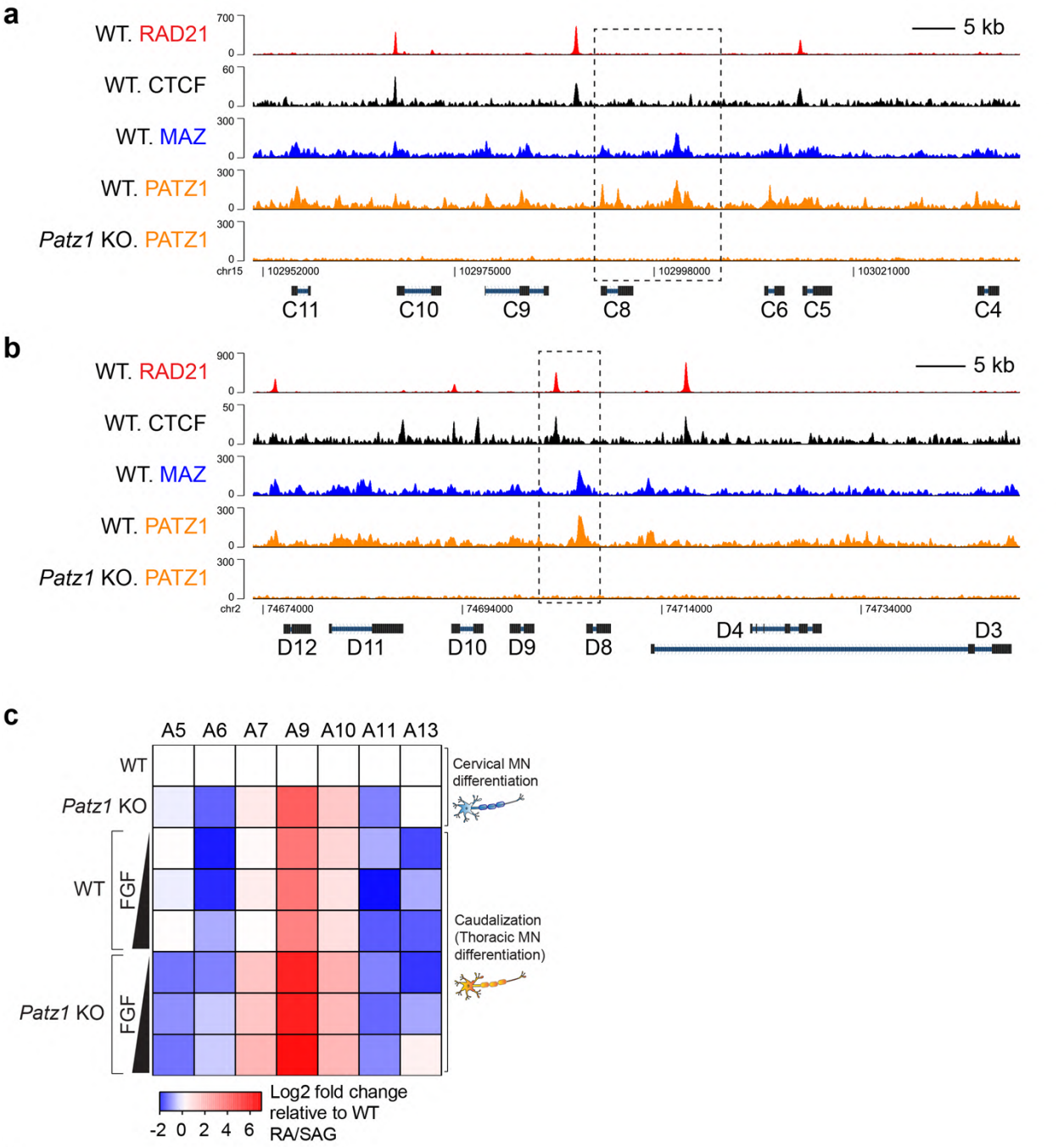

a
WT. RAD21
WT. CTCF
WT. MAZ
WT. PATZ1
Patz1 KO. PATZ1
5 kb
chr15
102952000
102975000
102998000
103021000
C11
C10
C9
C8
C6
C5
C4
b
WT. RAD21
WT. CTCF
WT. MAZ
WT. PATZ1
Patz1 KO. PATZ1
5 kb
chr2
74674000
74694000
74714000
74734000
D12
D11
D10
D9
D8
D4
D3
c
A5 A6 A7 A9 A10 A11 A13
WT
Patz1 KO
WT
FGF
Patz1 KO
FGF
Cervical MN differentiation
Caudalization (Thoracic MN differentiation)
Log2 fold change relative to WT RA/SAG
-2 0 2 4 6
d
WT
Patz1 KO
Log2 (normalized count + 1)
Hoxa1
Hoxa2
Hoxa3
Hoxa4
Hoxa5
Hoxa6
Hoxa7
Hoxa9
Hoxa10
Hoxa11
Hoxa13
Hoxc4
Hoxc5
Hoxc6
Hoxc8
Hoxc9
Hoxc10
Hoxc11
Hoxc12
Hoxc13
Hoxd1
Hoxd3
Hoxd4
Hoxd8
Hoxd9
Hoxd10
Hoxd11
Hoxd12
Hoxd13

Figure S7 | Loss of PATZ1 results in de-repression of *Hoxa9*, *Hoxc9* and *Hoxd9* in cervical MNs, indicating a rostro-caudal patterning defect in MNs. a, b. Normalized ChIP-seq densities for RAD21, CTCF, MAZ, and PATZ1 in WT and *Patz1* KO mESCs at the indicated region in the *HoxC* (**a**) and *HoxD* (**b**) cluster. **c.** Heat map of relative gene expression in WT versus *Patz1* KO MNs at the *HoxA* cluster in cervical MNs and thoracic MNs from one biological replicate. WNT/FGF signaling gradient was utilized for MN caudalization, as described previously (see Methods for details). **d.** RNA-seq normalized counts for *Hox* gene expression across *HoxA, C,* and *D* clusters in WT versus *Patz1* KO ESCs from three biological replicates.

Table S1. (separate file)

Public datasets analyzed in this study

Table S2. (separate file)

RNA-seq expression values in WT vs *Patz1* KO mESCs

Table S3. (separate file)

Gene Ontology Analysis in WT vs *Patz1* KO mESCs

Table S4. (separate file)

RNA-seq expression values in WT vs *Patz1* KO MNs

Table S5. (separate file)

Gene Ontology Analysis in WT vs *Patz1* KO MNs

Table S6. (separate file)

DNA oligos used in this study

Table S7. (separate file)

Antibodies used in this study

References

