Binding domain mutations provide insight into CTCF's relationship with chromatin and its contribution to gene regulation

Catherine Do[1,2,*], Guimei Jiang[1,2,*], Giulia Cova[1,2*], Christos C. Katsifis[3,4,5,φ], Domenic N. Narducci[3,4,5,φ], Theodore Sakellaropoulos[1,2], Raphael Vidal[1,2], Priscillia Lhoumaud[1,2], Aristotelis Tsirigos [1,2,6], Faye Fara D. Regis[1,2], Nata Kakabadze[1,2], Elphege P Nora[7], Marcus Noyes[8], Anders S. Hansen[3,4,5], Jane A Skok[1,2,†]

1. Department of Pathology, NYU Grossman School of Medicine, New York, NY 10016, USA
2. Perlmutter Cancer Center, NYU Langone Health, New York, NY 10016, USA
3. MIT Department of Biological Engineering, Cambridge, MA 02139, USA
4. Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA
5. Koch Institute for Integrative Cancer Research, Cambridge, MA, 02139, USA
6. Applied Bioinformatics Laboratories, Office of Science & Research, NYU Grossman School of Medicine, New York, NY 10016, USA
7. Cardiovascular Research Institute, and Department of Biochemistry and Biophysics, University of California San Francisco, CA 94143, USA
8. Institute for Systems Genetics, Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA

[*, φ] These authors contributed equally.
[†] Lead contact
Corresponding author:
Jane.Skok@nyulangone.org

Summary 150 words.

Here we used a series of CTCF mutations to explore CTCF's relationship with chromatin and its contribution to gene regulation. CTCF's impact depends on the genomic context of bound sites and the unique binding properties of WT and mutant CTCF proteins. Specifically, CTCF's signal strength is linked to changes in accessibility, and the ability to block cohesin is linked to its binding stability. Multivariate modelling reveals that both CTCF and accessibility contribute independently to cohesin binding and insulation, however CTCF signal strength has a stronger effect. CTCF and chromatin have a bidirectional relationship such that at CTCF sites, accessibility is reduced in a cohesin-dependent, mutant specific fashion. In addition, each mutant alters TF binding and accessibility in an indirect manner, changes which impart the most influence on rewiring transcriptional networks and the cell's ability to differentiate. Collectively, the mutant perturbations provide a rich resource for determining CTCF's site-specific effects.

Introduction

The CCCTC-binding factor (CTCF) is an eleven-zinc finger DNA-binding protein that together with its binding partner, cohesin plays a key role in organizing chromatin into TAD structures by promoting the formation of loops and boundaries that are important for gene regulation. This involves a loop-extrusion mechanism in which cohesin complexes create chromatin loops by actively extruding DNA until movement of the complex is blocked by two CTCF binding sites in convergent orientation[1,2]. CTCF, in conjunction with cohesin, is enriched at TAD boundaries that function as insulators, contributing to gene regulation by restricting the interaction of regulatory elements to promoters of target genes located within the same TAD[3,4].

CTCF has been degraded by auxin in numerous cell types, and although these studies provide important insight into its role in gene regulation, degron models are insufficient for (i) analyzing the contribution of genomic context to site-specific CTCF binding and function, or (ii) distinguishing direct from indirect effects that can be a confounding issue for interpreting site specific impact. Instead, information related to CTCF's direct effects on individual loci has come from the genetic ablation of CTCF binding sites at specific regions in the genome in a given cell type. However, interrogating the general mechanisms underlying the crosstalk between CTCF function and the cell- and locus-specific context of its binding by genetic manipulation of individual CTCF binding sites is laborious and has the limitation of providing insight into the control of only a limited number of loci in the neighboring region.

The *CTCF* locus, located on chromosome 16q band 22, corresponds to one of the smallest regions of overlap for common deletions in breast and prostate cancers[5,6]. Moreover, point mutations and deletions have been identified in many other tumors. Together these findings indicate that CTCF acts as a tumor suppressor[7]. Additionally, genetic alterations in *CTCF* and changes in dosage are associated to varying degrees with intellectual disability and microcephaly, and thus CTCF is thought to play an important role in brain development and neurological disorders[8-10].

CTCF binds to chromatin through a subset of its eleven zinc fingers (ZFs)[11,12]. Specifically, ZFs3-7 make sequence specific contacts with a 12-15 base-pair (bp) core consensus binding site. While the crystal structure of CTCF in complex with a known DNA binding site[13] has revealed new insights into the contribution of each ZF to sequence-specific binding in an *in vitro* setting, there has been no complementary, comprehensive analysis of cancer and neurological development

disorders (NDDs) associated CTCF mutants and it is thus not known how mutations within each ZF impact binding stability and DNA sequence specificity, and the consequences this has on binding profile, cohesin overlap, accessibility, chromosome architecture and gene regulation. Given this lack of information, we have no context in which to determine the functional relevance of mutations.

Using an inducible mouse ESC complementation system, we combined genetic, imaging, molecular, AlphaFold3 and bioinformatic analyses to examine the impact of nine, high frequency cancer associated CTCF mutations, a subset of which are also associated with NDDs. Selected mutations occur in eight amino acid residues of ZFs that contact the core consensus binding motif. Compared to a complete knockout of CTCF or deletion of a CTCF binding site, the mutants offer a more subtle and in-depth approach for teasing out (i) the relationship between WT and mutant CTCF and chromatin, and (ii) direct or indirect site-specific effects on accessibility, TF binding, cohesin overlap, chromatin interactivity and gene expression, since each mutation produces its own unique effect.

Here we demonstrate that the functional impact of WT and mutant CTCF depends both on the genomic context of bound sites and the unique binding properties of each WT and mutant CTCF protein. Analyses of the relationship between CTCF, cohesin and accessibility reveals that although CTCF can bind both accessible and inaccessible sites, inaccessible bound sites predominantly have weaker signals and are less competent at blocking cohesin. Multivariate modelling demonstrates that both CTCF and accessibility contribute independently to cohesin binding, however CTCF signal strength has a dominant effect, linked to a stronger insulation score. Indeed, CTCFs ability to block cohesin is linked to the residence time or ‘OFF’ rate of each mutant on chromatin as determined by FRAP experiments, while occupancy or ‘ON’ rate is correlated with their binding at common accessible sites, indicating that accessibility dominates the search space. Together, the mutants highlight the fact that occupancy and residence time can be uncoupled such that although some mutants have fewer binding sites than WT CTCF, they have a stronger residence time at these locations which increases their ability to block cohesin and contribute to chromatin folding. The reverse situation also occurs. Our analyses also highlight that in contrast to CTCF-bound sites, cohesin binding at CTCF-free sites is weak and does not track with accessibility even at enhancers and promoters, perhaps because TFs that block cohesin have considerably shorter residence times compared to CTCF.

114 The unique properties of WT and mutant CTCF perturb TF binding and cancer, brain, immune
115 and metabolic gene expression pathways. Mutants have a more pronounced effect on
116 transcriptional networks than auxin inducible degradation, because of indirect effects that can be
117 explained by altered TF binding profiles. Indeed, indirect effects that do not overlap changes in
118 CTCF binding impart the most influence on rewiring chromatin looping and transcriptional
119 networks. Changes in each mutant's transcriptional program in turn have an impact on the cell's
120 ability to exit from pluripotency and differentiate. Collectively, the mutants offer a rich resource to
121 investigate site specific CTCF-mediated effects on chromatin folding and gene regulation, while
122 distinguishing cause from consequence and direct from indirect outcomes.
123
124 **RESULTS**
125 **CTCF complementation system**
126 To explore CTCF's relationship with chromatin and its contribution to gene regulation, we made
127 use of a small-molecule auxin inducible degron (AID) mouse ESC (mESC) system[14,15]. In these
128 cells, both endogenous CTCF alleles are tagged with AID as well as eGFP (CTCF-AID-eGFP)
129 and the auxin-activated ubiquitin ligase TIR1 (from *Oryza sativa*) is constitutively expressed from
130 the *Rosa26* locus (**Figure 1A)**. Addition of indole acetic acid (IAA), an analogue of auxin leads to
131 rapid poly-ubiquitination and proteasomal degradation of the AID tagged protein[16].
132
133 To study the impact of CTCF mutations, we established a rescue system, modifying the mESC
134 degron cell line to express either a stable doxycycline-inducible control wild-type *Ctcf* or a mutant
135 *Ctcf* (*mCtcf*) transgene in the absence of endogenous CTCF[17]. Transgenes, knocked into the
136 *Tigre* locus, have a 3 x FLAG tag, which allows us to distinguish mutant or wild-type transgenic
137 CTCF from endogenous CTCF using a FLAG antibody in Western blot and ChIP-seq analysis.
138 An mRuby fluorescent tag enables us to accurately determine expression levels of mutant or wild-
139 type transgenic CTCF by flow cytometry (**Figure 1A**). The three conditions used for our analysis
140 are shown in **Figure 1B**, and an example of the FACS profile for cells expressing a WT CTCF
141 transgene in each condition is shown (**Figure 1C**).
142
143 We selected a subset of missense mutations (identified from cBioPortal [18] and COSMIC[19]) that
144 occur at high frequency in cancer patients (**Figure S1**), focusing on those found in ZFs that
145 contact the 12-15bp consensus CTCF binding motif (**Figure 1D** and highlighted with an asterisk
146 in **Figure S2**). We analyzed mutations in **(i)** amino acids that make base-specifying contacts with
147 the alpha helix of DNA (at locations -2, 2, 3 and 6 on the ZF), **(ii)** residues that coordinate the zinc

148 ion which are essential for providing stability to ensure the proper folding of the ZF domain, **(iii)**
149 two other classes of amino acid residues in the ZFs: boundary residues (that are important for
150 interactions between ZFs), and residues that contact the sugar phosphate backbone of DNA, and
151 **(iv)** other highly mutated residues in the region. Each class of mutation is color-coded as shown
152 in **Figure 1D** (left) and **Figures S1-2** and the color code is maintained throughout the figures.
153
154 In total, we analyzed mutations in eight amino acid residues and for R377, the most highly mutated
155 residue in cancer, we included two distinct mutations of R377, both of which (R377H and R377C)
156 are equally represented in patients (**Figure S2**). Selected mutations (R339Q, R342C, R448Q,
157 R377H and R377C) are also implicated in NDDs (**Figure S2**)[10,20]. Individual clones with
158 comparable levels of transgene expression (mutant and WT) were selected based on flow
159 cytometry and Western blot analysis (**Figure 1E, F and Figures S3-4**). To assess rescue of
160 endogenous CTCF by WT transgenic CTCF, we performed ChIP-seq using an anti-CTCF
161 antibody and RNA-seq in WT untreated and WT transgene expressing (ID) cells. This analysis
162 showed that about half of the overall CTCF binding sites (but 97% of the CTCF motif-containing
163 sites) and gene expression changes are rescued by transgenic CTCF (**Figure S5**), consistent
164 with the lower expression of the transgene reported previously[17]. This data provides a rationale
165 for comparing the effect of the mutant and WT transgenic CTCF.
166
167 **Retained binding sites are enriched in accessible regions with strong CTCF signals**
168 We next analyzed the profiles of mutant versus WT CTCF using cells expressing transgenic WT
169 and mutant CTCF in the absence of endogenous CTCF (ID). Briefly, we degraded endogenous
170 CTCF-mAID-GFP by IAA addition, induced WT or mutant transgene expression with Dox for 48
171 hours, and then performed FLAG ChIP-seq. For each mutation we identified WT only sites (bound
172 by WT CTCF that mutants no longer bind), common sites (bound by both WT and mutant proteins)
173 and mutant only sites (*de novo* sites that only mutant proteins bind)**.**
174
175 The location of each mutation and the complementary DNA triplet that the relevant ZF binds is
176 marked by a box in the consensus motif shown in **Figure 2A** and **Figure S6A**. Below that we
177 highlight the most frequent motif identified by MEME[21] for the set of *de novo* sites**.** These motifs
178 differ from that of the consensus sequence by either (i) a diminished requirement for DNA bases
179 that contact the mutated zinc finger (**Figure 2A**), or (ii) an alteration to the sequence of the
180 consensus motif, which could be explained by the impact of mutations on DNA-base interacting
181 zinc finger residues. For example, K365 binds to the second guanine in the triplet GGC of the

182 consensus motif and the *de novo* motif found in the K365T mutant is consistent with loss of
183 hydrogen bonds with the central guanine due to the threonine substitution, as previously
184 demonstrated[13]. Motifs found at WT only and common sites were similar to the consensus motif
185 (**Figure S6B**), with the exception of R377C for which few WT only sites were observed. On
186 average, the percentages of CTCF containing motifs at common sites were higher than at WT
187 only sites, suggesting that the latter might be lower affinity sites with degenerate motifs, consistent
188 with the weaker CTCF signal observed. Each mutant displays their own unique binding profiles
189 including the two distinct mutations of residue R377. R377C binds to less common and more *de*
190 *novo* sites compared to R377H and each mutant binds a motif that is similar to the consensus
191 motif but lacking bases associated with ZF3 (**Figure 2A** and **Figure S6**).
192
193 The bar graphs show the percentage of WT only, common and mutant only binding sites, while
194 the heatmaps reveal the number of binding sites for each subset and highlight differences in
195 CTCF ChIP-seq signal between WT and mutants at the three locations. Differential binding
196 analysis using DiffBind (**Figure S7)** showed that only a fraction of the mutant and WT only peaks
197 reached statistical significance, which can partly be explained by the fact that these peaks are
198 weaker and have a smaller effect size compared to the stronger, common peaks. However, the
199 ratios of gained versus lost CTCF binding sites are consistent with the proportion of 'WT only' and
200 'mutant only' peaks (**Figure 2A** and **S6A**). CTCF signal strength is strongest at common sites,
201 however, CTCF this is altered in a mutant specific manner, suggesting that a 'binding dosage
202 effect' contributes to their functional impact. As expected, the H455R mutation in the zinc
203 coordinating residue of ZF7 loses the most binding sites. OmniATAC[22] profiles for the different
204 binding groups indicate that CTCF ChIP-seq signal strength is a function of chromatin
205 accessibility, with common sites being the most accessible (**Figure 2A and Figure S6A**).
206
207 In sum, ChIP-seq analysis reveals that CTCF mutants lose, retain and gain a subset of binding
208 sites throughout the genome with mutant specific profiles. Loss and gain of CTCF binding occurs
209 predominantly at weak inaccessible CTCF sites, while binding is retained at the stronger common
210 binding sites which have high accessibility, although CTCF ChIP-seq signal and accessibility can
211 vary depending on the mutant.
212
213 **CTCF reduces ATAC-seq signal in a cohesin-dependent manner**
214 To further investigate the link between accessibility and CTCF binding, we compared the ATAC-
215 seq signal at CTCF-bound versus unbound sites in the presence or absence of SMC3, a subunit

216 of the cohesin complex (**Figure 2B** and **Figure S8**). Accessibility was reduced in an SMC3-
217 dependent manner at both WT and mutant CTCF-bound sites. Each CTCF mutant exhibited a
218 differential ability to decrease the height and breadth of ATAC-seq peaks, suggesting that CTCF
219 is driving the change in accessibility and that the relationship between CTCF and accessibility is
220 bidirectional. We do not observe the same phenomenon at sites without SMC3, which lends
221 support to the idea that loop formation decreases accessibility, indicating a unique chromatin
222 structure at loop bases. The same phenomenon is observed at enhancer and promoter sites,
223 although the effect is weaker, particularly at promoters (**Figure S9**). The weakened effect may be
224 due to interference of TF binding which can affect CTCF-dependent nucleosome phasing and
225 insulation at CTCF bound sites or competitive binding with CTCF[23]. These results highlight the
226 usefulness of the mutants in distinguishing cause from effect.
227
228 **Mutations reduce the chromatin residence time and bound fraction of CTCF**
229 Fluorescence recovery after photobleaching (FRAP) was performed on the 9 mutants and WT
230 CTCF[24,25], using the same auxin and Dox conditions as described for the ChIP-seq and ATAC-
231 seq. For the FRAP a circular 1 μm diameter circle was bleached, and recovery monitored over 10
232 minutes, recording 30 movies per condition in three replicates. While the recovery of WT CTCF
233 largely matched prior CTCF FRAP results[24], all mutants exhibited faster recovery consistent with
234 reduced and/or less stable DNA-binding (**Figure 3A and S10**). We fit the FRAP curves to a
235 reaction-dominant kinetic model [24,25] and estimated the residence time (duration of CTCF binding
236 to DNA) and bound fraction (proportion of total CTCF bound to DNA) of each mutant with
237 subsequent comparison to the WT (**Figure 3B, C**). We found a general reduction in the residence
238 time and/or bound fraction across each mutant, with WT CTCF exhibiting the largest specific
239 bound fraction and residence time (**Figure 3B, C**). Each mutant displayed unique parameters,
240 regardless of the mutant category. The R377H and R377C mutants – mutations in the phosphate
241 contacts – have similar bound fractions, but distinct residence times. The mutations in the amino
242 acids making direct contact with DNA (R339Q, K365T, Q418R, and R448Q) have varied
243 residence times and bound fractions, despite belonging to the same group. R448Q showed the
244 second lowest bound fraction, despite exhibiting a residence time comparable to WT. Q418R had
245 the lowest residence time of all mutants examined, with an intermediate bound fraction.
246 Conversely, K365T exhibited a bound fraction and residence time most closely comparable to
247 WT. H455R, which assists in coordinating the zinc, had the lowest bound fraction and the second
248 lowest residence time of all examined mutants, accurately reflecting the most deleterious mutation
249 to canonical CTCF function. R342C, which occurs in a boundary residue between the two zinc-

250 ligand histidine residues of ZF3, showed a residence time similar to that of wild type, but had a
251 lower bound fraction, possibly suggesting a role for boundary residues in binding efficiency.
252 Lastly, S354F, located between the two zinc-ligand cysteine residues of ZF4, which substitutes a
253 bulky phenylalanine in place of a serine in ZF4, showed both a moderate decrease in bound
254 fraction as well as residence time.
255
256 In summary, our FRAP results demonstrate that all mutants show changes in DNA binding either
257 through an altered residence time or bound fraction. Surprisingly, some mutations strongly affect
258 either the residence time or the bound fraction, with minimal or less of an effect on the other. This
259 suggests that the CTCF target search (ON-rate) can be decoupled from binding stability (OFF
260 rate) [26], such that some mutants affect the efficiency of the search for binding sites, without
261 affecting the residence time of CTCF.
262
263 **The impact of mutations on the DNA-CTCF complex**
264 To assess the impact of each mutation on the DNA-CTCF complex, we generated predictions
265 using AlphaFold3. First, we compared the WT and K365T mutant DNA-CTCF complexes between
266 the predicted and experimental structure and found a high concordance (**Figure S11**). We then
267 predicted the DNA-CTCF (ZF1-11) structure for each mutant and compared the alignment against
268 WT (**Figure S12**). Like K365T, the mutations have little effect on the overall conformation of
269 CTCF. However, Q418R the mutant with the lowest residence time has the highest deviation from
270 WT (RMSD=1.6 Å), although the RMSD is still <2Å, which is considered a good alignment with
271 the WT structure. It is of note that AlphaFold3 does not take account of the effect of ions, therefore,
272 the prediction for H455R might not be accurate since the mutation disrupts the coordination of
273 H455 with the zinc ion. Some of the mutations found at residues making direct contact with the
274 DNA, have lost or weakened hydrogen bonds (K365T, R339Q and R448Q). Q418R gains
275 hydrogen bonds (**Figure S12**), which could explain the paradox of this mutation: that it has the
276 lowest residence time but strong binding at common sites.
277
278 **CTCF mutations have a graded impact on binding and function**
279 To understand the relationship between FRAP and ChIP-seq data, we asked whether the
280 residence time is linked to CTCF's binding stability and ability to block cohesin. In **Figure S13A**
281 we show the stratified overlap of cohesin/FLAG at WT only, common and mutant only CTCF sites.
282 Overlap with cohesin is lower at WT only and mutant only sites, consistent with their weaker CTCF
283 and ATAC-seq signals. These changes do not reflect changes in cohesin protein levels as

284 Western blots demonstrate that CTCF mutants do not impact the levels of RAD21 or SMC1, two
285 subcomponents of the cohesin complex (**Figure S13B**). The proportion of global CTCF-cohesin
286 versus CTCF only binding sites shown in the heatmaps (**Figure 3D**), demonstrate that CTCF-
287 cohesin overlap is graded across mutants, to some extent mirroring a gene dosage effect that
288 could occur as a consequence of altered residence time (**Figure 3C**). Indeed, analysis of SMC3
289 ChIP-seq revealed that the residence time was strongly associated (R^2=0.8, p=0.002) with the
290 percentage of CTCF-cohesin overlap for each mutant (**Figure 3E**) which reflects the impact of
291 binding stability on CTCF's function in loop extrusion. Conversely, the proportion of SMC3
292 overlapping CTCF might reflect changes in the number of CTCF peaks across the mutants since
293 the expression level of components of the cohesin complex, SMC1 and RAD 21 do not change
294 in CTCF mutant expressing cells. Indeed, **Figure S14A** shows a strong correlation between the
295 percentage of SMC3 peaks overlapping CTCF and the number of CTCF peaks (**Figure S14B**).
296
297 We next compared the specific bound fraction estimated by FRAP, to the bound fraction using
298 FLAG ChIP-seq and ATAC-seq peaks to define all potential binding sites. While we did not find a
299 good correlation when assessing the overall CTCF ChIP-seq bound fraction (**Figure S14C**), there
300 was a significant correlation (R^2=0.6, p=0.008) between the chromatin bound fraction of the
301 mutants and the fraction of mutant binding at common CTCF sites (**Figure 3F**). These data
302 suggest that the FRAP bound fraction mostly captures the effect of the mutations on the strong,
303 accessible binding sites. Furthermore, we found a strong correlation with the fraction bound and
304 the overall ability of CTCF-cohesin overlapping sites to reduce accessibility, (R^2=0.7, p=0.003),
305 in line with CTCF functioning in this context in a dose dependent manner (**Figure 3G**).
306
307 **The relationship between cohesin and accessibility at CTCF-bound and CTCF-free sites**
308 Consistent with previous findings[27], our data links accessibility with CTCF binding signal strength.
309 Furthermore, the FRAP data support an association between CTCF residence time and cohesin
310 overlap, which is also supported by simulation analysis[28]. To separate out the contribution of
311 CTCF and ATAC strength on SMC3 signals, we performed multivariate modelling to determine (i)
312 the effect of CTCF strength on SMC3 overlap, adjusted for the effect of ATAC signal, and (ii) the
313 effect of ATAC strength on SMC3 overlap, adjusted for the effect of CTCF signal. **Figure 4A** (top)
314 shows that SMC3 enrichment independent of ATAC-seq signal, is strongly associated with CTCF
315 signal strength in a dose-dependent manner. Only strong accessibility significantly influences
316 SMC3 binding independent of CTCF, however, accessibility has much less contribution compared
317 to CTCF binding stability **(Figure 4A**, bottom). Thus, CTCF signal strength plays a dominant role

318 in determining the probability of CTCF-cohesin overlap, consistent with the FRAP and ChIP-seq
319 data (**Figure 3E**). The heatmap in **Figure 4B** and **Table S1**, further shows that, in contrast to
320 CTCF-bound sites, the percentage of SMC3 overlap at CTCF-free sites is increased with
321 increasing accessibility, however signal strength at highly accessible sites is much lower than
322 sites overlapping even weak CTCF-bound sites. The same effect is observed for the mutants, but
323 in some cases SMC3 overlap at CTCF-bound sites is weaker than WT CTCF, in line with their
324 reduced residence times.
325
326 Overall SMC3 profiles at CTCF-bound sites confirmed that CTCF is most effective at blocking
327 cohesin at accessible sites enriched with strong CTCF signals, and further highlight that each
328 mutant CTCF performs this function with variable ability depending on their residence time
329 (**Figure 3E, 4C and Figure S15**). A comparison of SMC3 profiles at CTCF bound versus unbound
330 sites, demonstrates that other factors such as TFs, contribute very little compared to CTCF likely
331 because residence times of TFs on chromatin are much shorter than CTCF's[29]. Indeed, at CTCF-
332 free sites, the SMC3 profiles are similar at CTCF-free accessible and inaccessible sites,
333 suggesting that accessibility has little contribution to cohesin binding in the absence of CTCF. A
334 similar trend was observed at enhancers and promoters where accessibility is higher and TFs are
335 known to be enriched (**Figure S16**). These data suggest that for a given cell type, the subset of
336 strong and accessible CTCF binding sites are predominantly involved in loop extrusion. However,
337 weaker, inaccessible CTCF binding sites might also have a role in more transient, dynamic
338 looping events with a reduced loop lifetime[30,31]. Moreover, a subset of inaccessible sites (16% in
339 the WT) have strong CTCF signals.
340
341 **Each CTCF mutation alters gene expression and TF binding in a unique manner.**
342 To determine how CTCF mutations impact gene expression we performed RNA-seq and did an
343 unsupervised clustering analysis, comparing gene expression in cells in which endogenous CTCF
344 was degraded in the absence (IAA) or presence of the WT or mutant CTCF transgenes (ID).
345 Changes in transcriptional output were determined by comparison with cells expressing WT
346 transgenic CTCF (ID) (**Figure 5A**). Overall, 6474 genes had altered gene expression across all
347 mutants including the IAA condition (**Table S2**). **Figure S17A** shows that the majority of
348 differentially expressed genes (DEGs) are found only in one mutant.
349
350 Gene set enrichment analysis using the KEGG database[32] showed a strong enrichment in
351 functional pathways related to cancer, brain, immune and metabolic processes among DEGs

352 across mutants (color coded in **Figure 5B**, with genes outside these pathways represented in
353 black**)**. This finding is compatible with the human diseases associated with CTCF mutations. The
354 top DEGs of these pathways are shown in the bar graphs below. Examples in **Figure 5B** show
355 changes in pathway and gene expression in the absence of CTCF (IAA condition) and in cells
356 expressing the mutant CTCF transgenes, H455R, S354F, R448Q and R339Q (ID condition). All
357 other mutants are shown in **Figure S17B**.
358
359 We also observed an enrichment of dysregulated genes observed in patients with CTCF
360 associated neurodevelopmental disorders (NDDs[9]) (**Figure S17C and Table S2**). CTCF plays an
361 important role in imprinting, and alterations in imprinted genes could contribute to the CTCF-
362 dependent neurodevelopmental phenotype as suggested by congenital imprinting disorders[33].
363 Indeed, we found a strong enrichment of imprinted genes across all mutants (**Figure S17D and**
364 **Table S2**).
365
366 Together, our analyses demonstrate that although each mutant exhibits a unique impact on
367 binding and gene expression, all mutants affect common pathways, consistent with phenotypes
368 linked to human CTCF-associated disorders. Differences in the impact of each mutant provide
369 insight into CTCF-associated NDD inter-patient variability. Interestingly, cells in which
370 endogenous CTCF is degraded (IAA condition) show fewer gene expression changes compared
371 to mutant expressing cells, except for the R377C mutation (**Figure S17B**). Some mutants have
372 alterations in many pathways (R399Q), while others highlight changes in only a few (Q418R and
373 R377C) as shown in **Figure 5B and Figure S17B**, suggesting that in the latter, the alteration in
374 gene expression may not accumulate in specific pathways.
375
376 **Mutant specific alterations in transcriptional networks impact exit from pluripotency**
377 To investigate, whether mutant specific transcriptional networks could potentially affect exit from
378 pluripotency and the ability to differentiate we compared expression profiles of WT and a selection
379 of mutant CTCF expressing cells before and after 72 hours of LIF withdrawal[34]. The mutants
380 analyzed were selected because they cover a range of CTCF residence times as measured by
381 FRAP. R342C exhibits a similar residence time to WT, K365T an intermediate residence time and
382 Q418R has one of the lowest residence times. RNA-seq revealed that each mutant demonstrated
383 a distinct differentiation trajectory compared to WT cells, which are most skewed towards a
384 neuroectodermal fate (**Figure 5C** and **Table S3**). A subset of the DEGs is shown in **Figure S18**.
385

386 In the mutants, we identified several differentially expressed imprinted genes , such as H19 and
387 PEG10, as well as ZFP57, which plays an important role in regulating imprinting. To further assess
388 the biological relevance of the differentiation behavior, we annotated DEGs using the DECIPHER
389 developmental disorder database (https://www.deciphergenomics.org/ddd) (**Table S3**). We found
390 44 genes associated with developmental disorders, including NDDs such as *Camk2a*, *Cnksr2*,
391 *Sobp* that are involved in intellectual disability, and *Dchs1* which is responsible for periventricular
392 neuronal heterotopia. In addition, we detected genes responsible for congenital connective tissue
393 syndromes (*Fras1* in Fraser syndrome and *Fbn1* in Marfan syndrome) as well as genes involved
394 in skeletal dysplasia (*Sox9*) and congenital cardiomyopathy (*Speg*). While CTCF-related
395 congenital disorders are most commonly characterized by neurodevelopmental symptoms,
396 patients can also exhibit other congenital anomalies such as heart defects, craniofacial
397 malformations or cleft palate[10], suggesting the biological relevance of the changes in
398 differentiation behavior observed in our system.
399
400 **Transcriptional changes correspond to predicted alterations in TF binding at promoters**
401 To investigate whether mutant specific transcriptional changes could be explained by altered TF
402 binding, we performed a footprinting analysis of ATAC-seq data using the TOBIAS pipeline [35] in
403 cells which have endogenous CTCF degraded in the absence (IAA) or presence of WT or mutant
404 transgenes (ID). For this analysis replicates were merged and comparisons made with WT ID
405 cells. Overall, 195 TFs had predicted differential binding across all mutants including the IAA
406 condition. As a proof of principle, we detected predicted differential binding for CTCF, which has
407 altered binding in the mutants. Numerous other predicted differentially bound TFs (n = 194) were
408 identified as shown in **Figure 5D, S19A** and **Table S4.** About one third were found in only one
409 mutant (**Figure S19B and Table S4**). Annotation of differentially bound TFs revealed a significant
410 enrichment for TFs associated with NDDs (Odd Ratio =2.2, p=0.02) (**Table S4**).
411
412 We overlapped predicted differentially bound TFs with a region of 2kb around the promoters of
413 DEGs in cells expressing different CTCF mutants, to analyze the impact of predicted differential
414 binding of TFs on gene expression and observed a strong enrichment (**Figure 5E, Figure S20**
415 and **Table S4**)**.** Differentially expressed target genes were also significantly enriched in genes
416 associated with developmental disorders (OR=3.6, $p=7x10^{-36}$) and in differentially expressed
417 genes found in CTCF associated NDDs (OR=4, $p=10^{-63}$). It is of note, that the majority of the
418 predicted changes in TF binding at the promoters of DEGs do not overlap CTCF ChIP-seq peaks
419 or predicted differential CTCF binding. This data suggests that many of the mutant-mediated

420 transcriptional changes could be explained by indirect effects resulting from the disruption of
421 specific TF pathways.
422
423 **Figure 5E and Figure S20** highlight the impact of two of the top predicted differentially bound
424 TFs (CTCF and MBD2) at promoters of DEGs in different mutants. These are shown in volcano
425 plots for cells in which endogenous CTCF is degraded (IAA) and in cells expressing different
426 CTCF mutations (ID). MBD2 is a methyl-sensitive factor that belongs to the NurD chromatin
427 remodeling complex, that has been implicated in cancer and NDDS. NurD is also required for exit
428 from pluripotency and lineage commitment[36-39]. For each CTCF mutant, predicted differentially
429 bound TFs target the promoters of a distinct set of up (red) and down (blue) regulated genes.
430 Both CTCF and MBD2 exhibit an overall activator profile with a positive association between
431 binding and expression (**Figure 5E and Figure S20**). However, for some mutants, a weaker but
432 significant negative association was also observed suggesting that both factors can act as
433 activators or repressors depending on the cell-specific genomic context, consistent with previous
434 reports[38,40].
435
436 Examples of altered CTCF binding at the promoter of differentially expressed *Rerg* and predicted
437 differentially bound MYC at the promoter of differentially expressed *Brdt*, are shown in **Figure 5F.**
438 Loss of CTCF binding at the promoter of the *Rerg* gene leads to a decrease in its expression,
439 while expression of *Brdt* increases in line with increased predicted binding of MYC at its promoter.
440 The latter is an example of an indirect effect because CTCF is not bound at the promoter of this
441 gene. In sum, combined analysis of RNA-seq and ATAC-seq highlight the variable effects of each
442 CTCF mutation, connecting gene expression changes with 195 predicted differentially bound TFs,
443 to provide a deeper understanding of the factors that underlie the unique transcriptomic profile of
444 each CTCF mutation and their impact on developmental processes particularly relevant to brain
445 disorders.
446
447 **Chromatin interactivity and insulation are linked to mutant binding properties**
448 Hi-C demonstrated that each mutant has a distinct interaction profile and exhibits (i) variable loss
449 of intra TAD interactions and boundary strength as well as (ii) aggregated loop strength. The IAA
450 condition, in which CTCF is degraded shows a loss of intra TAD interactions and a gain of inter
451 TAD interactions as well as the most reduction in aggregated loop strength (**Figure 6A**), while
452 mutants display their own unique effects. No major compartment changes were observed as
453 shown by the interaction matrices at the chromosome level (**Figure S21A**).

454 Consistent with our finding that cohesin overlap is correlated with CTCF mutant residence time
455 as determined by FRAP analysis (**Figure 3E**), we found that residence time is also associated
456 with chromatin interactivity and insulation score (**Figure 6B, C**).The correlation with the latter was
457 modest and did not reach significance, probably because of the lower Hi-C resolution (10 kb)
458 compared to ChIP-seq. The loop extrusion length (P'max[41]) for each mutant is associated with
459 the fraction that is bound to chromatin (**Figure 6D**), with the longest loops associated with mutants
460 that have the lowest bound fraction, indicating a lower density of CTCF-cohesin anchors in these
461 mutants (**Figure 3B**). The variability in the interaction profile of each mutant can be clearly seen
462 across a 10Mb region of chromosome (**Figure 6E and Figure S21B**).
463
464 To further examine the impact of each mutant on chromatin organization, we first compared the
465 aggregated insulation score at lost sites upon CTCF degradation (comparison of IAA versus WT-
466 ID condition) (**Figure S21C**) using a 2Mb window centered on CTCF peak summits. The insulation
467 scores were generated from Hi-C data at a 10kb resolution. This analysis confirmed that the
468 mutants partially retain insulation at lost sites depending on their binding stability. We then
469 analyzed the insulation score at WT only, common and mutant only binding sites (**Figure 6F and**
470 **Figure S22A**). This analysis revealed that the strong, common accessible binding sites with the
471 highest cohesin overlap, have the highest insulation compared to WT only and mutant only sites.
472 At the WT only sites (comparison of WT-ID versus mutant-ID), some mutants exhibit weak
473 insulation which is not detected when CTCF is degraded (**Figure 6F, Figure S22A** and **Figure**
474 **S22B**), suggesting there might be weak mutant CTCF binding (non-significantly detected by ChIP-
475 seq) at some of these sites. Whether insulation scores at WT only and mutant only sites are lower
476 or higher than each other depends on the residence time and cohesin overlap of each mutant, in
477 line with what we observed in **Figure 3C, E and 6C**. These studies further demonstrate that
478 imaging and molecular analyses can be functionally integrated to provide new insight into mutant
479 specific effects.
480
481 Since WT and mutant only sites are associated with weaker CTCF peaks and lower accessibility,
482 we examined the insulation score at CTCF binding sites, stratified by CTCF and ATAC-seq
483 strength. The profiles in **Figure 6G, 6H, S22C and S22D** indicate that stronger CTCF signals are
484 associated with stronger insulation, while accessibility has a more modest impact. To quantify
485 their independent effects, we performed a mixed multivariate model, the results of which are
486 consistent with the analysis of SMC3 overlap shown in **Figure 4A**.
487

488 **Changes in gene expression are linked to changes in chromatin interactivity**
489 To determine if changes in gene expression are linked to altered chromatin folding, we analyzed
490 interactions involving the promoters of genes that were differentially expressed and found that the
491 promoters of overexpressed genes were enriched in gained chromatin loops (**Figure 7A**).
492 Enrichment was not observed among the under-expressed genes, suggesting that under-
493 expression observed in mESC expressing mutant CTCF proteins might be an indirect effect, while
494 overexpression might reflect a direct CTCF-mediated effect on loop extrusion. This is supported
495 by the sub-analysis of overexpressed and underexpressed genes shown in **Figure S23,** focusing
496 on loops with both anchors overlapping a CTCF binding site, which show a similar result. These
497 loops are more likely to capture the direct effect of CTCF on loop extrusion. This is shown in
498 examples in **Figure 7B** and **Figure S24A** which depict gain of interactions from CTCF binding
499 sites toward the promoters of overexpressed genes, including *Msh6*, *Epcam*, *Foxn2* and *Cox7a2l*
500 (**Figure 7B** and **Figure S24B)**. Interestingly, all these genes are implicated in tumorigenesis[42-48]
501
502 **Most of the change in expression are the result of indirect effects.**
503 Our data show that most CTCF mutations have more impact on gene expression than the IAA
504 condition. Indeed, CTCF degradation has been shown to induce limited gene expression changes
505 [14,15]. To ask whether this could be explained by indirect effects which depend on cell-specific
506 transcriptomes, we estimated the direct versus indirect effects of CTCF mutations on gene
507 expression. Since CTCF can act as a traditional transcription and chromatin regulator [40], we
508 defined indirect CTCF effects as (i) differentially expressed genes (DEGs) with a predicted
509 differentially bound TF at the promoters that do not overlap a CTCF ChIP-seq peak or a predicted
510 differentially bound CTCF site, and (ii) DEGs with differential looping of the promoter without
511 CTCF binding at either loop anchor. Conversely, for direct effects, we required (i) the presence of
512 a CTCF peak and/or a predicted differentially bound CTCF site at the DEG promoter or (ii) loops
513 anchoring the DEG promoter with at least one loop anchor overlapping a CTCF binding site.
514
515 **Figure 7C** shows that indirect effects account for approximately twice as many DEGs compared
516 to the direct effects, except for Q418R. This mutant has a short residence time but a high number
517 of peaks, supporting the hypothesis that unstable peaks might be involved in dynamic and
518 functional looping. Furthermore, a higher percentage of DEGs can be explained by changes in
519 TF or CTCF binding at gene promoters, rather than by changes in chromatin interactivity in all the
520 mutants. For the IAA condition, the contribution of differential loops to expression changes is
521 greater. A similar result was observed when segregating up- and down-regulated genes but with

522 a higher proportion of direct versus indirect effects in the IAA condition for down-regulated genes
523 consistent with loss of loops upon CTCF degradation (**Figure S24B**). These findings suggest that
524 although CTCF mutations induce changes in chromatin looping, the primary mechanism driving
525 alterations in gene expression is loop-independent and mediated by changes in transcription
526 factor binding.
527
528 **DISCUSSION**
529 Using a combination of genetic, imaging, molecular, AlphaFold3 and bioinformatic approaches
530 we have examined the impact of nine, high frequency cancer associated CTCF mutations in ZFs
531 that make contact with the core consensus binding motif. Our studies reveal that the specificity of
532 interaction between the ZFs in CTCFs binding domain and its target DNA sequence is a key
533 determinant of attachment. However, if the same sequence is found in both accessible and
534 inaccessible chromatin, CTCF will preferentially bind accessible sites as highlighted by the fact
535 that mutants bind common accessible sites in preference to less accessible WT only sites. This
536 is underscored by the strong correlation between the fraction of chromatin bound mutants
537 detected by FRAP, and mutant binding at common, accessible CTCF sites. These findings are
538 consistent with studies from Dirk Schubeler's group showing that CTCF can bind both accessible
539 and inaccessible sites, but that the latter are less likely to be involved in insulation [27,49]. We also
540 observed a good correlation between each mutant's residence time and CTCF-cohesin overlap,
541 indicating that binding stability is important for blocking cohesin. A similar, but more modest
542 correlation was found between residence time and chromatin interactivity as well as insulation
543 score, linking binding stability with cohesin overlap as well as loop and boundary formation.
544
545 By dissecting out the contribution of CTCF signal strength and accessibility to SMC3 overlap, we
546 were able to reveal the dominant effect of CTCF. SMC3 overlap at accessible and inaccessible
547 CTCF-bound sites is stronger than overlap at accessible, CTCF-free sites, including enhancers
548 and promoters, likely because TFs have a short residence time compared to CTCF [29]. Our data
549 also demonstrate that strong accessible CTCF bound sites are associated with higher insulation,
550 consistent with higher SMC3 stalling and more stable loops.
551
552 In principle, these features could have been uncovered by analyzing WT CTCF in the context of
553 cohesin overlap and accessibility. However, collectively the graded effects of the mutants act as
554 perturbations that lend support to our models, underscoring how different aspects of CTCFs
555 binding properties contribute. Moreover, they reveal that each mutant's search for binding sites

556 can be decoupled from its chromatin bound residence time, as previously demonstrated for a
557 mutant form of CTCF with ZF8 deleted [50]. For example, the Q418R mutant binds an intermediate
558 number of sites but at these sites it is has a very low residence time and is less competent at
559 blocking cohesin. In contrast, R448Q binds fewer sites with a similar residence time compared to
560 WT CTCF, and it is slightly better at blocking cohesin than WT protein. The low chromatin bound
561 fractions of H455R and R448Q, which have very different chromatin residence times, are
562 correlated to an increased loop length of both mutants (P'-max), which can be explained by the
563 unobstructed passage of cohesin past what would normally be a bound CTCF site.
564
565 The crosstalk between accessibility and CTCF binding goes in both directions, such that
566 accessibility affects binding and binding affects accessibility. Binding at accessible sites, narrows
567 and reduces the ATAC-seq peak. Evidence for this being a CTCF driven effect comes from
568 analysis of the CTCF mutants. Those with low residence times (H455R and Q418R) are unable
569 to function in this capacity, while the R342C mutant has a stronger effect compared to WT CTCF.
570 We speculate that these differences reflect decreased accessibility at the bases of CTCF-SMC3-
571 mediated loops, because in the absence of SMC3, where loops cannot form there is no reduction
572 in accessibility.
573
574 While loss or gain of CTCF binding at promoters can account for direct mutant specific effects on
575 accessibility and gene expression, we could only demonstrate a direct effect of gained CTCF
576 binding associated with gained loops involving the promoters of over-expressed genes. Under-
577 expression could mostly reflect indirect effects of CTCF binding alterations. Indeed, all mutants
578 have their own unique indirect impact in globally changing accessibility at sites where CTCF is
579 not bound. This effect links changes in TF footprinting with altered transcriptional output,
580 explaining most of the mutant specific effects that we observed. We also identified a higher
581 proportion of DEGs mediated by changes in TF or CTCF binding at the gene promoters than by
582 changes in chromatin looping, suggesting that the primary mechanism driving alterations in gene
583 expression might be loop independent. Indeed, the impact of each mutant on TF expression and
584 altered TF binding at the promoters of differentially expressed genes is less likely to give rise to
585 detectable changes in interactivity by Hi-C, consistent with our finding that SMC3 signals at CTCF-
586 free sites are far weaker than at CTCF-bound sites, and thus likely to be involved in weaker, more
587 dynamic loops.
588

589 Aside from the R377C mutant, we observed fewer changes in gene expression in IAA treated
590 cells compared to CTCF mutant expressing cells, which can be explained by the predominant
591 contribution of indirect changes mediated by TF binding on the mutant-specific transcriptome.
592 Although the mutants all uniquely affect gene expression pathways, we detected enrichment in
593 pathways affecting the brain, immune system, cancer and metabolism, which is consistent with
594 the clinical setting in which the mutations are found, namely cancer and brain disorders. The
595 added effect on immune and metabolic pathways provides insight into other changes that could
596 occur in CTCF mutant-mediated diseases.
597
598 CTCF mutations also affect the trajectory of exit from pluripotency. Indeed, while both WT and
599 CTCF mutant cells tend to acquire a neuroectodermal expression profile upon LIF withdrawal, we
600 observed mutant-specific transcriptional signatures of other germ layers, suggesting a more
601 heterogenous differentiation trajectory, which could be confirmed in future work by scRNA-seq.
602 This finding provides a potential mechanism for CTCF mutation-associated neurodevelopmental
603 disorders. Indeed, while CTCF has been involved in neuro-progenitor maintenance and survival[51],
604 our analyses suggest that CTCF mutations might also alter brain development by affecting germ
605 layer commitment.
606
607 Taken together the binding domain mutants we have analyzed here, provide a new appreciation
608 of CTCF's bidirectional relationship with chromatin, its ability to bind and function in different
609 contexts as well as the potential impact of each mutation in clinical settings. Furthermore, our
610 analyses provide a better understanding of how any genetic or epigenetic disorder that alters the
611 landscape of chromatin can in turn impact CTCF binding and function.
612
613 **LIMITATIONS OF THE STUDY**
614
615 In our complementation system, the expression of the transgene is lower than the endogenous
616 CTCF. Since CTCF binding and function is dose-dependent, our analyses of CTCF binding sites
617 might be biased toward higher affinity sites. To distinguish the binding of mutant transgene from
618 residual endogenous WT CTCF in the ID condition, we performed ChIP-seq using a FLAG
619 antibody which resulted in multiple non-specific binding sites. To address this issue, we used a
620 stringent peak calling resulting in a lower number of CTCF binding sites compared to previous
621 reports[15].
622

In this study, we found decreased accessibility at CTCF-SMC3 sites and hypothesize a link with the loop extrusion process. However, despite the evidence from the mutant perturbation, the reduction in overall accessibility at these sites could reflect a preference for CTCF binding involved in insulation at less open chromatin regions, which we cannot entirely rule out.