1. Ortabozkoyun, H. *et al.* CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at Hox clusters. *Nat Genet* **54**, 202-212 (2022).
2. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
3. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-5 (2012).
4. Phillips-Cremins, J.E. & Corces, V.G. Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**, 461-74 (2013).
5. Van Bortle, K. *et al.* Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res* **22**, 2176-87 (2012).
6. Moon, H. *et al.* CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep* **6**, 165-70 (2005).
7. Gurudatta, B.V. & Corces, V.G. Chromatin insulators: lessons from the fly. *Brief Funct Genomic Proteomic* **8**, 276-82 (2009).
8. Bell, A.C., West, A.G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396 (1999).
9. Lobanenkov, V.V. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. (1990).
10. Nora, E.P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e22 (2017).

11. Rao, S.S.P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e24 (2017).

12. Davidson, I.F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345 (2019).

13. Li, Y. *et al.* The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472-476 (2020).

14. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).

15. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**, 676-84 (2015).

16. Nora, E.P. *et al.* Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* **11**, 5612 (2020).

17. Xiao, T., Li, X. & Felsenfeld, G. The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proc Natl Acad Sci U S A* **118**(2021).

18. Wellik, D.M. Chapter 9 Hox Genes and Vertebrate Axial Pattern. in *Hox Genes* 257-278 (2009).

19. Narendra, V., Bulajic, M., Dekker, J., Mazzoni, E.O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev* **30**, 2657-2662 (2016).

20. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* **19**, 24-32 (2009).

21. Valton, A.L. *et al.* A cohesin traffic pattern genetically linked to gene regulation. *Nat Struct Mol Biol* **29**, 1239-1251 (2022).