We found that most of the differences in transcriptomic profiles observed in CTCF mutant cells were the result of indirect and loop-independent effects. However, our Hi-C data has a resolution of 10 kb, which could lead to an underestimation of CTCF loop-dependent expression changes.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jane Skok (Jane.Skok@nyulangone.org).

Materials availability

The WT and mutant CTCF mESC lines are available upon reasonable request.

Data and code availability

All raw and processed sequencing data files will be deposited at NCBI's Gene Expression Omnibus (GEO) upon publication. FRAP analysis was performed using custom Python scripts available at https://github.com/ahansenlab/ZNF143_analysis_code/tree/main and https://zenodo.org/doi/10.5281/zenodo.14056922 [52]. No other custom code/software was used in the study. The publicly available software used is indicated in the Method details.

Acknowledgements

This work was supported by 1R35GM122515 (J.S) and NIH P01CA229086 (J.S). A.T. was supported by the American Cancer Society (RSG-15-189-01-RMC) and St. Baldrick's foundation (581357). ASH additionally acknowledges support from US National Institutes of Health grants R00GM130896, DP2GM140938, R33CA257878 and UM1HG011536, National Science Foundation grant 2036037, the and a Pew-Stewart Cancer Research Scholar grant. GC and GJ were supported by fellowships from the NCC.

The authors thank Skok lab members for helpful scientific discussions, New York University School of Medicine High Performance Computing Facility (HPCF) for computing technical

support, Adriana Heguy and the Genome Technology Center (GTC) core for sequencing efforts, NYU Flow Cytometry and Cell Sorting Center for FACS analysis and sorting. GTC is a shared resource partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center.

Author contributions

These studies were designed by J.A.S., C.D. and M.N. The analysis was performed by C.D and T.S. G.J, P.L and F.F.D.R. established the CTCF mutant cell lines with the guidance of E.P.N. Sequencing libraries were performed by G.J., G.C., R.V., N.K. A.T. reviewed the manuscript. FRAP analysis was performed by C.C.K, D.N.N and A.S.H. The paper was written by J.A.S and C.D.

Declaration of interests

The authors declare no competing interests.

MAIN FIGURE TITLES AND LEGENDS

Figure 1: **CTCF complementation system**. (**A)** Scheme of the CTCF doxycycline-inducible degron system. (**B**) Experimental strategy for the expression of WT and mutant transgenic CTCF. (**C**) Flow cytometry showing the level of GFP (endogenous CTCF) and mRuby (transgenic WT CTCF). (**D**) Left, scheme showing the locations of the different types of CTCF mutations within a ZF. Amino acids making contacts with the DNA are shown in shades of pink, residues that coordinate the zinc ion in red, boundary residues in purple, and residues that contact the sugar phosphate backbone of DNA in blue. Right, representation of CTCF showing the locations of each mutation under investigation. (**E**) Bar graph showing the expression levels (mean fluorescence intensity) of endogenous CTCF (GFP) and transgenic WT or mutant CTCF (mRuby2) after ID treatment. The error bars represent the standard deviation between the two replicates. (**F**) Western blot of endogenous CTCF (CTCF antibody) and transgenic CTCF (FLAG antibody) in untreated (WT-U), IAA, ID conditions. See also **Figures S1-5**.

Figure 2: **CTCF mutations have unique chromatin binding profiles**. (**A)** Scheme showing the locations of the CTCF mutations within the ZF. The consensus CTCF motif highlights the triplet to which each mutant ZF binds (top). The most common motif for the *de novo* binding sites is shown below.. Each bar graph shows the percentage of WT only, common and mutant only CTCF

691 binding sites. The heatmaps show the profile of CTCF and ATAC-seq signal at those sites. The
692 UN condition corresponds to the FLAG control. (**B**) Profiles of ATAC-seq in WT and mutants.
693 Wilcoxon p-values were coded as follow: NS, not significant, * $5x10^{-2}$-$5x10^{-3}$, ** $5x10^{-3}$-$5x10^{-4}$, ***
694 $5x10^{-4}$-$5x10^{-5}$, **** <$5x10^{-5}$. Data were generated on 2 replicates. See also **Figures S6-9.**
695
696 **Figure 3**: **Each mutation uniquely impacts CTCF's chromatin bound fraction, residence**
697 **time and interaction with DNA.** (**A)** Plots of FRAP dynamics for WT and mutant CTCF. The bold
698 lines show the fitted model of the average recovery, and the outlines give the 95% confidence
699 intervals (95%CIs). (**B**) Violin plots of specific bound fractions. (**C**) Violin plots of specific residence
700 times (min). p-values were determined by bootstrapping (n=2500). (**D**) Heatmaps show the
701 proportion of CTCF-cohesin versus CTCF only binding sites. UN corresponds to the FLAG control
702 in untreated cells. (**E**) Correlation between residence time and the percentage of CTCF-cohesin
703 overlap. (**F**) Correlation between the FRAP specific bound fraction relative to WT and the fraction
704 of common CTCF sites relative to all potential binding sites. (**G**) Correlation between the FRAP
705 specific bound fraction relative to WT and the effect of CTCF binding on ATAC-seq signal at
706 CTCF-SMC3 sites (**Figure 2B**). For **E**, **F** and **G,** data were generated in 2 replicates, the p-values
707 were calculated using linear regression, the shaded area corresponds to the 95% CI. See also
708 **Figures S10-14**.
709
710 **Figure 4. Effect of CTCF binding and accessibility on SMC3 overlap. (A)** Bar graph showing
711 the independent effect of CTCF (top) and ATAC-seq (bottom) signal on SMC3 enrichment in
712 WT. The error bars correspond to the 95%CIs. The p-values were calculated using a
713 multivariate logistic model. (**B**) Percentage of SMC3 overlap in WT and CTCF mutants, stratified
714 by ATAC-seq and CTCF signals. **(C)** SMC3 profiles in WT and mutant CTCF. p-values were
715 calculated using Wilcoxon tests. See also **Figures S15-16** and **Table S1**.
716
717 **Figure 5**: **CTCF mutations alter gene expression, cellular reprogramming and TF binding.**
718 (**A)** Heatmap showing supervised clustering of the cell lines based on the expression levels of
719 DEGs identified across the comparisons to WT. (**B**) Gene set enrichment analysis of DEGs in IAA
720 condition and in H455R. The volcano plots below highlight the DEGs belonging to these enriched
721 pathways. (**C**) Radar plots showing the averaged expression of developmental germ layer genes
722 in WT and mutant mESCs cultured in LIF and no LIF conditions. (**D**) Heatmap showing predicted
723 differentially bound TFs in WT, mutants and IAA. CTCF is highlighted with a *. (**E**) Volcano plots
724 highlighting the differentially expressed target genes of CTCF and MBD2 in IAA and CTCF

725 mutants. The metrics for enrichment of the target genes among the DEGs are reported on top on
726 the volcanos (ORs and logistic p-values). (**F**) Examples of altered CTCF binding and footprinting
727 at *Rerg* promoter (left) and altered MYC footprinting at *Brdt* promoter (right). All data in this figure
728 were generated on 2 replicates. See also **Figures S17-20** and **Table S2-4**.
729
730 **Figure 6: CTCF mutations alter chromatin interactivity. (A)** The top panels show the
731 aggregated differential TAD analysis between mutants and WT**. (a)** Reflects the intra-TAD
732 interaction, **(b)** and **(c)** highlight the inter-TAD interactions**.** The lower panels show the aggregated
733 differential peak analysis. The data was generated by Hi-C on 2 replicates. (**B**) Correlation
734 between the interaction counts and the FRAP residence times. (**C**) Correlation between the
735 insulation score at CTCF peaks and the FRAP residence times. (**D**) Correlation between the loop
736 extrusion length and the FRAP bound fractions. (**E**) Example of differential interactions between
737 mutants and WT (right). The left matrix shows the interaction in WT within a 10 Mb region (40 kb
738 resolution) with the insulation score on the side. (**F**) Profiles show the averaged insulation score
739 at WT only, common and mutant only binding sites. (**G**) Profiles show the insulation score at
740 CTCF binding sites stratified by CTCF signals. The bar graph shows the independent effect of
741 CTCF signal on insulation score. (**H**) Profiles show the insulation score at CTCF binding sites
742 stratified by ATAC signals. The bar graph shows the independent effect of ATAC signal on
743 insulation score. For **F, G** and **H,** p-values reported in the profiles were calculated using Kruskal-
744 Wallis tests. The fitted estimates for the bar graphs in (**G**) and (**H**) were obtained using a mixed
745 multivariate model. The error bars correspond to the 95%CI. See also **Figures S21-24**.
746
747 **Figure 7: Changes in gene expression are linked to changes in chromatin interactivity but**
748 **to a lesser extent) than changes in TF binding at gene promoters:** (**A**) Bar graphs showing
749 the enrichment of over-expressed (top) or under-expressed (bottom) genes in gained (top) or lost
750 (bottom) loops in IAA and mutants compared to WT. (**B**) Example of 2 loci (blue and red
751 rectangles) with a direct effect of gain in CTCF binding and chromatin interactivity. The interaction
752 matrices (left) show both gain of intra- and inter-TAD interactions in some mutants compared to
753 WT. The left panels show the zoom-in tracks of these loci. Only significant differential chromatin
754 loops are shown. Overexpressed genes are highlighted in red. **C**) Bar graph showing the
755 percentage of DEGs resulting from direct or indirect effect of CTCF binding distinguishing loop
756 dependent and independent effect.
757
758 **STAR Methods**

759 **Key resources table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
mouse anti FLAG	Sigma	Cat#F1804
mouse anti GAPDH	BioLegend	Cat#W17079A
mouse anti CTCF (WB)	Active Motif	Cat#61311
mouse anti CTCF (Chip-seq)	Cell signaling	Cat#3418
FLAG M2 Magnetic Beads	Sigma	Cat#M8823
mouse anti SMC3	abcam	Cat#ab9263
Mouse anti RAD21	abcam	Cat#ab992
Mouse anti SMC1	ThermoFisher	Cat#MA5-15583
Rabbit IgG	abcam	Cat#ab37415
Critical commercial assays		
Tagment DNA Enzyme and Buffer Kit	Illumina	Cat#20034198
Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus		Cat#20040529
Arima Hi-C kit	Arima	Cat#A510008
Oligonucleotide		
Nextera barcodes	Schmidl *et al* [54]	NA
Deposited data		
RNA-seq	This paper	GEO accession number available upon publication
HiC	This paper	GEO accession number available upon publication
OmniATAC-seq	This paper	GEO accession number available upon publication
SMC3 Chipmentation	This paper	GEO accession number available upon publication
FLAG Chipmentation	This paper	GEO accession number available upon publication
H3K27ac peaks from 120/Ola mESCs	ENCODE	ENCFF519QMV
HOCOMOCO v11	Kulakovskiy *et al.* [74]	https://hocomoco11.autosome.org/
Developmental germ layer gene sets	Hutchins *et al.* [68]	NA
CTCF associated NDD genes	Konrad *et al.* [9]	NA
Developmental genes	DECIPHER	https://www.decipher genomics.org/ddd/overview
Imprinted genes	geneimprint	https://www.geneimprint.com/
ENCODE blacklisted regions	Amemiya *et al.* [61]	https://github.com/Boyle-Lab/Blacklist
KEGG genesets	Kanehisa *et al.* [32]	https://www.genome.jp/kegg/genes.html
CTCF-DNA complex (WT ZF1-7)	PDB	8SSS
CTCF-DNA complex (K365T ZF1-7)	PDB	8SST

CTCF-DNA complex (WT ZF3-11)	PDB	8SSQ
CTCF-DNA complex (K365T ZF3-11)	PDB	8SSR
Experimental models: Cell lines		
Mouse embryonic stem cells E14Tg2a (129/Ola isogenic background) expressing transgene WT or mutant CTCF	This paper	NA
pEN366 - pTRE3G-CTCF-mRuby2-BGHpA-CAGGS-rtta3G-rbgpA-Frt-PGK-EM7-PuroR-bpA-Frt TIGRE donor	Nora EP *et al.* [15]	Addgene #156432
pX330-EN1201	Nora EP *et al.* [15]	Addgene #92144
pEN114 - pCAGGS-Tir1-V5-BpA-Frt-PGK-EM7-PuroR-bpA-Frt-Rosa26	Nora EP *et al.* [15]	Addgene #92143
pASH40-H2B-mRuby-neo (H2B-mRuby)	Hansen *et al*. [24]	NA
Recombinant DNA		
Super piggyBac Transposase expression vector (replaces PB200A-1)	SBI (System Biosciences)	PB210PA-1
Software and algorithms		
FlowJo	BD	https://www.bdbiosciences.com/en-us/products/software/flowjo-v10-software
PyMol 3.1.0	Schrödinger	https://www.pymol.org/
Alphafold3	Abramson *et al* [53]	https://alphafoldserver.com/
Bowtie2	Tarasov et al [59]	https://github.com/BenLangmead/bowtie2
MACS2	Amemiya et al [61]	https://pypi.org/project/MACS2/
trimGalore	Krueger [57]	https://zenodo.org/record/7598955
Sambamba	Tarasov *et al.* [59]	https://lomereiter.github.io/sambamba/
MEME	Bailey *et al.* [21]	https://meme-suite.org/meme/
Deeptools	Ramirez *et al.* [63]	https://deeptools.readthedocs.io/en/develop/
pheatmap	Kolde *et al.* [66]	https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12/topics/pheatmap
fgsea	Sergushichev *et al.* [67]	https://github.com/ctlab/fgsea
STAR aligner	Dobin *et al.* [64]	https://github.com/alexdobin/STAR
Seq-N-Slide pipeline	Dolgalev [56]	https://doi.org/10.5281/zenodo.5550459
HiC-Pro	Servant *et al.* [69]	https://github.com/nservant/HiC-Pro

HiCExplorer	Wolff *et al.* [70]	https://hicexplorer.readthedocs.io/en/latest/
GENOVA	van der Weide *et al.* [44]	https://github.com/robinweide/GENOVA
FitHic	Ay *et al.* [77]	https://github.com/ay-lab/fithic
pairToBed	Quinlan *et al.* [78]	https://bedtools.readthedocs.io/en/latest/content/tools/pairtopair.html
samtools	Danecek *et al.* [62]	https://github.com/samtools/samtools
DESeq2	Love *et al.* [65]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
cooler	Abdennur *et al.* [71]	https://cooler.readthedocs.io/en/latest/
STATA v17	stataCorp	https://www.stata.com/stata17/
R	CRAN	https://www.r-project.org/
TOBIAS	Bentsen *et al.* [35]	https://github.com/loosolab/TOBIAS
UROPA	Kondili *et al.* [75]	https://uropa-manual.readthedocs.io/custom.html
MultiHiCCompare	Stansfield *et al.* [79]	https://github.com/dozmorovlab/multiHiCcompare

760

761 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

762 **Cell lines**

763 Mouse embryonic stem cells E14Tg2a (karyotype 19, XY; 129/Ola isogenic background) and all
764 clones derived from these were cultured under feeder-free conditions in 0.1% gelatin (Sigma ES-
765 006-B) coated dishes (Falcon, 353003) at 37°C and 5% CO_2 in a humidified incubator. The cells
766 were grown in DMEM (Thermo Fisher, 11965-118) supplemented with 15% fetal bovine serum
767 (Thermo Fisher, SH30071.03), 100 U/ml penicillin - 100 µg/ml streptomycin (Sigma, P4458), 1 X
768 GlutaMax supplement (Thermo Fisher, 35050-061), 1 mM sodium pyruvate (Thermo Fisher,
769 11360-070), 1 X MEM non-essential amino-acids (Thermo Fisher, 11140-50), 50 µM b-
770 mercaptoethanol (Sigma, 38171), 10^4 U/ml leukemia inhibitory factor (Millipore, ESG1107), 3 µM
771 CHIR99021 (Sigma, SML1046) and 1 µM MEK inhibitor PD0325901 (Sigma, PZ0162). The cells
772 were passaged every alternate day by dissociation with TrypLE (Thermo Fisher, 12563011). For
773 the exit from pluripotency experiments, mESCs were cultured for 3 days in the same medium but

774 without the addition of LIF (leukemia inhibitory factor, CHIR99021 (Sigma, SML1046)) and MEK
775 inhibitor (PD0325901).
776
777 **METHOD DETAILS**

778 **DNA constructs**

779 Construction of vector for cloning transgenic, doxycycline-inducible expression of WT and mutant
780 mouse *Ctcf* cDNA were obtained from GenScript in pUC19 vectors. The cDNA was amplified such
781 that it harbors an AflII sequence at the 3' end of the gene and fused to a FLAG tag (that harbors
782 NotI sequence) at the 5' end with the help of a fusion PCR. The resultant fragment was digested
783 with NotI and AflII. The *Ctcf* gene was removed from pEN366[15] by digesting with the same
784 enzymes. This backbone was used for insertion of each *Ctcf* mutant.
785
786 In brief, the cDNA region corresponding to each of the C and N terminals and zinc fingers were
787 PCR amplified in such a way that it included a short stretch of the 5′ and/or 3′ region of the
788 neighboring fragment to be connected. The desired PCR products were then annealed, amplified
789 by PCR and cloned into the NotI and AflII sites of the pEN366 backbone (Addgene #156432). All
790 the constructs were verified by DNA sequence analysis. For all transgenes, the final vector
791 harbors an N terminal 3 X FLAG tag and a C terminal *mRuby* as in-frame fusion to WT and mutant
792 *Ctcf*. The vector also harbors a *TetO-3G* element and *rtTA3G* for doxycycline induced expression
793 of the transgene, and homology arms surrounding the sgRNA target site of the *Tigre* locus for
794 locus-specific insertion. The selection of stable integrants was achieved by virtue of *FRT-PGK-*
795 *puro-FRT* cassette. Further details of the vector are described elsewhere [15]. The vector pX330-
796 EN1201 [15] harboring spCas9 nuclease and *Tigre*-targeting sgRNA was used for targeting of *Tigre*
797 locus (Addgene #92144).
798

799 **Gene targeting**

800 Mouse embryonic stem cell E14Tg2a harboring *Ctcf-AID-eGFP* on both alleles and a knock-in of
801 pEN114 - *pCAGGS-Tir1-V5-BpA-Frt-PGK-EM7-PuroR-bpA-Frt-Rosa26* at *Rosa26* locus was
802 used as the parental cell line for making all the transgenes [15]. pEN366 derived vectors harboring
803 the rescue transgenes (WT and mutant *Ctcf*) were used for targeting transgenes to the *Tigre* locus
804 (clone ID# EN156.3.5) [14]. For nucleofections, 15 µg each of plasmids harboring the transgenes
805 and 2.5 µg of those with sgRNA targeting the *Tigre* locus was used. Nucleofection were performed
806 using Amaxa P3 Primary Cell kit (Lonza, V4XP-3024) and 4D-transfector. 2 million cells were
807 transfected with program CG-104 in each case. The cells were recovered for 48 h with no