22. Hsieh, T.S. *et al.* Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet* **54**, 1919-1932 (2022).

23. Davidson, I.F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated DNA loop extrusion. *bioRxiv*, 2022.09.08.507093 (2022).

24. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345-1349 (2019).

25. Bell, A.C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-5 (2000).

26. Chung, J.H., Whiteley, M. & Felsenfeld, G. A 5' Element of the Chicken Beta-Globin Domain Serves as an Insulator in Human Erythroid-Cells and Protects against Position Effect in Drosophila. *Cell* **74**, 505-514 (1993).

27. Recillas-Targa, F. *et al.* Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* **99**, 6883-8 (2002).

28. Dickson, J. *et al.* VEZF1 Elements Mediate Protection from DNA Methylation. *PLOS Genetics* **6**, e1000804 (2010).

29. Kobayashi, A. *et al.* A Combinatorial Code for Gene Expression Generated by Transcription Factor Bach2 and MAZR (MAZ-Related Factor) through the BTB/POZ Domain. **20**, 1733-1746 (2000).

30. Zlatanova, J. & Caiafa, P. CTCF and its protein partners: divide and rule? *Journal of Cell Science* **122**, 1275-1284 (2009).

31. Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737-743 (2020).

32. Dasen, J.S. & Jessell, T.M. Chapter Six Hox Networks and the Origins of Motor Neuron Diversity. in *Hox Genes* 169-200 (2009).

33. Mazzoni, E.O. *et al.* Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci* **16**, 1191-1198 (2013).

34. Bauer, B.W. *et al.* Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. *Cell* **184**, 5448-5464 e22 (2021).

35. Oldenkamp, R. & Rowland, B.D. A walk through the SMC cycle: From catching DNAs to shaping the genome. *Mol Cell* **82**, 1616-1630 (2022).

36. Pradhan, B. *et al.* SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Reports* **41**, 111491 (2022).

37. Narendra, V.R., P.P. ; An, D.; Raviram, R.; Skok, J.A.; Mazzoni,E.O.; Reinberg, D. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017-21 (2015).

38. Jung, H. *et al.* Global control of motor neuron topography mediated by the repressive actions of a single hox gene. *Neuron* **67**, 781-96 (2010).

39. Sweeney, L.B. *et al.* Origin and Segmental Diversity of Spinal Inhibitory Interneurons. *Neuron* **97**, 341-355 e3 (2018).

40. Xu, B. & Wellik, D.M. Axial Hox9 activity establishes the posterior field in the developing forelimb. *Proc Natl Acad Sci U S A* **108**, 4888-91 (2011).

41. Liu, N.Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat Genet* **53**, 100-109 (2021).

42. West, A.G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes Dev* **16**, 271-88 (2002).

43. Van Bortle, K. *et al.* Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* **15**, R82 (2014).

44. Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat Genet* **53**, 1064-1074 (2021).

45. Kang, J., Kim, Y.W., Park, S., Kang, Y. & Kim, A. Multiple CTCF sites cooperate with each other to maintain a TAD for enhancer-promoter interaction in the beta-globin locus. *FASEB J* **35**, e21768 (2021).

46. Kyrchanova, O. & Georgiev, P. Chromatin insulators and long-distance interactions in Drosophila. *FEBS Lett* **588**, 8-14 (2014).

47. Maksimenko, O. *et al.* Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res* **25**, 89-99 (2015).

48. Chetverina, D., Erokhin, M. & Schedl, P. GAGA factor: a multifunctional pioneering chromatin protein. *Cell Mol Life Sci* **78**, 4125-4141 (2021).

49. Melnikova, L. *et al.* Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proc Natl Acad Sci U S A* **101**, 14806-11 (2004).

50. Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O. & Corces, V.G. A Drosophila Protein That Imparts Directionality on a Chromatin Insulator Is an Enhancer of Position-Effect Variegation. *Cell* **82**, 587-597 (1995).

51. Bardwell, V.J. & Treisman, R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* **8**, 1664-77 (1994).

52. Aziati, I.D. *et al.* PATZ1 is required for efficient HIV-1 infection. *Biochem Biophys Res Commun* **514**, 538-544 (2019).