808 antibiotic followed by selection in puromycin (1 µg/mL) (Thermo Fisher, A1113803). Single
809 colonies were manually picked and expanded in 96 well plates. Clones were genotyped by PCR
810 and flow cytometry was performed to confirm that the level of expression of transgenes were
811 comparable. All the clones that were used for the analyses were homozygous for the integration
812 of the transgenes and their levels of expression were comparable.
813
814 **Induction of auxin inducible degradation of CTCF and doxycycline induced expression**
815 For degradation of endogenous CTCF, the auxin-inducible degron was induced by adding 500
816 µM indole-3-acetic acid (IAA, chemical analog of auxin) (Sigma, I5148) to the media. Expression
817 of transgenes was achieved by the addition of doxycycline (Dox, 1 µg/ml) (Sigma, D9891) to the
818 media. The cells were treated with IAA and/or Dox for 2 days.
819
820 **Western Blotting**
821 mESCs were dissociated using TrypLE, washed in PBS, pelleted and used for western blotting.
822 Approximately 2 million cells were used to prepare cell extract. Cell pellets were resuspended in
823 RIPA lysis buffer (Thermo Fisher, 89900) with 1X HALT protease inhibitors (Thermo Fisher,
824 78430), incubated on ice for 30 min, spun at 4°C at 13,000 rpm for 10 min and supernatant was
825 collected. For the western blot of CTCF, low salt lysis buffer (0.1 M NaCl, 25 mM HEPES, 1 mM
826 $MgCl_2$, 0.2 mM EDTA and 0.5% NP40) was used supplemented with 125 U/ml of benzonase
827 (Sigma E1014). Protein concentration was measured using the Pierce BCA assay kit (Thermo
828 Fisher, 23225). 20 µg of protein were mixed with Laemmli buffer (Biorad, 1610737) and β-
829 mercaptoethanol, heated at 95°C for 10 min and run on a Mini-protean TGX 4%-20%
830 polyacrylamide gel (Biorad, 456-1095). The proteins were transferred onto PVDF membranes
831 using the Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, 170-3930) at 80 V, 120 mA for
832 90 min. PVDF membranes were blocked with 5% BSA in 1 X TBST prior to the addition of
833 antibody. The membranes were probed with appropriate antibodies overnight at 4°C, mouse anti-
834 FLAG antibody (Sigma, F1804; 1: 1,000 dilution), GAPDH (BioLegend, W17079A; 1:5,000
835 dilution), CTCF (active motif, 61311; 1: 2,000 dilution), SMC1 (ThermoFischer, MA5-15583.
836 1:1000 dilution), RAD21 (abcam, ab992,1:2000 dilution). Membranes were washed five times in
837 PBST (1 × PBS and 0.1% Tween 20) for 5 min each and incubated with respective secondary
838 antibodies in 5% BSA at room temperature for 1 h. The blots were rinsed in PBST and developed
839 using enhanced chemiluminescence (ECL) and imaged by Odyssey LiCor Imager (Kindle
840 Biosciences).
841

842 **Flow cytometric analysis**

843 Cells were dissociated with TrypLE, washed and resuspended in MACS buffer for flow cytometric
844 analysis on LSRII UV (BD Biosciences). Analysis was performed using the FlowJo software.

845

846 **Cell culture – FRAP**

847 Control cells expressing H2B-mRuby were engineered using the same background as the CTCF
848 transgene expressing cells, E14Tg2a mESCs. Briefly, pEN114 - *pCAGGS-Tir1-V5-BpA-Frt-PGK-*
849 *EM7-PuroR-bpA-Frt-Rosa26* was used as the parental cell line to be nucleofected by the
850 pASH40-H2B-mRuby-neo with the PiggyBac transposase vector (SBI, PB210PA-1). Halo-H2B
851 plasmid was previously published in Hansen *et al,* [24] and we replaced the Halotag with mRuby.
852 For the FRAP experiment, CTCF and H2B expressing cells were cultured for two days on 35 mm
853 no 1.5H glass-bottom imaging dishes (MatTek, Ashland, MA, P35G-1.5-14C) coated with Geltrex
854 (Gibco, A1413201) according to manufacturer's instructions. 48 hours prior to the experiment, the
855 culture medium was changed to medium supplemented with 500 μM 3-Indoleacetic acid (IAA;
856 Sigma-Aldrich, I2886-5G) and 1 μg/mL doxycycline (Sigma-Aldrich, D3072-1ML). Just prior to
857 imaging, the medium was changed to imaging medium: phenol red free DMEM with all other
858 aspects of the medium the same.

859

860 **Fluorescence recovery after photobleaching (FRAP)**

861 FRAP experiments were performed on an LSM900 confocal microscope (Zeiss, Germany)
862 equipped with a full, humidified incubation chamber that was maintained at 37°C with 5.5% CO_2.
863 The time series were acquired in confocal mode using two-stage acquisition with the following
864 excitation parameters: 561 nm excitation laser at 0.8% power, 53 μm pinhole size corresponding
865 to 1 AU, scan speed 7, and 3x crop factor corresponding to a pixel dwell time of 3.06 μs. During
866 the first phase, we imaged 25 frames with 0.5 seconds between frames, and during the second
867 phase we imaged 160 frames with 4 seconds between frames, resulting in a total movie length of
868 657 seconds. For bleaching, a 1μm diameter ROI was selected. Bleaching was performed after
869 frame 8 using both the 488nm and 561nm lasers at 100% power, 53μm pinhole size, and scan
870 speed 7 corresponding to a pixel dwell time of 3.06μs. For each condition, 30 movies were
871 acquired and between 3 and 11 were excluded due to drift.

872

873 **ChIPmentation**

874 mESCs were dissociated using TrypLE (Thermofisher #12605010) and washed once in 1X PBS.
875 After counting, cells were divided in 10 million aliquots and resuspended in fresh 1X PBS (1million
876 cells/1ml). For double cross linking 25mM EGS (ethylene glycol bis(succinimidyl succinate);

877 Thermofisher #21565) were added and cells were put in rotation for 30 min at room temperature,
878 followed by addition of 1% formaldehyde (Tousimis #1008A) for 10 min also in rotation at room
879 temperature. Quenching was performed by adding glycine to a final concentration of 0.125 M
880 followed by incubations of 5 min at room temperature in rotation. Fixed cells were washed twice
881 with 5 ml of 1 X PBS containing 0.5% BSA and centrifuged at 3000 rpm for 5 min at 4°C. Pellets
882 were finally resuspended in 500 µl 1X PBS containing 0.5% BSA, transferred to 1.5 ml Eppendorf
883 and centrifuged at 3000rpm for 3 min at 4°C. Supernatant was completely removed, pellets were
884 snap-frozen in liquid nitrogen and stored at -80°C. Fixed cells (10 million) were thawed on ice,
885 resuspended in 350 µl ice cold lysis buffer (10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA
886 (pH 8.0), 0.5 mM EGTA (pH 8.0), 0.1% sodium deoxycholate, 0.5% N-lauroysarcosine and 1X
887 protease inhibitors) and lysed for 10 min by rotating at 4° C. Chromatin was sheared using a
888 bioruptor (Diagenode) for 15 minutes (30 sec on, 30 sec off, high output level). 100 µl of cold
889 lysis buffer and 50 µl of 10% Triton X-100 (final concentration of 1) were then added and the
890 samples were centrifuged for 5 min at full speed at 4°C. Supernatant was collected, transferred
891 to a new tube (Protein Low Binding tube) and shearing was continued for another 10 min, then
892 the chromatin was quantified. FLAG M2 Magnetic Beads (Sigma, M8823) were used for FLAG
893 immunoprecipitation. In other cases, SMC3 (abcam, ab9263), CTCF (cell Signaling Technology,
894 3418) and Rabbit IgG (abcam, ab37415) antibodies were bound to protein A magnetic beads by
895 incubation on a rotator for one hour at room temperature. 10 µl each of antibody was bound to 50
896 µl of protein-A magnetic beads (Dynabeads) and added to the sonicated chromatin for
897 immunoprecipitation at 4°C overnight. Next day, samples were washed and tagmentation were
898 performed as per the original ChIPmentation protocol [54]. In short, the beads were washed
899 successively twice in 500 µl cold low-salt wash buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2
900 mM EDTA (pH 8.0), 0.1% SDS, 1% tritonX-100), twice in 500 µl cold LiCl-containing wash buffer
901 (10 mM Tris-HCl (pH 8.0), 250 mM LiCl, 1 mM EDTA (pH 8.0), 1% triton X-100, 0.7% sodium
902 deoxycholate) and twice in 500 µl cold 10 mM cold Tris-Cl (pH 8.0) to remove detergent, salts
903 and EDTA. Subsequently, the beads were resuspended in 25 µl of the freshly prepared
904 tagmentation reaction buffer (10 mM Tris-HCl (pH 8.0), 5 mM MgCl2, 10% dimethylformamide)
905 and 1 µl Tagment DNA Enzyme from the Tagment DNA Enzyme and Buffer Kit (Illumina
906 #20034198) and incubated at 37°C for 10 min in a thermomixer. Following tagmentation, the
907 beads were washed successively twice in 500 µl cold low-salt wash buffer (20 mM Tris-HCl (pH
908 7.5), 150 mM NaCl, 2 mM EDTA (pH8.0), 0.1% SDS, 1% triton X-100) and twice in 500 µl cold
909 Tris-EDTA-Tween buffer (0.2% tween, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)). Chromatin
910 was eluted and de-crosslinked by adding 70 µl of freshly prepared elution buffer (0.5% SDS, 300

911 mM NaCl, 5 mM EDTA (pH 8.0), 10 mM Tris-HCl (pH 8.0) and 10 ug/ml proteinase K for 1 hour
912 at 55°C 850rpm and overnight at 65°C 850rpm. Next day, the supernatant was collected,
913 transferred to new DNA Low Binding tubes and supplemented with an additional 30 µl of elution
914 buffer. DNA was purified using MinElute Reaction Cleanup Kit (Qiagen #28204) and eluted in 20
915 ul. Purified DNA (20 µl) was amplified as per the ChIPmentation protocol (Schmidl, Rendeiro et
916 al. 2015) using indexed and non-indexed primers and NEBNext High-Fidelity 2X PCR Master Mix
917 (NEB M0541) in a thermomixer with the following program: 72°C for 5 m; 98°C for 30 s; 14 cycles
918 of 98°C for 10 s, 63°C for 30 s, 72°C for 30 s and a final elongation at 72°C for 1 m. DNA was
919 purified using Agencourt AMPure XP beads (Beckman, A63881) to remove fragments larger than
920 700 bp as well as the primer dimers. Library quality and quantity were estimatred using
921 Tapestation (Agilent High Sensitivity D1000 ScreenTape #5067-5584 and High Sensitivity D1000
922 reagents #5067-5585) and quantified by Qubit (Life Technologies Qubit™ 1X dsDNA High
923 Sensitivity (HS) #Q33230). Libraries were then sequenced with the Novaseq6000 Illumina
924 technology according to the standard protocols and with around 200 million 150bp paired-end
925 total per sample.
926
927 **Total RNA-seq**
928 mESCs were dissociated using TrypLE, washed in 1X PBS, pelletted and 2.5 million cells were
929 used for extracting RNA with RNeasy plus kit (Qiagen #74134) in each case. RNA quality was
930 checked in Tapestation (Agilent High Sensitivity RNA ScreenTape #5067-5579, High Sensitivity
931 RNA Sample Buffer #5067-5580 and High Sensitivity RNA Ladder #5067-5581) and quantified
932 by Nanodrop. 500 ng were used for Total RNA libraries preparation using Illumina kit (Illumina
933 Stranded Total RNA Prep, Ligation with Ribo-Zero Plus #20040529). Library concentrations were
934 estimated using Tapestation (Agilent High Sensitivity D1000 ScreenTape #5067-5584 and High
935 Sensitivity D1000 reagents #5067-5585) and quantified by Qubit (Life Technologies Qubit™ 1X
936 dsDNA High Sensitivity (HS) #Q33230). Libraries were then sequenced with the Novaseq6000
937 Illumina technology according to the standard protocols and with around 100 million 150bp paired-
938 end total per sample.
939
940 **Hi-C**
941 Hi-C was performed in duplicates using around 1 million cells each. mESCs were dissociated
942 using TrypLE (Thermofishe #12605010), washed once in 1X PBS and resuspended in fresh 1X
943 PBS (1 million cells/ 1 ml). For double cross linking 25 mM EGS (ethylene glycol bis(succinimidyl
944 succinate); Thermofisher #21565) were added and cells were put in rotation for 30 min at room

945 temperature, followed by addition of 1% formaldehyde (Tousimis #1008A) for 10 min also in
946 rotation at room temperature. Quenching was performed by adding glycine to a final concentration
947 of 0.125M followed by incubations of 5 min at room temperature in rotation. Fixed cells were
948 washed twice with 5 ml of 1X PBS containing 0.5% BSA and centrifuged at 3000 rpm for 5 min at
949 4°C. Pellets were finally resuspended in 500 µl 1X PBS containing 0.5% BSA, transferred to 1.5ml
950 Eppendorf and centrifuged at 3000 rpm for 3 min at 4°C. Supernatant was completely removed,
951 pellets were snap-frozen in liquid nitrogen and stored at -80°C. Samples were subsequently
952 processed using the Arima Hi-C kit (A510008) as per the manufacturer's protocol and sequenced
953 with the Novaseq6000 Illumina technology according to the standard protocols and with around
954 600 million 150bp paired-end reads per sample.
955
956 **ATAC-seq**
957 We used an improved ATAC-seq protocol (Omni-ATAC) adopted from Corces et al., [22] with small
958 adaptations. Briefly, cells in culture were treated with 200 U/ml DNase (Worthington # LS002007)
959 for 30 min at 37°C to remove free-floating DNA and any DNA from dead cells. Cells were then
960 harvested via trypsinization and resuspended in regular medium. After counting, 500,000 cells
961 were collected, resuspended in 1X cold PBS and spin down at 500 g for 5 min at 4°C in a fixed-
962 angle centrifuge. After centrifugation, the supernatant was removed and the pellet resuspended
963 in 500 µl of cold ATAC-seq resuspension buffer 1 (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM
964 $MgCl_2$, 0.1% NP-40 and 0.1% Tween-20 in water). 50 µl (50.000 cells) were transferred to a new
965 1.5 ml Eppendorf tube and 0.5 µl l 1% Digitonin (Promega #G9441) was added by pipetting well
966 few times. The cell lysis reaction was incubated on ice for 3 min. After lysis, 1 ml of cold ATAC-
967 seq resuspension buffer 2 (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% Tween-
968 20 in water) was added. Tube was inverted three times to mix and nuclei were pelleted by
969 centrifugation for 10 min at 500 g at 4°C. Supernatant was very carefully removed (pellet could
970 be almost invisible at this step) and nuclei resuspended in 45 µl of transposition mix (25 µl 2X TD
971 buffer (Illumina #20034198), 16.5 µl 1X PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20 and 2.5
972 µl water) by pipetting up and down few times. 5 µl of transposase enzyme (Illumina #20034198)
973 was then added and the transposition reaction was incubated at 37°C for 30 min in a thermomixer
974 with shaking at 1000 rpm. Reaction was cleaned up with Zymo DNA Clean and Concentrator-5
975 kit (Zymo#11-302C). Transposed DNA fragments were then amplified as described previously in
976 Buenrostro et al., [55]. Final libraries were purified with with Zymo DNA Clean and Concentrator-5
977 kit (Zymo#11-302C; Note: better to use two separate kits for pre- and post-amplification clean up),
978 checked by Tapestation (Agilent High Sensitivity D1000 ScreenTape #5067-5584 and High

979 Sensitivity D1000 reagents #5067-5585) and quantified by Qubit (Life Technologies Qubit™ 1X
980 dsDNA High Sensitivity (HS) #Q33230). Libraries were then sequenced with the Novaseq6000
981 Illumina technology according to the standard protocols and with around 50 million 150bp paired-
982 end reads per sample.
983
984 **QUANTIFICATION AND STATISTICAL ANALYSIS**
985
986 **Fluorescence recovery after photobleaching (FRAP) analysis**
987 FRAP analysis was performed using custom Python scripts. Briefly, movies were loaded into a
988 custom Python GUI (https://github.com/ahansenlab/FRAP-drift-correction-GUI). The bleached
989 nucleus was segmented using manually selected segmentation parameters and was used to
990 estimate the average nuclear intensity and the average background intensity. To correct for cell
991 drift, the ROI positions were manually updated at a number of frames and the ROI position was
992 linearly interpolated between manual updates. Cells with excessive drift were manually excluded
993 and the mean fluorescence intensity of the bleach was taken as the average intensity within the
994 ROI.
995 To correct for photobleaching, $I_{nonbleach}(t)$ was smoothed using an averaging filter, and
996 a series of correction factors ($C(t)$) were computed as the ratio of these values to their average
997 prebleach value as follows:
998
999 $$C(t) = \frac{I_{nonbleach,smooth}(t) - I_{background}(t)}{\langle I_{nonbleach}(t) - I_{background}(t)\rangle_{prebleach}}$$
1000
1001 Subsequently, the FRAP curves were normalized according to the following equation:
1002
1003
1004 $$\underline{I_{bleach}(t)} = \frac{C(t)(I_{bleach}(t) - I_{background}(t))}{\langle C(t)(I_{bleach}(t) - I_{background}(t))\rangle_{prebleach}}$$
1005
1006 where $\underline{I_{bleach}(t)}$ is the normalized, background-subtracted, and photobleaching corrected FRAP
1007 signal.
1008 Finally, since previous measurements of CTCF's free diffusion coefficient have indicated
1009 a reaction-dominant model is most appropriate, we fit the following, previously demonstrated
1010 model (Sprague, Pego et al. 2004, Hansen, Pustova et al. 2017).
1011
1012 $$FRAP(t) = 1 - C_{eq,fast}e^{-k_{off,fast}t} - C_{eq,slow}e^{-k_{off,slow}t}$$
1013

where $k_{off,slow} < k_{off,fast}$. $C_{eq,slow}$ was taken as the specific bound fraction and the specific residence time was taken as $\frac{1}{k_{off,slow}}$. To estimate parameter distributions, we implemented a bootstrapping routine in which a set of 16 movies were randomly resampled from the full dataset 2500 times with replacement. For each sample, the above equation was fit and the parameter values were taken as one data point. The 2500 resampled points were then used as distributions for plotting the specific bound fraction and residence time. For ease of comparison, the specific bound fractions were normalized to the average WT value when plotting.

ChIP-seq data processing and quality control

Data were processed using Seq-N-Slide pipeline [56]. Briefly, after trimming for NEXTERA adaptor sequence using trimGalore [57]. The reads were aligned to mm10 genome with Bowtie2 [58]. Ambiguous reads were filtered to use uniquely mapped reads in the downstream analysis. PCR duplicates were removed using Sambamba [59]). Narrow peaks were called using MACS2 [60] in pair-end mode and with IgG as control for SMC3 and CTCF ChIP-seq or WT untreated condition (in which FLAG CTCF is not expressed) for CTCF FLAG to control for unspecific binding, Peaks overlapping ENCODE blacklisted regions were filtered out [61]. These procedures allowed us to minimize the false positive calls, as reflected by the high percentage of called peaks containing CTCF consensus motifs as detected by MEME [21] with 63% for WT. To ensure comparability between conditions, FLAG and SMC3 chip-seq samples were down-sampled to the same number of aligned (100 millions reads per replicates), except for 2 samples R377H-rep1 and R339Q-rep which were under-sequenced (with 68 and 90 millions reads, respectively) and good quality reads before peak calling using samtools [62]. Bigwigs were obtained for visualization on individual as well as merged bam files using Deeptools [63]. For CTCF FLAG visualization, the differential signals were generated after subtracting non-specific signal present in the WT untreated condition. Heatmaps and average profiles were performed on merged bigwig files using Deeptools.

RNA-seq data processing and quality control

Data were processed using Seq-N-Slide pipeline [56]. Briefly, after trimming for the Illumina adaptor and the T overhang as per Illumina recommendation using TrimGalore [57], reads were aligned against the mouse reference genome (mm10) using the STAR [64] aligner and differentially expressed genes were called using DESeq2 [65] with an adjusted p-value of 0.05 and a fold change cutoff of 1.5. Heatmap was performed using pheatmap R package [66] and GSEA using fgsea R package [67] and KEGG genesets [32]. Developmental germ layer gene sets were downloaded from

1047 Hutchins *et a*l.,[68]. Radarplots were generated using the R package fsmb after averaging the
1048 expression level of the genes of each germ layers and of the 2 replicates. CTCF-associated
1049 neurodevelopmental disorder (NDD) genes were retrieved from Konrad *et al.* [9], while genes
1050 associated with developmental disorders were downloaded from DECIPHER
1051 (https://www.deciphergenomics.org/ddd) and imprinted genes from geneimprint
1052 (https://www.geneimprint.com/). Enrichment analyses for those sets were performed using
1053 logistic regressions. To compare the change in gene expression during the exit from pluripotency
1054 between WT and mutant (longitudinal analysis), linear regressions on the difference of log2
1055 between LIF versus no LIF condition were performed.
1056
1057 **ATAC-seq data processing and quality control**
1058 Data were processed using Seq-N-Slide pipeline [56]. Briefly, Reads were aligned to mm10 genome
1059 with Bowtie2 [58]. Ambiguous reads were filtered to use uniquely mapped reads in the downstream
1060 analysis. PCR duplicates were removed using Sambamba [59]. Mitochondrial contamination was
1061 assessed using samtools [62]. All the samples showed less than 0.6% of mitochondrial DNA, Peaks
1062 were called using MACS2 [60]and peaks overlapping ENCODE blacklisted regions were filtered out
1063 [61]. Bigwigs were obtained for visualization on individual as well as merged bam files using
1064 Deeptools [63]. Heatmaps and average profiles were performed on merged bigwig files using
1065 Deeptools.
1066
1067 **Hi-C Processing and Quality Control.**
1068 HiC-Pro [69] was used to align and filter the Hi-C data. To generate Hi-C filtered contact matrices,
1069 the Hi-C reads were aligned against the mouse reference genome (mm10) by bowtie2 (version
1070 2.3.1) using local mode for the second step of HiC-Pro alignment (--very-sensitive -L 20 --score-
1071 min G,20,8 --local) to ignore the molecular barcode UMI sequences and dark bases introduced
1072 by Arima library prep kit at the 5' end of the read. Singleton, ambiguous and duplicated reads
1073 were filtered out. Ligation sites for Arima HiC kit were set up using
1074 GATCGATC,GANTGATC,GANTANTC,GATCANTC. Low mapping quality (MAPQ<30), self
1075 circle, dangling ends and re-ligation reads were filtered out through the HiC-pro pipeline. Valid
1076 pairs represented more than 80% for all the samples. Samples were downscaled to the same
1077 number of good quality uniquely aligned reads using HiCExplorer [70]. Of note, we did not use the
1078 number of valid pairs to downscale the samples to not overcorrect some conditions such as IAA
1079 and mutant H455R which were expected to show a dramatic biological decrease in valid
1080 interactions.

1081

1082 Downscaled interaction matrices were generated separately for each replicate and merged for
1083 visualization and downprocessing as mcool files using cooler [71]. Samples were balanced using
1084 the ICE correction method [72] with HiCExplorer [70]. HiC maps were generated from the merged
1085 downscaled and corrected matrices using GENOVA and HiCExplorer [44,70]. RCP (relative contact
1086 probabilities) and insulation score were generated using GENOVA [44]. After log-log transformation
1087 and Loess smoothing using R, the first derivative (P'max) was estimated as described in Gassler
1088 et al., [41].

1089

1090 **CTCF peak motif analysis**

1091 Motifs were called using MEME [21] with a 1st order background model on CTCF peaks called using
1092 MACS2 after separating the peaks into 3 groups for each pairwise comparison between WT and
1093 mutant CTCF: WT only, common and mutant only peaks. Peak intersection sets were generated
1094 using the intervene package [73]and motifs were searched within 100 bp of each peak summit.

1095

1096 **AlphaFold3 prediction of the DNA-CTCF complexes in the mutants**

1097 As a quality control, to compare the prediction to experimental data, we downloaded the structure
1098 deposited by Yang *et al*. [13] from PDB for WT CTCF ZF1-ZF7 (accession number PDB 8SSS, with
1099 oligo CCAGCAGGGGGCGCTAGTGAG and CTCACTAGCGCCCCCTGCTGG) , K365T CTCF
1100 ZF1-ZF7 (accession number PDB 8SST with the same oligos as WT), WT CTCF ZF3-ZF11
1101 (accession number PDB 8SSQ with oligo
1102 GTGCAGTACCACATTTAACCAGCAGGGGGCGCTAA and
1103 TTAGCGCCCCCTGCTGGTTAAATGTGGTACTGCAC), K365T CTCF ZF3-ZF11 (accession
1104 number PDB 8SSR with oligo with the same oligos as WT). For the Alphafold3 predictions we
1105 used the same DNA sequences and the CTCF ZF1-11 positions obtained from Yang *et al* [13].
1106 AlphaFold3 predictions [53] were also carried out for WT and K365T CTCF ZF3-ZF11 and CTCF
1107 ZF1-7 using the same DNA and CTCF sequence to compare the prediction to the experimental
1108 data. We, then, predicted the CTCF (ZF1-11)-DNA complex for WT and each mutant CTCF using
1109 the concatenated DNA sequences from Yang *et al* [13]
1110 (GTGCAGTACCACATTTAACCAGCAGGGGGCGCTAGTGAG and
1111 CTCACTAGCGCCCCCTGCTGGTTAAATGTGGTACTGCAC). The predicted protein-DNA
1112 structures were visualized in PyMOL 3.1.0. The hydrogen bonds between DNA and CTCF were
1113 identified using PyMOL dist function with a cutoff of 5 Å. The root mean square deviation (RMSD)
1114 between the structures was calculated using PyMOL 3.1.0.

1115

1116 **Stratified analysis at enhancers and promoters**

1117 H3K27ac peaks from 120/Ola mESCs were downloaded from ENCODE (ENCFF519QMV).
1118 CTCF, cohesin and ATAC-seq peaks were classified as active promoter sites when they overlap
1119 with an H3K27ac peak and are located in a gene promoter region, defined as 2kb upstream and
1120 1 kb downstream gene TSSs. Active enhancer sites when they overlap with an H3K27ac peak
1121 and are located outside gene promoter regions.

1122

1123 **Multivariate CTCF/SMC3 overlap and insulation score analysis**

1124 To test the independent of CTCF and Accessibility on SMC3 overlap, CTCF ChIP-seq and ATAC-
1125 seq signal strengths were categorized into weak, intermediate and strong signal using the inter-
1126 quartile values of the signal across WT and mutant samples. Weak binding or accessibility was
1127 defined as sites with signals inferior to the 1st quartile, intermediate binding or accessibility as
1128 sites with signals between the 2nd and 3rd quartiles and strong binding or accessibility as sites with
1129 signals superior to the 3rd quartile. For SMC3 overlap, a multivariate logistic model testing the
1130 effect of CTCF strength, ATAC strength and their interaction was performed in WT mESCs after
1131 averaging the replicates. The log2 odd ratios were computed to estimate the enrichment in SMC3.
1132 For insulation score a mixed multivariate model testing the effect of CTCF strength, ATAC
1133 strength and their interaction was performed across sample. For both models the marginal effects
1134 and 95th confidence intervals were computed. These analyses were performed using STATA v17.

1135

1136 **ATAC-seq footprinting analysis**

1137 Footprinting analysis was performed on ATAC-seq signal using TOBIAS [35]. Briefly, ATAC-seq
1138 signals were first corrected for Tn5 cutting sequence bias using the ATACorrect function and
1139 footprinting score calculated using FootprintScores at peaks detected by MACS. Finally
1140 differential footprinting estimated pariwise between WT and mutant CTCF using BINDetect and
1141 HOCOMOCO v11 core motif probability weight matrices [74]. Peaks were annotated for gene
1142 promoter using UROPA [75] and GENCODE version M25 [76]. Gene promoters were defined as
1143 regions within 2kb upstream and 1 kb downstream the gene transcription starting site.

1144

1145 **Loop calling and annotation**

1146 Significant loops (q-value<0.05) were called on the downscaled and corrected interaction
1147 matrices using FitHIC at 10kb resolution [77] and annotated for CTCF peaks and gene TSS using
1148 pairToBed [78].

1149

1150 **Differential aggregated peak and TAD analyses**

1151 Aggregated peak and TAD analysis on downscaled and corrected interaction matrices was
1152 performed using GENOVA at 10kb resolution [44]. For aggregated peaks analyses, the union of
1153 loops called by FitHIC [77] in WT and mutant CTCF for each pairwise comparison was used to
1154 generate and compare aggregated interaction at loop anchors. For aggregated TAD analyses,
1155 the insulation score was first calculated to identify TAD and TAD boundaries. The union of TADs
1156 called by GENOVA in WT and mutant CTCF was used to generate and compare aggregated intra
1157 and inter-TAD interactions.

1158

1159 **Differential loop analysis**

1160 Differential loops (adjusted p-value<0.05) were called for each pairwise comparison between WT
1161 and mutant CTCF using MultiHiCCompare which performed a cross-sample normalization before
1162 testing for differential loops using an exact test [79]. Enrichment for differentially expressed genes
1163 among differential loops was tested using logistic regression after annotating loop anchors with
1164 gene TSS.

1165

1166 **Direct vs indirect effect analysis**

1167 Direct effect was defined as differential TF binding with a predicted differential CTCF binding
1168 identified using TOBIAS [35] or ChIP-seq peak at the gene promoter and/or differential loops
1169 identified using MultiHiCCompare [79] with a CTCF peak at either anchor. Gene promoters were
1170 defined as regions spanning 2 kb upstream and 1 kb downstream gene TSSs. Overlap between
1171 gene promoters and differential loop were performed using pairToBed [78].

1172

1173 **SUPPLEMENTAL INFORMATION:**

1174

1175 **Excel table title and legends**

1176

1177 **Table S1:** Number and percentage of SMC3 peaks overlapping a CTCF or ATAC-seq peak,
1178 stratified by CTCF and ATAC signal strength. This table is related to **Figure 4B.**

1179

1180 **Table S2:** Differentially expressed genes (DEGs) identified using DE-seq. DEGs were defined as
1181 adjusted genes with p-value<0.05 and absolute log2 fold change $\geq$ log2(1.5) compared to WT.
1182 This table is related to **Figure 5A**.

1183

1184 **Table S3: CTCF mutations altered cell differentiation. Tab A,** Expression of germ layer genes
1185 (TPM) in LIF versus no LIF condition in WT and the mutant. **Tab B,** Differentially expressed genes
1186 during in LIF condition in mutant compared to WT. **Tab C,** Differentially expressed genes during
1187 in no LIF condition in mutant compared to WT. **Tab D,** Differentially expressed genes during the
1188 exit from pluripotency (LIF to no LIF) in mutant compared to WT. This table is related to **Figure**
1189 **5C**.

1190

1191 **Table S4: CTCF mutations altered TF binding. Tab A,** Differentially bound transcription factors
1192 predicted by footprinting analyses. This tab is related to **Figure 5D**. **Tab B,** Enrichment of the
1193 target genes of differentially bound TFs among the DEGs. This tab is related to **Figure 5E**. **Tab**
1194 **C,** List of the target genes of differentially bound TFs with differential expression. This tab is
1195 related to **Figure 5E**.

1196 **References:**

1197

1198 1. Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.M. (2019). DNA
1199 loop extrusion by human cohesin. Science *366*, 1338-1345. 10.1126/science.aaz3418.

1200 2. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A.
1201 (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep *15*, 2038-2049.
1202 10.1016/j.celrep.2016.04.085.

1203 3. Hansen, A.S. (2020). CTCF as a boundary factor for cohesin-mediated loop extrusion:
1204 evidence for a multi-step mechanism. Nucleus *11*, 132-148.
1205 10.1080/19491034.2020.1782024.

1206 4. Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal
1207 Communication. Cell *164*, 1110-1121. 10.1016/j.cell.2016.02.007.

1208 5. Filippova, G.N., Lindblom, A., Meincke, L.J., Klenova, E.M., Neiman, P.E., Collins, S.J.,
1209 Doggett, N.A., and Lobanenkov, V.V. (1998). A widely expressed transcription factor with
1210 multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1
1211 within one of the smallest regions of overlap for common deletions in breast and prostate
1212 cancers. Genes Chromosomes Cancer *22*, 26-36.

1213 6. Xiang, Y., Zhou, X., Hewitt, S.L., Skok, J.A., and Garrard, W.T. (2011). A multifunctional
1214 element in the mouse Igkappa locus that specifies repertoire and Ig loci subnuclear
1215 location. J Immunol *186*, 5356-5366. 10.4049/jimmunol.1003794.

1216 7. Debaugny, R.E., and Skok, J.A. (2020). CTCF and CTCFL in cancer. Curr Opin Genet
1217 Dev *61*, 44-52. 10.1016/j.gde.2020.02.021.

1218 8. Gregor, A., Oti, M., Kouwenhoven, E.N., Hoyer, J., Sticht, H., Ekici, A.B., Kjaergaard, S.,
1219 Rauch, A., Stunnenberg, H.G., Uebe, S., et al. (2013). De novo mutations in the genome
1220 organizer CTCF cause intellectual disability. Am J Hum Genet *93*, 124-131.
1221 10.1016/j.ajhg.2013.05.007.

1222 9. Konrad, E.D.H., Nardini, N., Caliebe, A., Nagel, I., Young, D., Horvath, G., Santoro, S.L.,
1223 Shuss, C., Ziegler, A., Bonneau, D., et al. (2019). CTCF variants in 39 individuals with a

1224 variable neurodevelopmental disorder broaden the mutational and clinical spectrum.
1225 Genet Med *21*, 2723-2733. 10.1038/s41436-019-0585-z.

1226 10. Valverde de Morales, H.G., Wang, H.V., Garber, K., Cheng, X., Corces, V.G., and Li, H.
1227 (2023). Expansion of the genotypic and phenotypic spectrum of CTCF-related disorder
1228 guides clinical management: 43 new subjects and a comprehensive literature review. Am
1229 J Med Genet A *191*, 718-729. 10.1002/ajmg.a.63065.

1230 11. Chen, H., Tian, Y., Shu, W., Bo, X., and Wang, S. (2012). Comprehensive identification
1231 and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human
1232 genome. PLoS One *7*, e41374. 10.1371/journal.pone.0041374.

1233 12. Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and
1234 Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription
1235 Factor Occupancy. Cell Rep *12*, 1184-1195. 10.1016/j.celrep.2015.07.024.

1236 13. Yang, J., Horton, J.R., Liu, B., Corces, V.G., Blumenthal, R.M., Zhang, X., and Cheng, X.
1237 (2023). Structures of CTCF-DNA complexes including all 11 zinc fingers. Nucleic Acids
1238 Res *51*, 8447-8462. 10.1093/nar/gkad594.

1239 14. Nishana, M., Ha, C., Rodriguez-Hernaez, J., Ranjbaran, A., Chio, E., Nora, E.P., Badri,
1240 S.B., Kloetgen, A., Bruneau, B.G., Tsirigos, A., and Skok, J.A. (2020). Defining the relative
1241 and combined contribution of CTCF and CTCFL to genomic regulation. Genome Biol *21*,
1242 108. 10.1186/s13059-020-02024-0.

1243 15. Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N.,
1244 Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF
1245 Decouples Local Insulation of Chromosome Domains from Genomic
1246 Compartmentalization. Cell *169*, 930-944 e922. 10.1016/j.cell.2017.05.004.

1247 16. Holland, A.J., Fachinetti, D., Han, J.S., and Cleveland, D.W. (2012). Inducible, reversible
1248 system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl
1249 Acad Sci U S A *109*, E3350-3357. 10.1073/pnas.1216880109.

1250 17. Nora, E.P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., Uebersohn,
1251 A., Hajj, B., Saux, A.L., Coulon, A., et al. (2020). Molecular basis of CTCF binding polarity
1252 in genome folding. Nat Commun *11*, 5612. 10.1038/s41467-020-19283-x.

1253 18. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A.,
1254 Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an
1255 open platform for exploring multidimensional cancer genomics data. Cancer Discov *2*,
1256 401-404. 10.1158/2159-8290.CD-12-0095.

1257 19. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H.,
1258 Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic
1259 Mutations In Cancer. Nucleic Acids Res *47*, D941-D947. 10.1093/nar/gky1015.

1260 20. Price, E., Fedida, L.M., Pugacheva, E.M., Ji, Y.J., Loukinov, D., and Lobanenkov, V.V.
1261 (2023). An updated catalog of CTCF variants associated with neurodevelopmental
1262 disorder phenotypes. Front Mol Neurosci *16*, 1185796. 10.3389/fnmol.2023.1185796.

1263 21. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic
1264 Acids Res *43*, W39-49. 10.1093/nar/gkv416.

1265 22. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A.,
1266 Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved
1267 ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat
1268 Methods *14*, 959-962. 10.1038/nmeth.4396.

1269 23. Do, C., Jiang, G., Zappile, P., Heguy, A., and Skok, J.A. (2024). A genome wide code to
1270 define cell-type specific CTCF binding and chromatin organization. bioRxiv.
1271 10.1101/2024.11.02.620823.

1272 24. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and
1273 cohesin regulate chromatin loop stability with distinct dynamics. Elife *6*.
1274 10.7554/eLife.25776.

1275 25. Sprague, B.L., Pego, R.L., Stavreva, D.A., and McNally, J.G. (2004). Analysis of binding
1276 reactions by fluorescence recovery after photobleaching. Biophys J *86*, 3473-3495.
1277 10.1529/biophysj.103.026765.

1278 26. Hansen, A.S., Amitai, A., Cattoglio, C., Tjian, R., and Darzacq, X. (2020). Guided nuclear
1279 exploration increases CTCF target search efficiency. Nat Chem Biol *16*, 257-266.
1280 10.1038/s41589-019-0422-3.

1281 27. Iurlaro, M., Masoni, F., Flyamer, I.M., Wirbelauer, C., Iskar, M., Burger, L., Giorgetti, L.,
1282 and Schubeler, D. (2024). Systematic assessment of ISWI subunits shows that NURF
1283 creates local accessibility for CTCF. Nat Genet *56*, 1203-1212. 10.1038/s41588-024-
1284 01767-x.

1285 28. Rahmaninejad, H., Xiao, Y., Tortora, M.M.C., and Fudenberg, G. (2024). Dynamic barriers
1286 modulate cohesin positioning and genome folding at fixed occupancy. bioRxiv.
1287 10.1101/2024.10.08.617113.

1288 29. de Jonge, W.J., Patel, H.P., Meeussen, J.V.W., and Lenstra, T.L. (2022). Following the
1289 tracks: How transcription factor binding dynamics control transcription. Biophys J *121*,
1290 1583-1592. 10.1016/j.bpj.2022.03.026.

1291 30. Mach, P., Kos, P.I., Zhan, Y., Cramard, J., Gaudin, S., Tunnermann, J., Marchi, E.,
1292 Eglinger, J., Zuin, J., Kryzhanovska, M., et al. (2022). Cohesin and CTCF control the
1293 dynamics of chromosome folding. Nat Genet *54*, 1907-1918. 10.1038/s41588-022-01232-
1294 7.

1295 31. Gabriele, M., Brandao, H.B., Grosse-Holz, S., Jha, A., Dailey, G.M., Cattoglio, C., Hsieh,
1296 T.S., Mirny, L., Zechner, C., and Hansen, A.S. (2022). Dynamics of CTCF- and cohesin-
1297 mediated chromatin looping revealed by live-cell imaging. Science *376*, 496-501.
1298 10.1126/science.abn6583.

1299 32. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes.
1300 Nucleic Acids Res *28*, 27-30. 10.1093/nar/28.1.27.

1301 33. Isles, A.R. (2022). The contribution of imprinted genes to neurodevelopmental and
1302 neuropsychiatric disorders. Transl Psychiatry *12*, 210. 10.1038/s41398-022-01972-4.

1303 34. Hirai, H., Karian, P., and Kikyo, N. (2011). Regulation of embryonic stem cell self-renewal
1304 and pluripotency by leukaemia inhibitory factor. Biochem J *438*, 11-23.
1305 10.1042/BJ20102152.

1306 35. Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A.,
1307 Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics
1308 of transcription factor binding during zygotic genome activation. Nat Commun *11*, 4267.
1309 10.1038/s41467-020-18035-1.

1310 36. Lax, E., Do Carmo, S., Enuka, Y., Sapozhnikov, D.M., Welikovitch, L.A., Mahmood, N.,
1311 Rabbani, S.A., Wang, L., Britt, J.P., Hancock, W.W., et al. (2023). Methyl-CpG binding
1312 domain 2 (Mbd2) is an epigenetic regulator of autism-risk genes and cognition. Transl
1313 Psychiatry *13*, 259. 10.1038/s41398-023-02561-9.

1314 37. Mahmood, N., Arakelian, A., Szyf, M., and Rabbani, S.A. (2024). Methyl-CpG binding
1315 domain protein 2 (Mbd2) drives breast cancer progression through the modulation of
1316 epithelial-to-mesenchymal transition. Exp Mol Med *56*, 959-974. 10.1038/s12276-024-
1317 01205-2.

1318 38. Menafra, R., and Stunnenberg, H.G. (2014). MBD2 and MBD3: elusive functions and
1319 mechanisms. Front Genet *5*, 428. 10.3389/fgene.2014.00428.

1320 39. Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A.,
1321 Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012). NuRD suppresses
1322 pluripotency gene expression to promote transcriptional heterogeneity and lineage
1323 commitment. Cell Stem Cell *10*, 583-594. 10.1016/j.stem.2012.02.020.

1324 40. Holwerda, S.J., and de Laat, W. (2013). CTCF: the protein, the binding partners, the
1325 binding sites and their chromatin loops. Philos Trans R Soc Lond B Biol Sci *368*,
1326 20120369. 10.1098/rstb.2012.0369.

1327 41. Gassler, J., Brandao, H.B., Imakaev, M., Flyamer, I.M., Ladstatter, S., Bickmore, W.A.,
1328 Peters, J.M., Mirny, L.A., and Tachibana, K. (2017). A mechanism of cohesin-dependent
1329 loop extrusion organizes zygotic genome architecture. EMBO J *36*, 3600-3618.
1330 10.15252/embj.201798083.

1331 42. Edelmann, W., Yang, K., Umar, A., Heyer, J., Lau, K., Fan, K., Liedtke, W., Cohen, P.E.,
1332 Kane, M.F., Lipford, J.R., et al. (1997). Mutation in the mismatch repair gene Msh6 causes
1333 cancer susceptibility. Cell *91*, 467-477. 10.1016/s0092-8674(00)80433-x.