53. Saldana-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev* **28**, 723-34 (2014).

54. Kaye, E.G. *et al.* Differential Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage Compensation Complex to the Male X Chromosome. *Cell Rep* **22**, 3227-3239 (2018).

55. Bag, I., Dale, R.K., Palmer, C. & Lei, E.P. Correction: The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila (doi:10.1242/jcs.226092). *J Cell Sci* **132**(2019).

56. Jordan, W., 3rd & Larschan, E. The zinc finger protein CLAMP promotes long-range chromatin interactions that mediate dosage compensation of the Drosophila male X-chromosome. *Epigenetics Chromatin* **14**, 29 (2021).

57. Mazzoni, E.O. *et al.* Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci* **16**, 1191-8 (2013).

58. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research* **11**, 1475-1489 (1983).

59. LeRoy, G. *et al.* LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. *Sci Adv* **5**, eaay3068 (2019).

60. LeRoy, G., Rickards, B. & Flint, S.J. The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Mol Cell* **30**, 51-60 (2008).

61. Oksuz, O. *et al.* Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol Cell* **70**, 1149-1162 e5 (2018).

62. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

63. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).

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

65. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P.D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**, D419-D426 (2019).

66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).

67. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

68. Lerdrup, M., Johansen, J.V., Agrawal-Singh, S. & Hansen, K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nature Structural & Molecular Biology* **23**, 349-357 (2016).

69. Zhu, L.J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).

70. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

71. Nguyen, N.T.T. *et al.* RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res* **46**, W209-W214 (2018).

72. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).

73. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity between motifs. *Genome Biol* **8**, R24 (2007).

74. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications* **9**, 189 (2018).

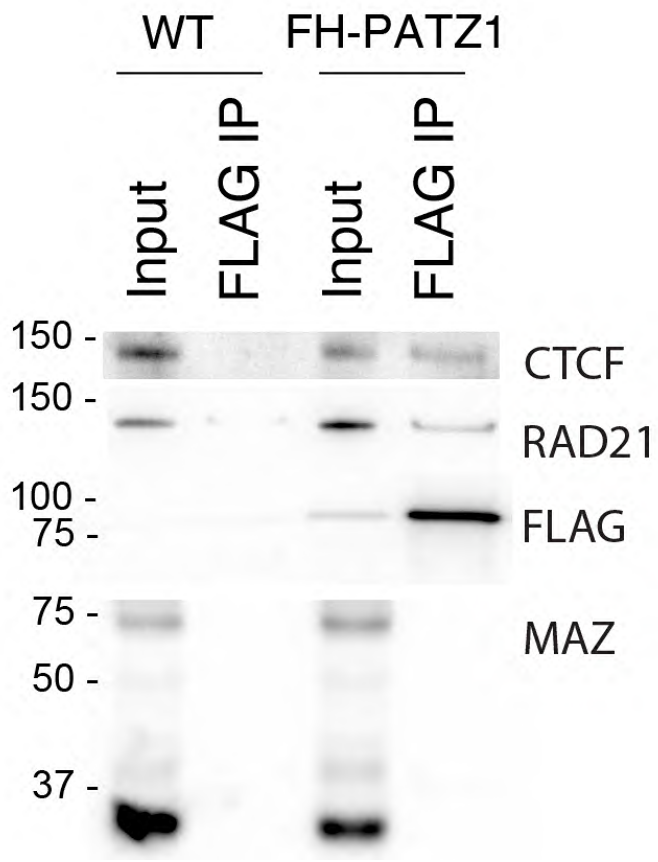
75. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets *Bioinformatics* **37**, 422-423 (2020).

76. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).

77. Durand, N.C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-8 (2016).

78. Greenwald, W.W. *et al.* Pgltools: a genomic arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. *BMC Bioinformatics* **18**, 207 (2017).

Supplementary Figures (Biological Replicate for Figure 3c)



Supplementary Figure 1 | PATZ1 co-immunoprecipitated with RAD21 in 293FT cells. Western blot analysis of CTCF, RAD21, FLAG, and MAZ upon FLAG-PATZ1 immunoprecipitation from nuclear extract of 293FT cells.