1334 43. Chen, Y., Liu, P., Sun, P., Jiang, J., Zhu, Y., Dong, T., Cui, Y., Tian, Y., An, T., Zhang, J.,
1335 et al. (2019). Oncogenic MSH6-CXCR4-TGFB1 Feedback Loop: A Novel Therapeutic
1336 Target of Photothermal Therapy in Glioblastoma Multiforme. Theranostics *9*, 1453-1473.
1337 10.7150/thno.29987.

1338 44. van der Weide, R.H., van den Brand, T., Haarhuis, J.H.I., Teunissen, H., Rowland, B.D.,
1339 and de Wit, E. (2021). Hi-C analyses with GENOVA: a case study with cohesin variants.
1340 NAR Genom Bioinform *3*, lqab040. 10.1093/nargab/lqab040.

1341 45. Liu, X.H., Liu, L.P., Xu, X.M., Hua, M., Kang, Q., Li, A., and Huang, L. (2021). FOXN2
1342 suppresses the proliferation and invasion of human hepatocellular carcinoma cells. Eur
1343 Rev Med Pharmacol Sci *25*, 731-737. 10.26355/eurrev_202101_24634.

1344 46. Ma, J., Lu, Y., Zhang, S., Li, Y., Huang, J., Yin, Z., Ren, J., Huang, K., Liu, L., Yang, K.,
1345 et al. (2018). beta-Trcp ubiquitin ligase and RSK2 kinase-mediated degradation of FOXN2
1346 promotes tumorigenesis and radioresistance in lung cancer. Cell Death Differ *25*, 1473-
1347 1485. 10.1038/s41418-017-0055-6.

1348 47. Ye, H., and Duan, M. (2019). FOXN2 is downregulated in breast cancer and regulates
1349 migration, invasion, and epithelial- mesenchymal transition through regulation of SLUG.
1350 Cancer Manag Res *11*, 525-535. 10.2147/CMAR.S176938.

1351 48. Ikeda, K., Horie-Inoue, K., Suzuki, T., Hobo, R., Nakasato, N., Takeda, S., and Inoue, S.
1352 (2019). Mitochondrial supercomplex assembly promotes breast and endometrial
1353 tumorigenesis by metabolic alterations and enhanced hypoxia tolerance. Nat Commun *10*,
1354 4108. 10.1038/s41467-019-12124-6.

1355 49. Do, C., and Skok, J.A. (2024). Factors that determine cell type-specific CTCF binding in
1356 health and disease. Curr Opin Genet Dev *88*, 102244. 10.1016/j.gde.2024.102244.

1357 50. Soochit, W., Sleutels, F., Stik, G., Bartkuhn, M., Basu, S., Hernandez, S.C., Merzouk, S.,
1358 Vidal, E., Boers, R., Boers, J., et al. (2021). CTCF chromatin residence time controls three-
1359 dimensional genome organization, gene expression and DNA methylation in pluripotent
1360 cells. Nat Cell Biol *23*, 881-893. 10.1038/s41556-021-00722-w.

1361 51. Watson, L.A., Wang, X., Elbert, A., Kernohan, K.D., Galjart, N., and Berube, N.G. (2014).
1362 Dual effect of CTCF loss on neuroprogenitor differentiation and survival. J Neurosci *34*,
1363 2860-2870. 10.1523/JNEUROSCI.3769-13.2014.

1364 52. Narducci, D.N., and Hansen, A.S. (2025). Putative looping factor ZNF143/ZFP143 is an
1365 essential transcriptional regulator with no looping function. Mol Cell *85*, 9-23 e29.
1366 10.1016/j.molcel.2024.11.032.

1367 53. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O.,
1368 Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of
1369 biomolecular interactions with AlphaFold 3. Nature *630*, 493-500. 10.1038/s41586-024-
1370 07487-w.

1371 54. Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast,
1372 robust, low-input ChIP-seq for histones and transcription factors. Nat Methods *12*, 963-
1373 965. 10.1038/nmeth.3542.

1374 55. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method
1375 for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-
1376 21 29 29. 10.1002/0471142727.mb2129s109.

1377 56. Dolgalev, I. (2022). Seq-N-Slide [Computer software].
1378 https://doi.org/10.5281/zenodo.5550459.

1379 57. Krueger, F. (2021). TrimGalore. https://zenodo.org/record/7598955.

1380 58. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat
1381 Methods *9*, 357-359. 10.1038/nmeth.1923.

1382 59. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast
1383 processing of NGS alignment formats. Bioinformatics *31*, 2032-2034.
1384 10.1093/bioinformatics/btv098.

1385 60. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum,
1386 C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-
1387 Seq (MACS). Genome Biol *9*, R137. 10.1186/gb-2008-9-9-r137.

1388 61. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist:
1389 Identification of Problematic Regions of the Genome. Sci Rep *9*, 9354. 10.1038/s41598-
1390 019-45839-z.

1391 62. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham,
1392 A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools
1393 and BCFtools. Gigascience *10*. 10.1093/gigascience/giab008.

1394 63. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S.,
1395 Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-
1396 sequencing data analysis. Nucleic Acids Res *44*, W160-165. 10.1093/nar/gkw257.

1397 64. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
1398 Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner.
1399 Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.

1400 65. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
1401 dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-
1402 0550-8.

1403 66. Kolde, R. (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12.

1404 67. Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis
1405 using cumulative statistic calculation. bioRxiv. doi: 10.1101/060012.
1406 http://biorxiv.org/content/early/2016/06/20/060012.

1407 68. Hutchins, A.P., Yang, Z., Li, Y., He, F., Fu, X., Wang, X., Li, D., Liu, K., He, J., Wang, Y.,
1408 et al. (2017). Models of global gene expression define major domains of cell type and
1409 tissue identity. Nucleic Acids Res *45*, 2354-2367. 10.1093/nar/gkx054.

1410 69. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E.,
1411 Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C
1412 data processing. Genome Biol *16*, 259. 10.1186/s13059-015-0831-x.

1413 70. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke,
1414 T., Backofen, R., Ramirez, F., and Gruning, B.A. (2018). Galaxy HiCExplorer: a web
1415 server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids
1416 Res *46*, W11-W16. 10.1093/nar/gky504.

1417 71. Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other
1418 genomically labeled arrays. Bioinformatics *36*, 311-316. 10.1093/bioinformatics/btz540.

1419 72. Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R.,
1420 Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of
1421 chromosome organization. Nat Methods *9*, 999-1003. 10.1038/nmeth.2148.

1422 73. Khan, A.M., A (2017). Intervene: a tool for intersection and visualization of multiple
1423 genomic region sets.

1424 74. Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D.,
1425 Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., et al.
1426 (2018). HOCOMOCO: towards a complete collection of transcription factor binding models
1427 for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res *46*, D252-
1428 D259. 10.1093/nar/gkx1106.

1429 75. Kondili, M., Fust, A., Preussner, J., Kuenne, C., Braun, T., and Looso, M. (2017). UROPA:
1430 a tool for Universal RObust Peak Annotation. Sci Rep *7*, 2593. 10.1038/s41598-017-
1431 02464-y.

1432 76. Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu,
1433 C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). Gencode 2021. Nucleic Acids Res
1434 *49*, D916-D923. 10.1093/nar/gkaa1087.

1435 77. Ay, F., Bailey, T.L., and Noble, W.S. (2014). Statistical confidence estimation for Hi-C data
1436 reveals regulatory chromatin contacts. Genome Res *24*, 999-1011.
1437 10.1101/gr.160374.113.

1438 78. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing
1439 genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.

1440 79. Stansfield, J.C., Cresswell, K.G., and Dozmorov, M.G. (2019). multiHiCcompare: joint
1441 normalization and comparative analysis of complex Hi-C experiments. Bioinformatics *35*,
1442 2916-2923. 10.1093/bioinformatics/btz048.
1443

Figure 1

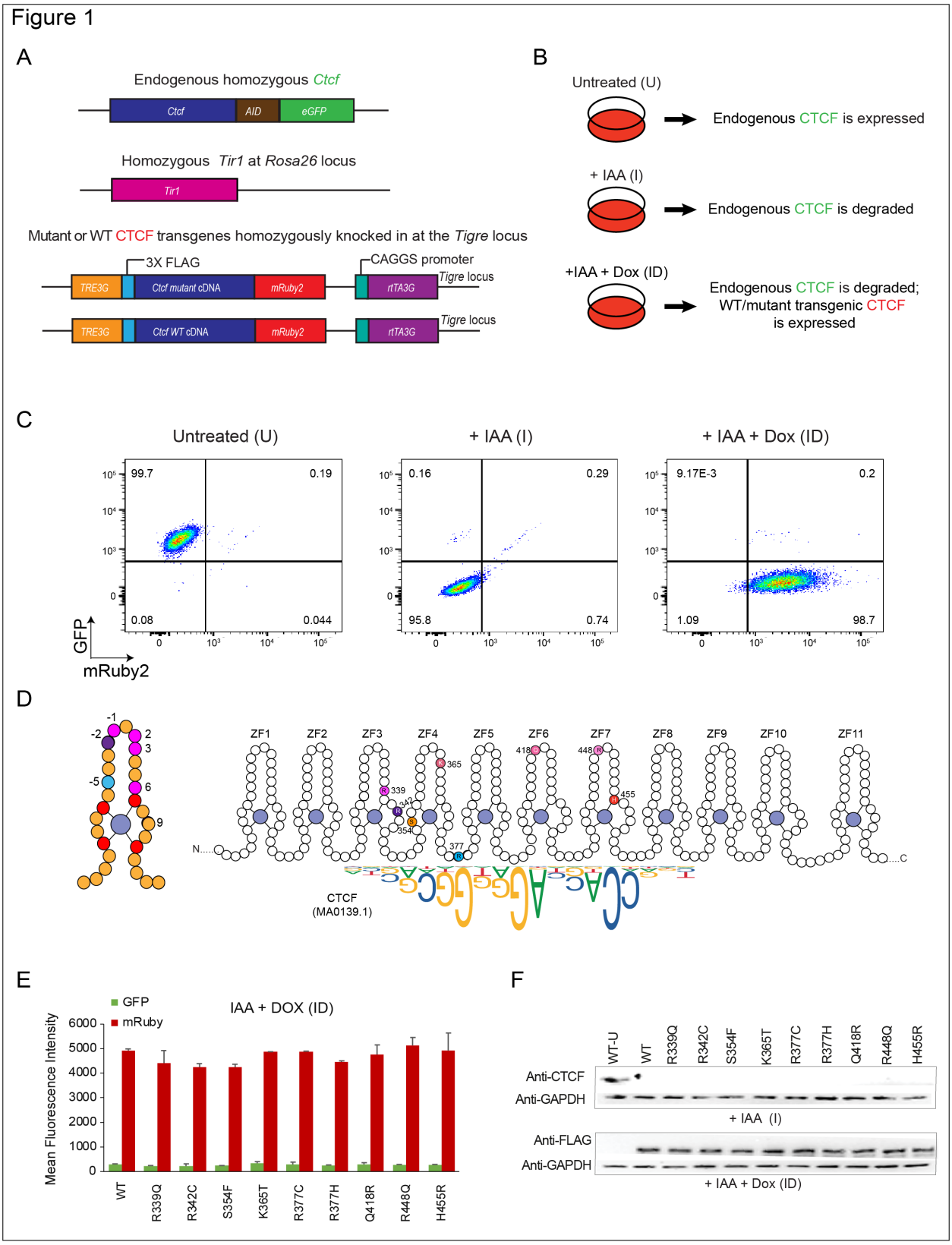
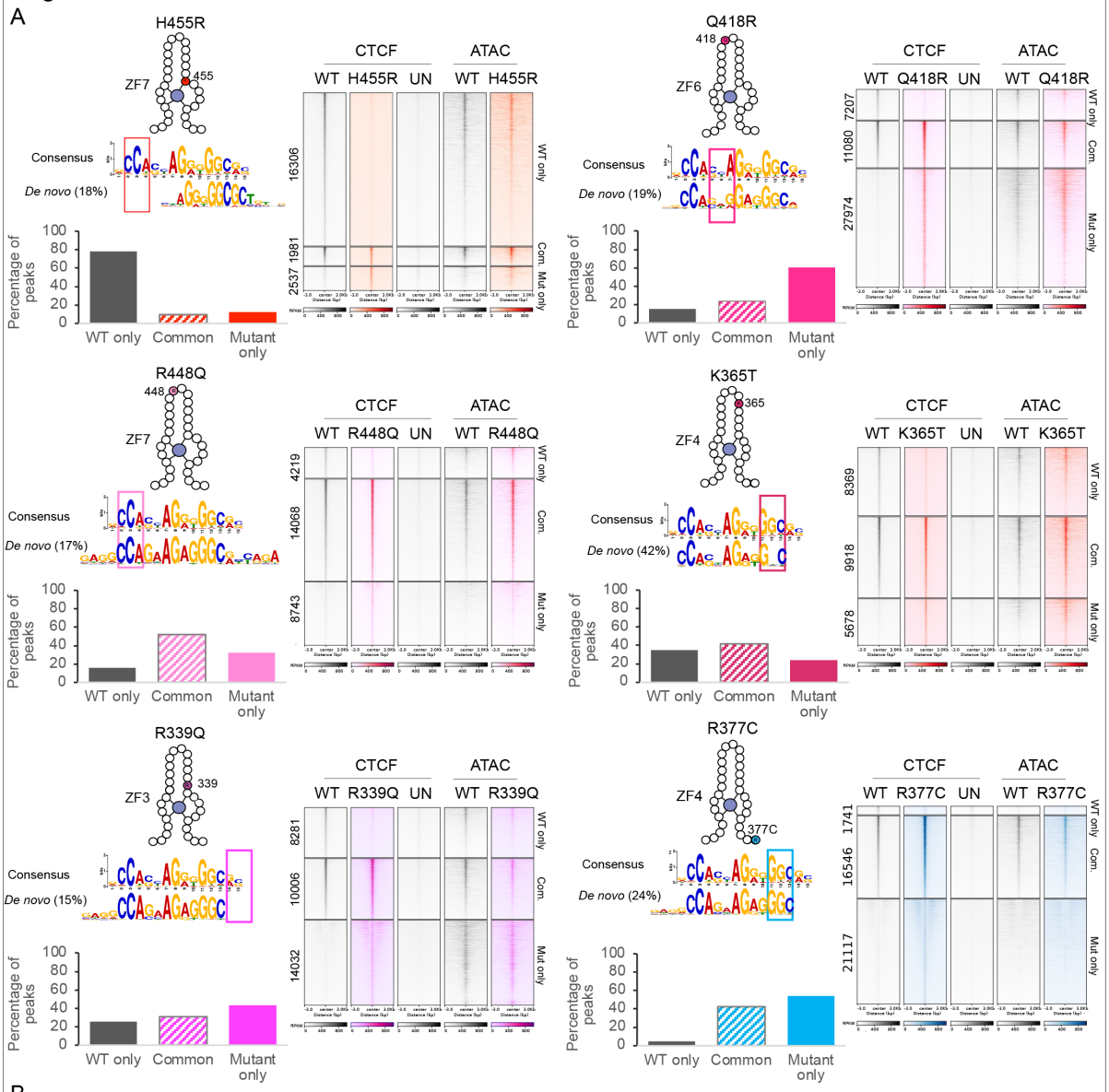

A
Endogenous homozygous Ctcf
Ctcf
AID
eGFP
Homozygous Tir1 at Rosa26 locus
Tir1
Mutant or WT CTCF transgenes homozygously knocked in at the Tigre locus
3X FLAG
CAGGS promoter
TRE3G
Ctcf mutant cDNA
mRuby2
rtTA3G
Tigre locus
TRE3G
Ctcf WT cDNA
mRuby2
rtTA3G
Tigre locus
B
Untreated (U)
Endogenous CTCF is expressed
+ IAA (I)
Endogenous CTCF is degraded
+IAA + Dox (ID)
Endogenous CTCF is degraded;
WT/mutant transgenic CTCF
is expressed
C
Untreated (U)
+ IAA (I)
+ IAA + Dox (ID)
99.7
0.19
0.08
0.044
0.16
0.29
95.8
0.74
9.17E-3
0.2
1.09
98.7
GFP
mRuby2
D
ZF1
ZF2
ZF3
ZF4
ZF5
ZF6
ZF7
ZF8
ZF9
ZF10
ZF11
N
C
CTCF
(MA0139.1)
E
IAA + DOX (ID)
GFP
mRuby
Mean Fluorescence Intensity
WT
R339Q
R342C
S354F
K365T
R377C
R377H
Q418R
R448Q
H455R
F
WT-U
WT
R339Q
R342C
S354F
K365T
R377C
R377H
Q418R
R448Q
H455R
Anti-CTCF
Anti-GAPDH
+ IAA (I)
Anti-FLAG
Anti-GAPDH
+ IAA + Dox (ID)

Figure 2

A


H455R
ZF7
455
Consensus
De novo (18%)
Percentage of peaks
WT only
Common
Mutant only
CTCF
ATAC
WT H455R UN
WT H455R
16306
1981
2537
WT only
Com.
Mut only
Q418R
418
ZF6
De novo (19%)
WT Q418R UN
WT Q418R
7207
11080
27974
R448Q
448
ZF7
De novo (17%)
WT R448Q UN
WT R448Q
4219
14068
8743
K365T
365
ZF4
De novo (42%)
WT K365T UN
WT K365T
8369
9918
5678
R339Q
339
ZF3
De novo (15%)
WT R339Q UN
WT R339Q
8281
10006
14032
R377C
377C
ZF4
De novo (24%)
WT R377C UN
WT R377C
1741
16546
21117

B

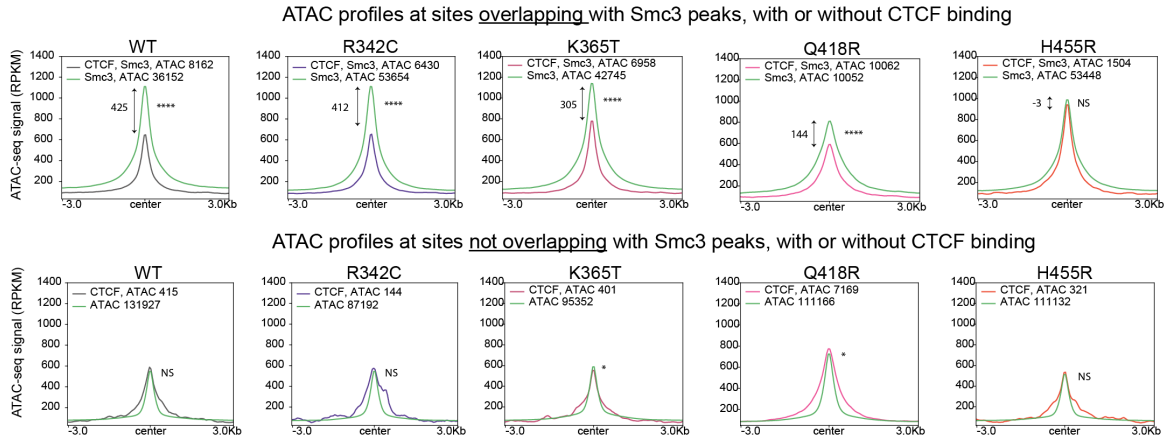

ATAC profiles at sites overlapping with Smc3 peaks, with or without CTCF binding
WT
CTCF, Smc3, ATAC 8162
Smc3, ATAC 36152
425
****
R342C
CTCF, Smc3, ATAC 6430
Smc3, ATAC 53654
412
****
K365T
CTCF, Smc3, ATAC 6958
Smc3, ATAC 42745
305
****
Q418R
CTCF, Smc3, ATAC 10062
Smc3, ATAC 10052
144
****
H455R
CTCF, Smc3, ATAC 1504
Smc3, ATAC 53448
-3
NS
ATAC-seq signal (RPKM)
-3.0
center
3.0Kb
ATAC profiles at sites not overlapping with Smc3 peaks, with or without CTCF binding
WT
CTCF, ATAC 415
ATAC 131927
NS
R342C
CTCF, ATAC 144
ATAC 87192
NS
K365T
CTCF, ATAC 401
ATAC 95352
*
Q418R
CTCF, ATAC 7169
ATAC 111166
*
H455R
CTCF, ATAC 321
ATAC 111132
NS

Figure 3

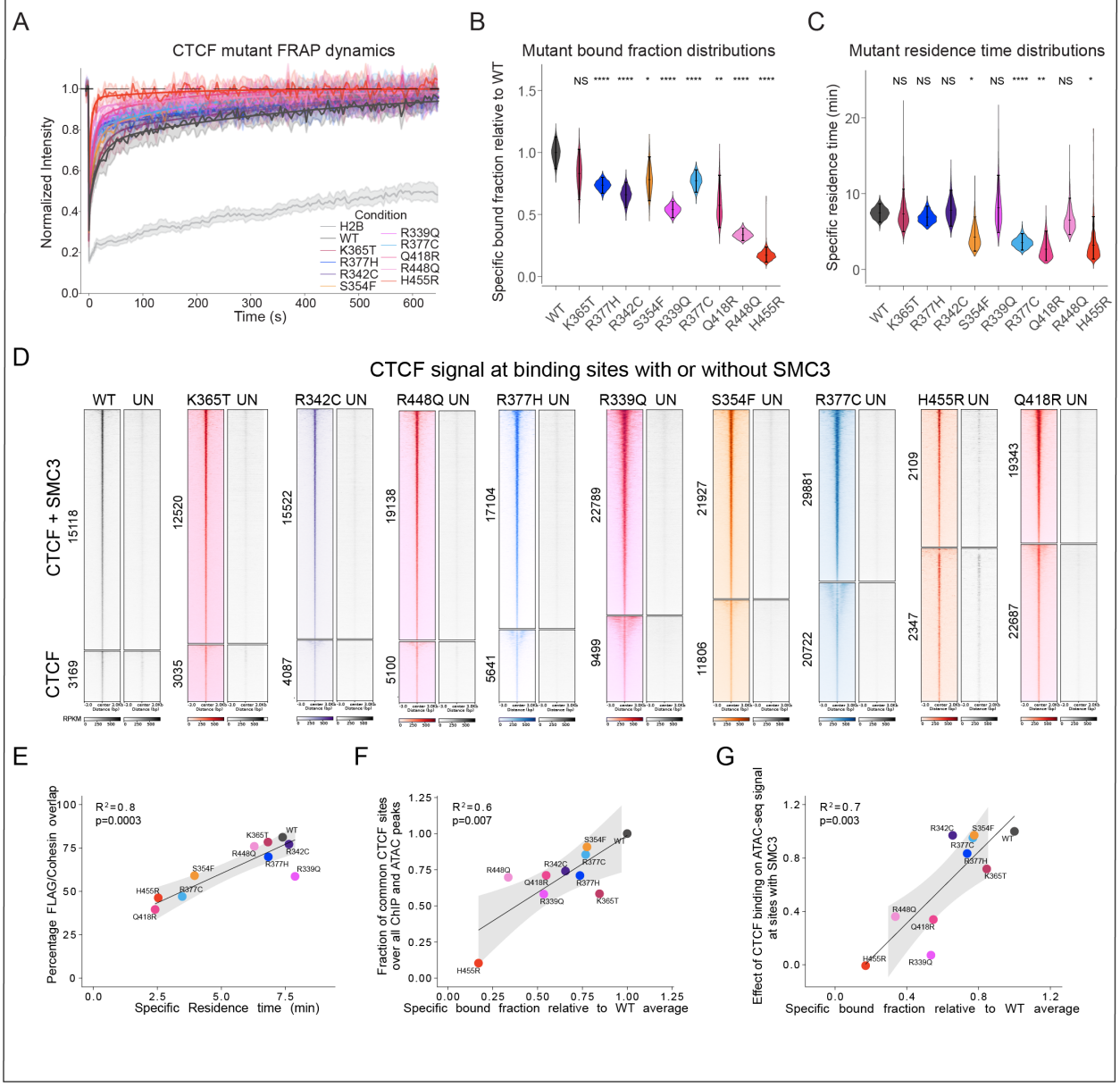
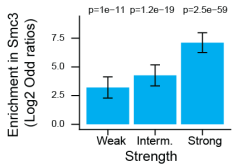
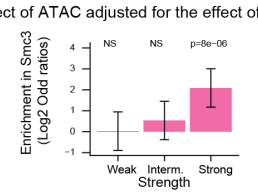

A
CTCF mutant FRAP dynamics
Normalized Intensity
Time (s)
Condition
H2B
WT
K365T
R377H
R342C
S354F
R339Q
R377C
Q418R
R448Q
H455R
B
Mutant bound fraction distributions
Specific bound fraction relative to WT
C
Mutant residence time distributions
Specific residence time (min)
D
CTCF signal at binding sites with or without SMC3
CTCF + SMC3
CTCF
E
Percentage FLAG/Cohesin overlap
Specific Residence time (min)
F
Fraction of common CTCF sites over all ChIP and ATAC peaks
Specific bound fraction relative to WT average
G
Effect of CTCF binding on ATAC-seq signal at sites with SMC3
Specific bound fraction relative to WT average

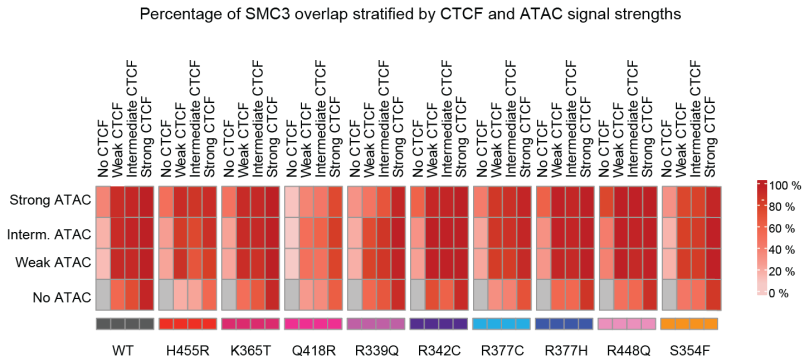
Figure 4

A

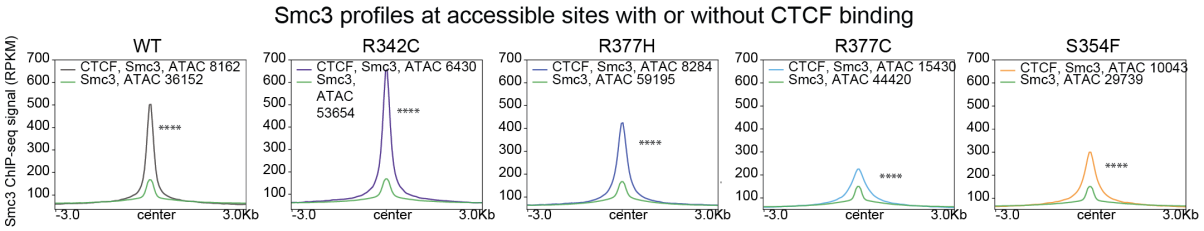

Effect of CTCF adjusted for the effect of ATAC
p=1e−11 p=1.2e−19 p=2.5e−59
Enrichment in Smc3 (Log2 Odd ratios)
7.5
5.0
2.5
0.0
Weak
Interm.
Strong
Strength


Effect of ATAC adjusted for the effect of CTCF
NS
NS
p=8e−06
Enrichment in Smc3 (Log2 Odd ratios)
4
3
2
1
0
−1
Weak
Interm.
Strong
Strength

B


Percentage of SMC3 overlap stratified by CTCF and ATAC signal strengths
No CTCF
Weak CTCF
Intermediate CTCF
Strong CTCF
Strong ATAC
Interm. ATAC
Weak ATAC
No ATAC
100 %
80 %
60 %
40 %
20 %
0 %
WT
H455R
K365T
Q418R
R339Q
R342C
R377C
R377H
R448Q
S354F

C


Smc3 profiles at accessible sites with or without CTCF binding
Smc3 ChIP-seq signal (RPKM)
WT
CTCF, Smc3, ATAC 8162
Smc3, ATAC 36152
R342C
CTCF, Smc3, ATAC 6430
Smc3, ATAC 53654
R377H
CTCF, Smc3, ATAC 8284
Smc3, ATAC 59195
R377C
CTCF, Smc3, ATAC 15430
Smc3, ATAC 44420
S354F
CTCF, Smc3, ATAC 10043
Smc3, ATAC 29739
****
-3.0
center
3.0Kb

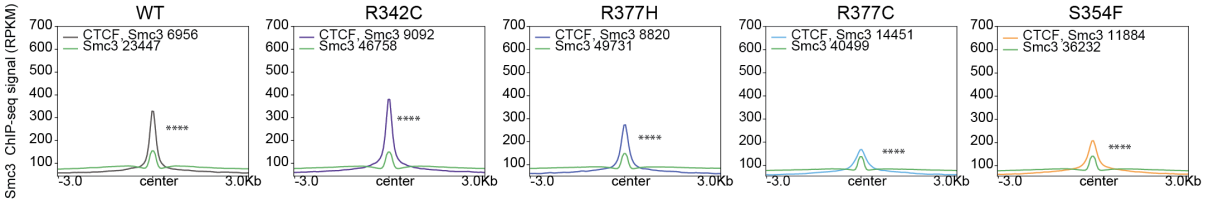

Smc3 profiles at inaccessible sites with or without CTCF binding
Smc3 ChIP-seq signal (RPKM)
WT
CTCF, Smc3 6956
Smc3 23447
R342C
CTCF, Smc3 9092
Smc3 46758
R377H
CTCF, Smc3 8820
Smc3 49731
R377C
CTCF, Smc3 14451
Smc3 40499
S354F
CTCF, Smc3 11884
Smc3 36232
****
-3.0
center
3.0Kb

Figure 5

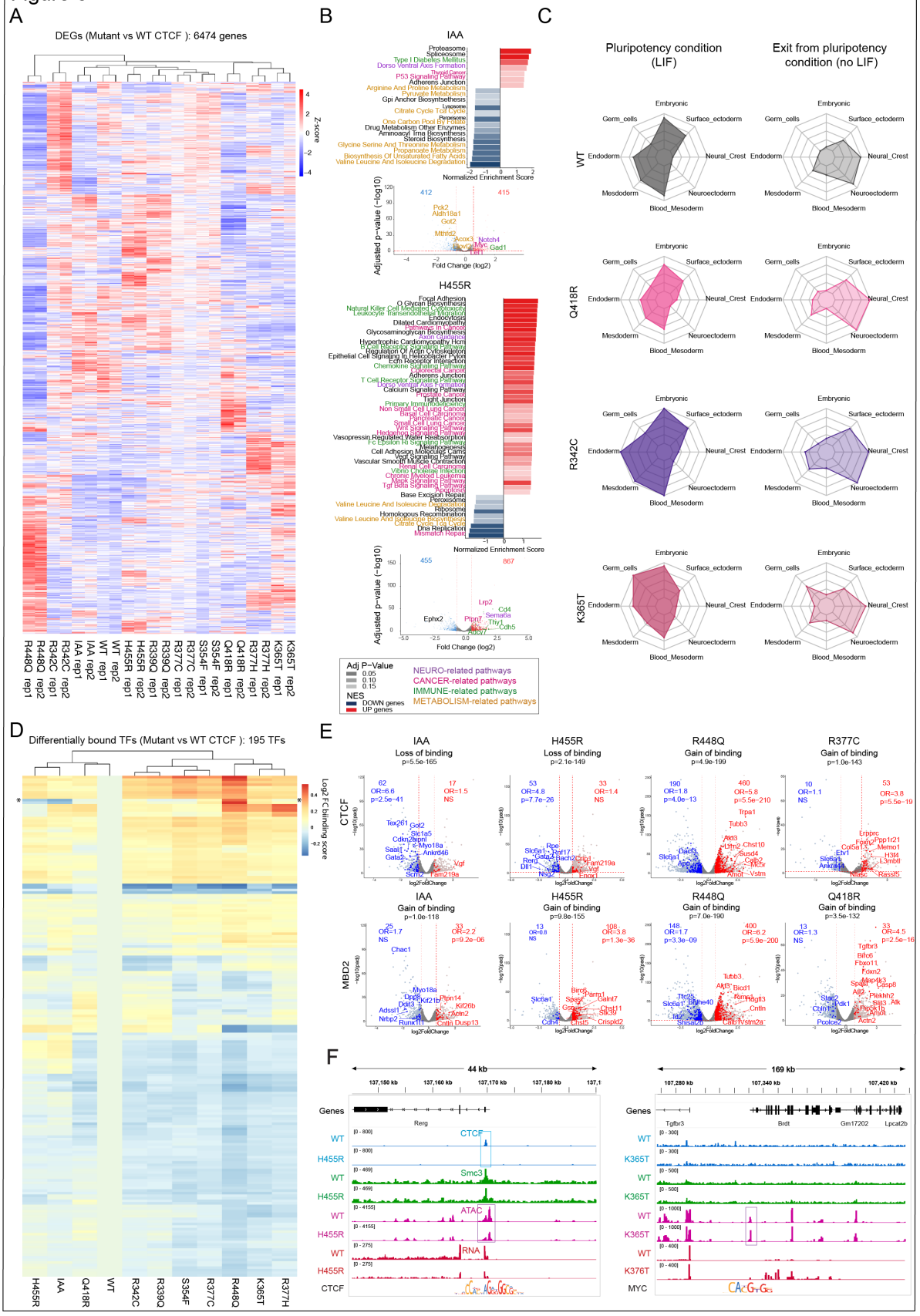

A
DEGs (Mutant vs WT CTCF): 6474 genes
Z-score
R448Q rep1
R448Q rep2
R342C rep1
R342C rep2
IAA rep1
IAA rep2
WT rep1
WT rep2
H455R rep1
H455R rep2
R339Q rep1
R339Q rep2
R377C rep1
R377C rep2
S354F rep1
S354F rep2
Q418R rep1
Q418R rep2
R377H rep1
R377H rep2
K365T rep1
K365T rep2
B
IAA
Normalized Enrichment Score
Adjusted p-value (−log10)
Fold Change (log2)
H455R
Normalized Enrichment Score
Adjusted p-value (−log10)
Fold Change (log2)
Adj P-Value
0.05
0.10
0.15
NES
DOWN genes
UP genes
NEURO-related pathways
CANCER-related pathways
IMMUNE-related pathways
METABOLISM-related pathways
C
Pluripotency condition (LIF)
Exit from pluripotency condition (no LIF)
WT
Q418R
R342C
K365T
Embryonic
Surface_ectoderm
Neural_Crest
Neuroectoderm
Blood_Mesoderm
Mesdoderm
Endoderm
Germ_cells
D
Differentially bound TFs (Mutant vs WT CTCF): 195 TFs
Log2 FC binding score
H455R
IAA
Q418R
WT
R342C
R339Q
S354F
R377C
R448Q
K365T
R377H
E
IAA
Loss of binding
p=5.5e-165
H455R
Loss of binding
p=2.1e-149
R448Q
Gain of binding
p=4.9e-199
R377C
Gain of binding
p=1.0e-143
IAA
Gain of binding
p=1.0e-118
H455R
Gain of binding
p=9.8e-155
R448Q
Gain of binding
p=7.0e-190
Q418R
Gain of binding
p=3.5e-132
CTCF
MBD2
log2FoldChange
F
44 kb
169 kb
Genes
Rerg
Tgfbr3
Brdt
Gm17202
Lpcat2b
WT
H455R
K365T
K376T
CTCF
Smc3
ATAC
RNA
MYC

Figure 6

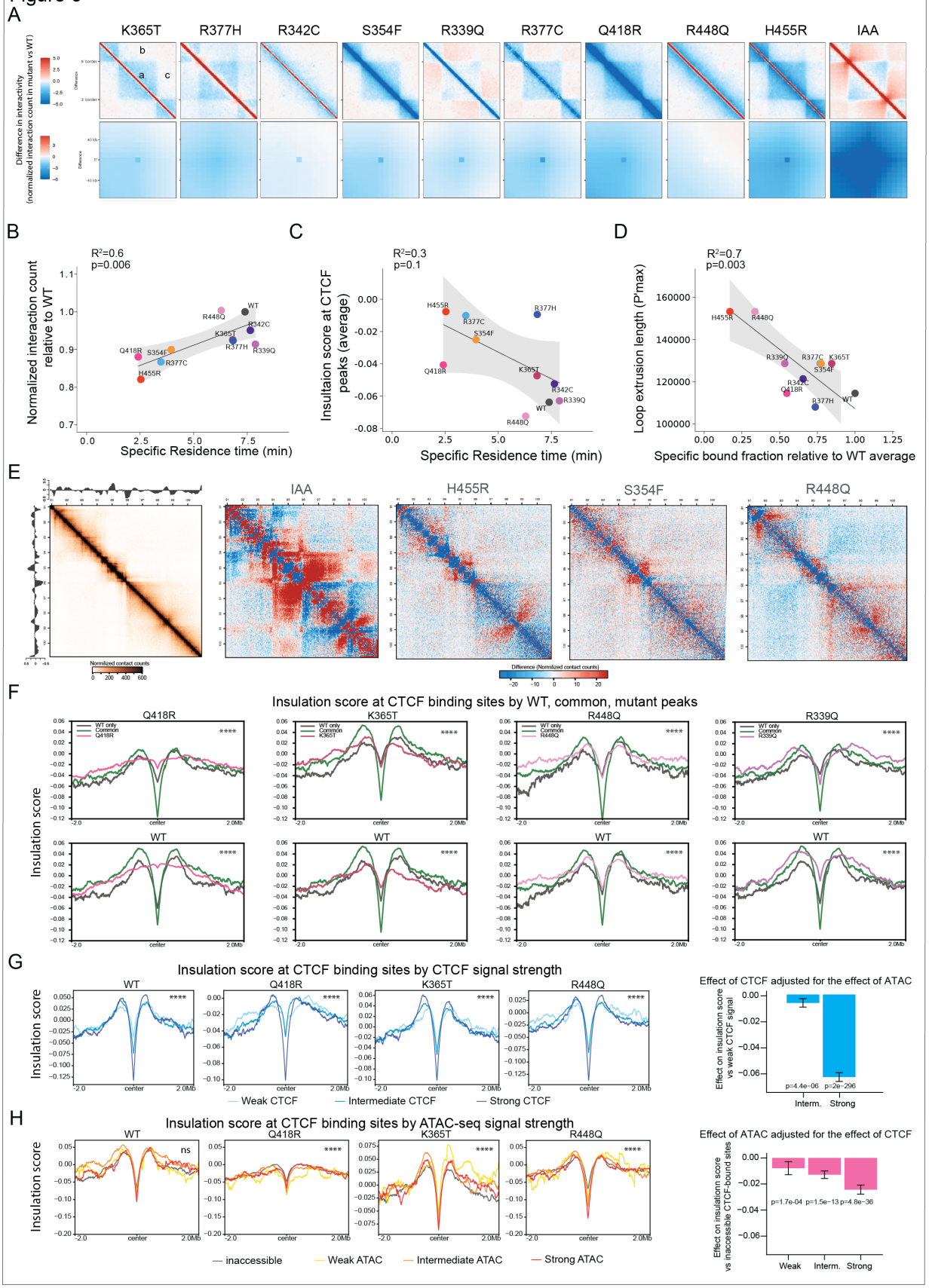
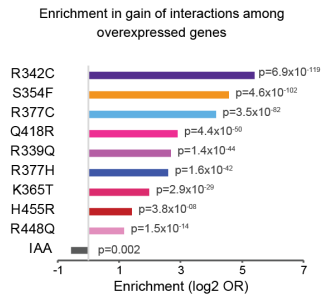
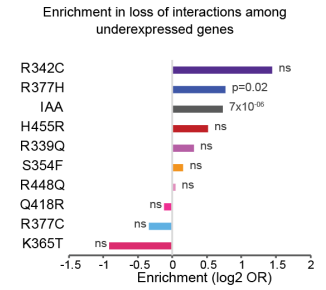

A
K365T
R377H
R342C
S354F
R339Q
R377C
Q418R
R448Q
H455R
IAA
Difference in interactivity
(normalized interaction count in mutant vs WT)
B
R²=0.6
p=0.006
Normalized interaction count relative to WT
Specific Residence time (min)
C
R²=0.3
p=0.1
Insultaion score at CTCF peaks (average)
Specific Residence time (min)
D
R²=0.7
p=0.003
Loop extrusion length (P'max)
Specific bound fraction relative to WT average
E
IAA
H455R
S354F
R448Q
Normalized contact counts
Difference (Normalized contact counts)
F
Insulation score at CTCF binding sites by WT, common, mutant peaks
Q418R
K365T
R448Q
R339Q
WT
Insulation score
G
Insulation score at CTCF binding sites by CTCF signal strength
Weak CTCF
Intermediate CTCF
Strong CTCF
Effect of CTCF adjusted for the effect of ATAC
Effect on insulationn score vs weak CTCF signal
p=4.4e-06
p=2e-296
Interm.
Strong
H
Insulation score at CTCF binding sites by ATAC-seq signal strength
inaccessible
Weak ATAC
Intermediate ATAC
Strong ATAC
Effect of ATAC adjusted for the effect of CTCF
Effect on insulationn score vs inaccessible CTCF-bound sites
p=1.7e-04
p=1.5e-13
p=4.8e-36
Weak
Interm.
Strong

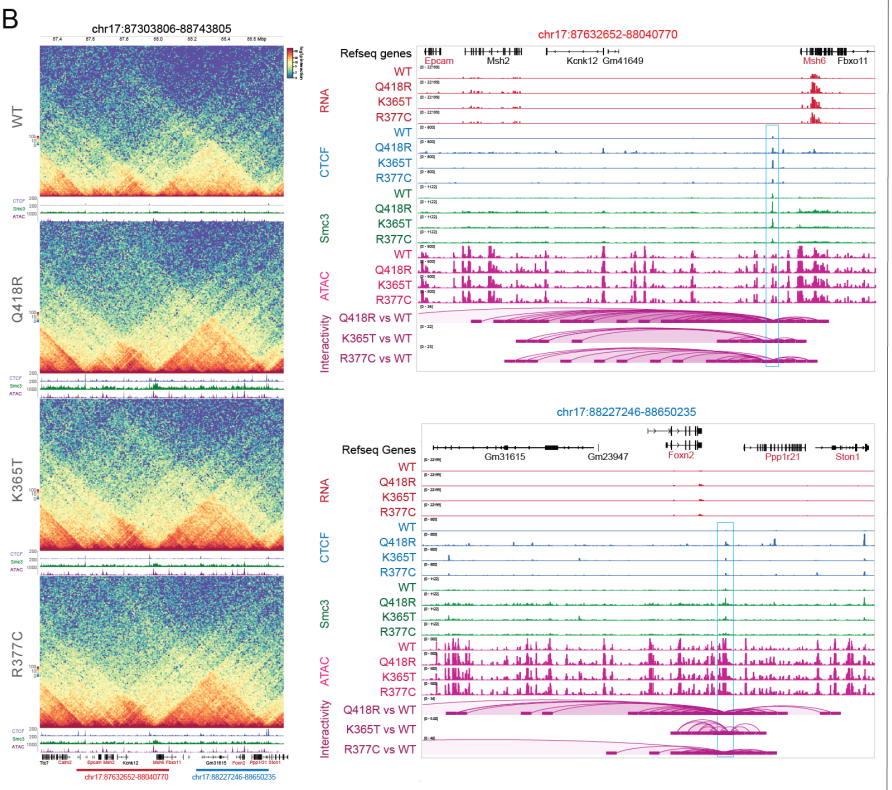
Figure 7

A

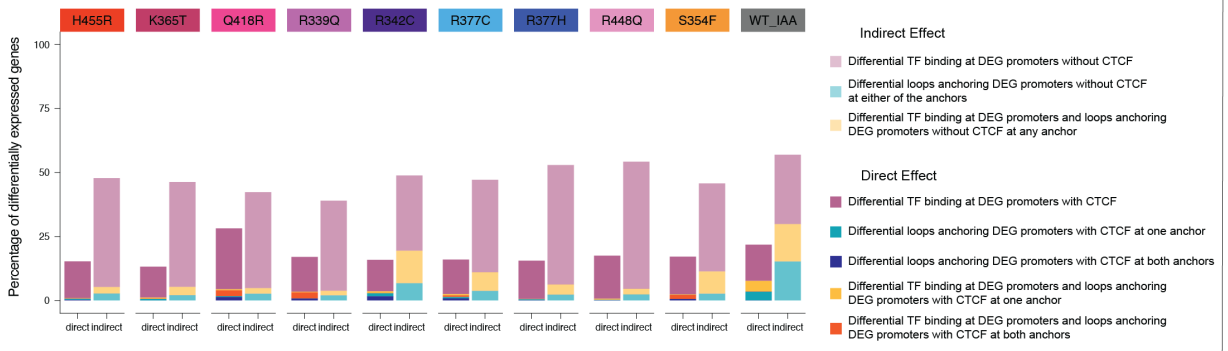

Enrichment in gain of interactions among overexpressed genes
R342C
S354F
R377C
Q418R
R339Q
R377H
K365T
H455R
R448Q
IAA
p=6.9x10-119
p=4.6x10-102
p=3.5x10-82
p=4.4x10-50
p=1.4x10-44
p=1.6x10-42
p=2.9x10-29
p=3.8x10-08
p=1.5x10-14
p=0.002
-1
1
3
5
7
Enrichment (log2 OR)


Enrichment in loss of interactions among underexpressed genes
R342C
R377H
IAA
H455R
R339Q
S354F
R448Q
Q418R
R377C
K365T
ns
p=0.02
7x10-06
-1.5
-1
-0.5
0
0.5
1
1.5
2
Enrichment (log2 OR)

B


chr17:87303806-88743805
WT
Q418R
K365T
R377C
chr17:87632652-88040770
Refseq genes
Epcam
Msh2
Kcnk12
Gm41649
Msh6
Fbxo11
RNA
CTCF
Smc3
ATAC
Interactivity
Q418R vs WT
K365T vs WT
R377C vs WT
chr17:88227246-88650235
Refseq Genes
Gm31615
Gm23947
Foxn2
Ppp1r21
Ston1

C


H455R
K365T
Q418R
R339Q
R342C
R377C
R377H
R448Q
S354F
WT_IAA
Percentage of differentially expressed genes
100
75
50
25
0
direct indirect
Indirect Effect
Differential TF binding at DEG promoters without CTCF
Differential loops anchoring DEG promoters without CTCF at either of the anchors
Differential TF binding at DEG promoters and loops anchoring DEG promoters without CTCF at any anchor
Direct Effect
Differential TF binding at DEG promoters with CTCF
Differential loops anchoring DEG promoters with CTCF at one anchor
Differential loops anchoring DEG promoters with CTCF at both anchors
Differential TF binding at DEG promoters and loops anchoring DEG promoters with CTCF at one anchor
Differential TF binding at DEG promoters and loops anchoring DEG promoters with CTCF at both anchors

SUPPLEMENTAL FIGURES

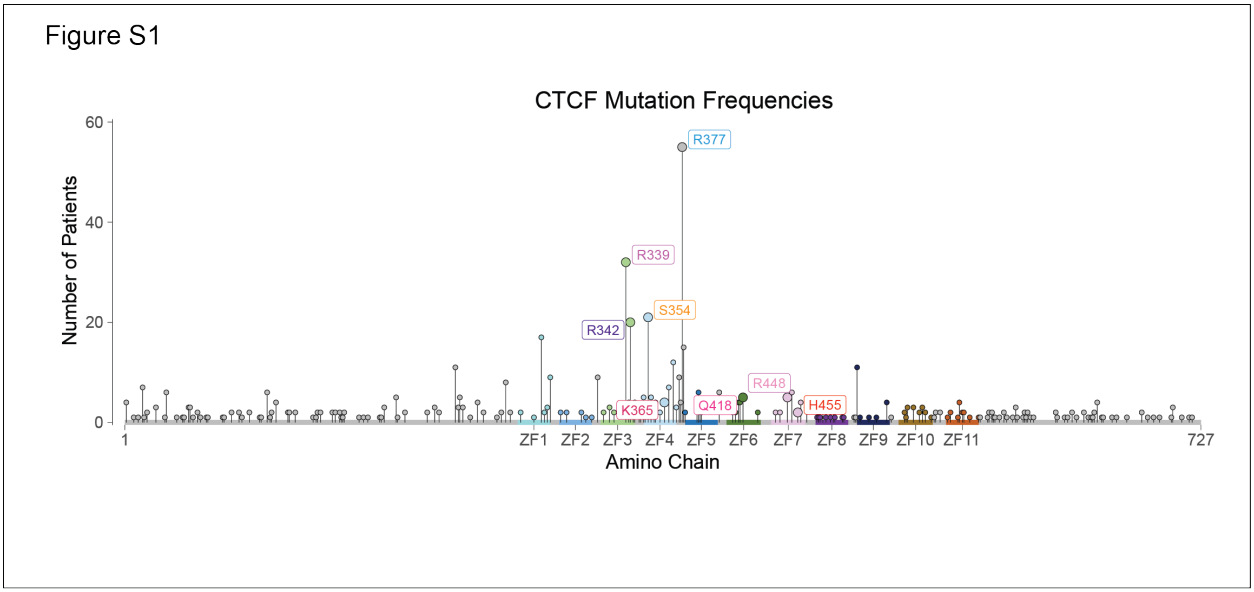


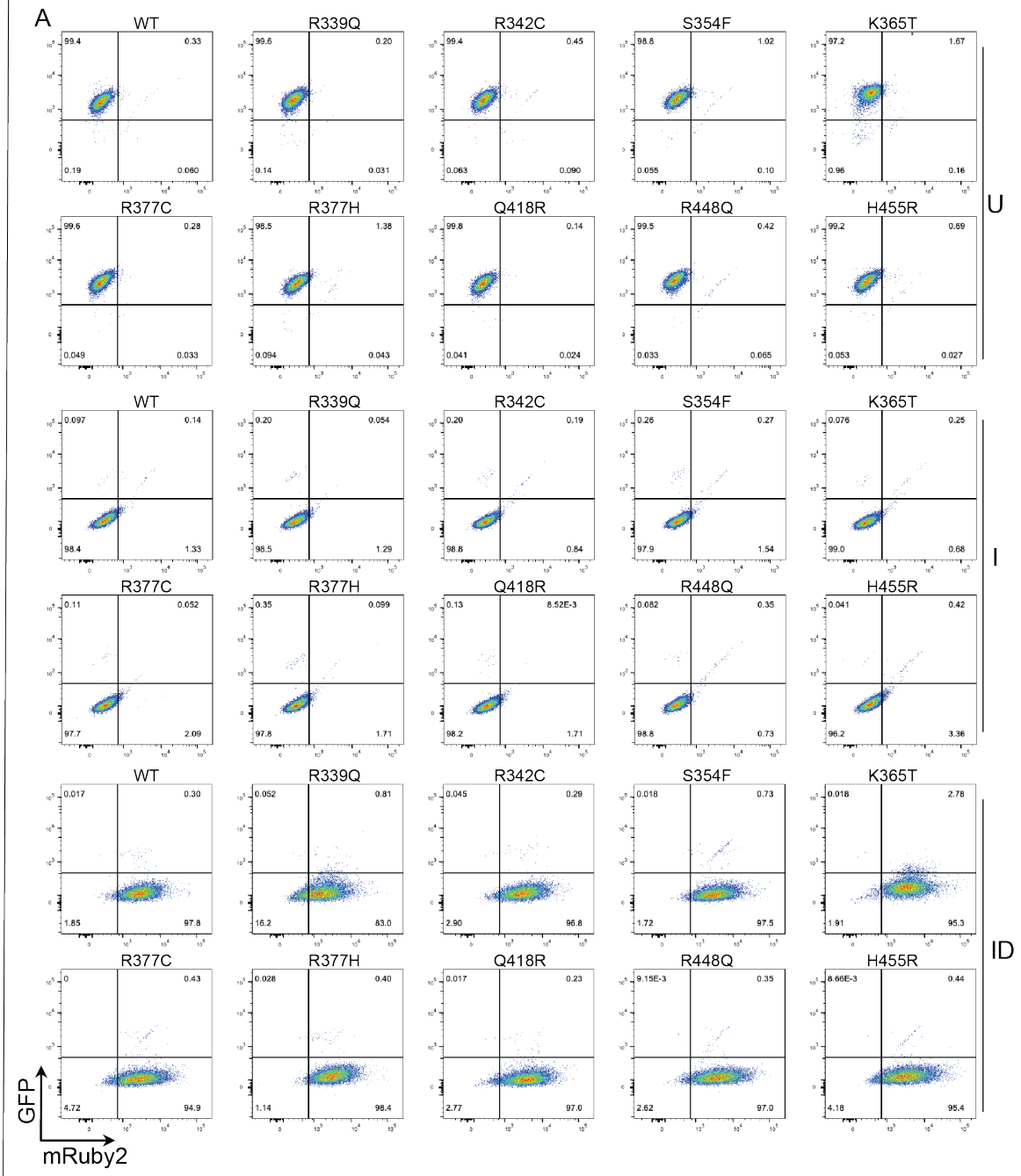
Figure S1. Prevalence of missense CTCF mutations on ZF3-ZF7 in human cancers.
Schematic representation of CTCF showing the location and frequency of each mutation under investigation. This figure relates to **Figure 1**.

Figure S2

Mutations reported in > than 2 samples	Zinc Finger	No. of samples	Associated cancers	Associated neurological disorder
Mutations in phosphate contacts (positions -5, 4th position of the linker between adjacent ZFs)				
R377C*	5	27	Breast, Colorectal, Endometrial, Haematopoietic And Lymphoid, Skin Cancer	ASD, CRD
R377H*	5	26	Breast, Colorectal, Endometrial, Esophagogastric, Haematopoietic And Lymphoid, Urinary Tract	ASD, CRD
R320C	3	6	Colorectal, Endometrial, Prostate Cancer	
R320H	3	3	Colorectal, Esophagogastric, Head And Neck Cancer	
K349E	4	2	Endometrial Cancer	
R377L	5	2	Breast, Esophagogastric Cancer	
Y358C	4	2	Colorectal, Endometrial Cancer	
Mutations in the amino acid that make direct contact with DNA (positions -1, +2, +3 and +6)				
R339W	3	20	Colorectal, Endometrial, Haematopoietic And Lymphoid, Renal Cancer	
R339Q*	3	12	Endometrial, Esophagogastric, Haematopoietic And Lymphoid Cancer	CRD,ASD,DD
Q418R*	6	5	Embryonal Tumor, Glioma, Hepatobiliary Cancer	
R448Q*	7	5	Endometrial, Haematopoietic And Lymphoid, Renal Cancer	CRD,ASD,DD
K365T*	4	4	Endometrial Cancer	
R368C	4	4	Colorectal, Esophagogastric, Haematopoietic And Lymphoid Cancer	DD,NDD,CRD,ASD
D451H	7	3	Skin, Urinary Tract Cancer	
D451V	7	3	Haematopoietic And Lymphoid Cancer	
R368H	4	3	Breast, Ovary, Prostate Cancer	CRD,ASD,DD
E362K	4	2	Haematopoietic And Lymphoid Cancer	
D390N	5	2	Haematopoietic And Lymphoid Cancer	CRD,ASD
Mutations in the amino acids in the other regions of the zinc finger				
S354F*	4	21	Breast, Endometrial, Esophagogastric, Head And Neck, Hepatobiliary, Skin, Urinary Tract Cancer	
P378L	5	10	Endometrial, Haematopoietic And Lymphoid Cancer	NDD, CRD, DD
S388N	5	6	Haematopoietic And Lymphoid, Hepatobiliary, Urinary Tract Cancer	
S402L	5	6	Breast Cancer	
P378A	5	5	Breast, Urinary Tract Cancer	
F351L	4	5	Breast, Urinary Tract Cancer	
G375V	4	5	Endometrial, Haematopoietic And Lymphoid Cancer	
A359T	4	4	Colorectal, Ovary, Prostate Cancer	
A387V	5	4	Breast, Endometrial Cancer	
R415W	6	4	Colorectal, Endometrial, Penile Cancer	
F416L	6	4	Endometrial, Glioma Cancer	
E376K	4	4	Breast, Haematopoietic And Lymphoid, Urinary Tract Cancer	
K352N	4	3	Endometrial Cancer	
D328N	3	3	Colorectal, Endometrial Cancer	
S461F	7	3	Skin Cancer	
E348K	3	3	Colorectal, Haematopoietic And Lymphoid Cancer	
F379I	5	2	Haematopoietic And Lymphoid Cancer	
P440S	7	2	Skin Cancer	
I411V	6	2	Urinary Tract Cancer	
H413Y	6	2	Colorectal Cancer	
D443G	7	2	Hepatobiliary Cancer	
F331Y	3	2	Haematopoietic And Lymphoid Cancer	
R341C	3	2	Colorectal, Skin Cancer	
R341H	3	2	Head And Neck Cancer	
Q428E	6	2	Haematopoietic And Lymphoid Cancer	
K344E	3	2	Breast, Haematopoietic And Lymphoid Cancer	
G375A	4	2	Haematopoietic And Lymphoid, Urinary Tract Cancer	
G375R	4	2	Head And Neck, Lung Cancer	
Mutations in boundary residues (position -2 and +9)				
R342C*	3	13	Breast, Endometrial, Esophagogastric, Haematopoietic And Lymphoid, Skin, Urinary Tract Cancer	NDD,CRD,DD
R371H	4	8	Colorectal, Endometrial, Haematopoietic And Lymphoid, Lung Cancer	
R342H	3	7	Endometrial, Haematopoietic And Lymphoid Cancer	CRD
R371C	4	4	Endometrial, Skin, Urinary Tract Cancer	
R457Q	7	4	Colorectal, Esophagogastric Cancer	
R389S	5	2	Breast, Skin Cancer	
Mutations in the amino acid that co-ordinate the zinc (position -10, -7, +7, +11 or +12 depending on ZF)				
C356R	4	3	Esophagogastric Cancer	
H373L	4	3	Endometrial Cancer	NDD,ASD
H455R*	7	2	Haematopoietic And Lymphoid Cancer	
C324Y	3	2	Haematopoietic And Lymphoid Cancer	
C356F	4	2	Haematopoietic And Lymphoid, Head And Neck Cancer	
H345P	3	2	Breast Cancer	
H345R	3	2	Colorectal, Prostate Cancer	

Figure S2: Table of CTCF mutations. Table showing CTCF mutations identified from cBioPortal [S1] and COSMIC [S2]. We selected a subset of mutations that occur at high frequency in cancer patients, focusing on those found in ZFs that make contact with the 12-15bp consensus CTCF binding motif. These are highlighted with an asterisk and color coded as in **Figure 1D.** Mutations analyzed included those found in **(i)** amino acids that make base-specifying contacts with the alpha helix of DNA (at locations -2, 2, 3 and 6 on the ZF), **(ii)** residues that coordinate the zinc ion which are essential for providing stability to ensure the proper folding of the ZF domain, **(iii)** two other classes of amino acid residues in the ZFs: boundary residues (that are important for interactions between ZFs), and residues that contact the sugar phosphate backbone of DNA, and **(iv)** other highly mutated residues in the region. The table identifies which zinc finger each mutation is found in, the number of patients it was identified in and the cancers each is associated with. Mutations were annotated for neurodevelopmental or neurological disorders using the CTCF variant catalog published in Price et al [S3]. ASD, autism spectrum disorder; CRD, CTCF related neurodevelopmental disorder; DD developmental disorder; NDD, neurodevelopmental disorder. This figure relates to **Figure 1**.

Figure S3


A
WT
R339Q
R342C
S354F
K365T
R377C
R377H
Q418R
R448Q
H455R
U
I
ID
GFP
mRuby2

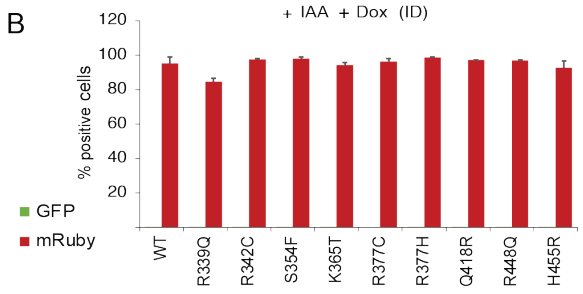

B
+ IAA + Dox (ID)
% positive cells
120
100
80
60
40
20
GFP
mRuby
WT
R339Q
R342C
S354F
K365T
R377C
R377H
Q418R
R448Q
H455R

Figure S3: **Flow cytometry analysis of mutant expression.** (**A**) Flow cytometry showing the level of GFP (endogenous CTCF) and mRuby (transgenic CTCF) in the IAA and Dox treated cells (ID). (**B**) Bar graph showing the percentage of GFP (endogenous CTCF) and mRuby2 (transgene WT and mutant CTCF) positive cells after treatment with IAA and Dox (ID). The presence of the IAA leads to degradation of GFP labelled endogenous CTCF. The error bars represent the standard deviation between the 2 replicates. This Figure relates to **Figure 1.**

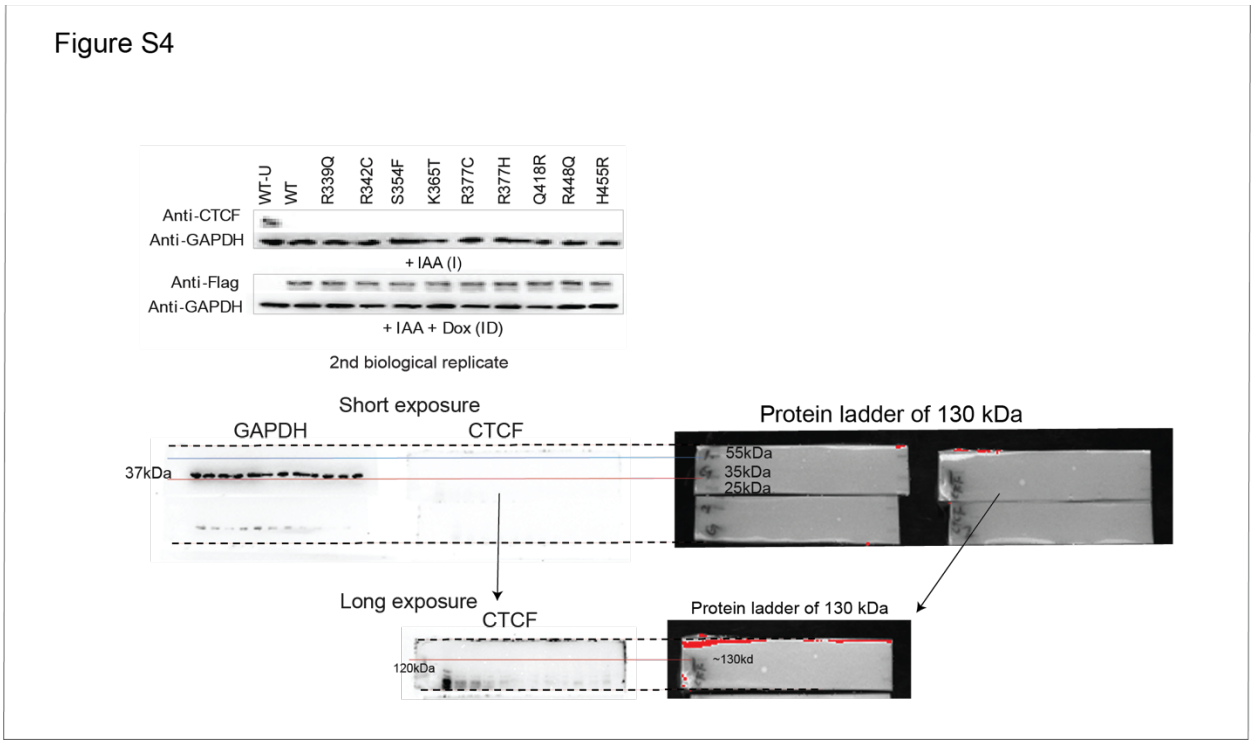


Figure S4: **Western Blot of endogenous and transgenic CTCF.** Western blot showing degradation of endogenous CTCF as detected with an antibody to CTCF, and the induction of transgenic CTCF as detected using an antibody to FLAG. The first replicate is shown in **Figure 1**. The uncropped pictures of the WB gels are shown below. This figure relates to **Figure 1**.

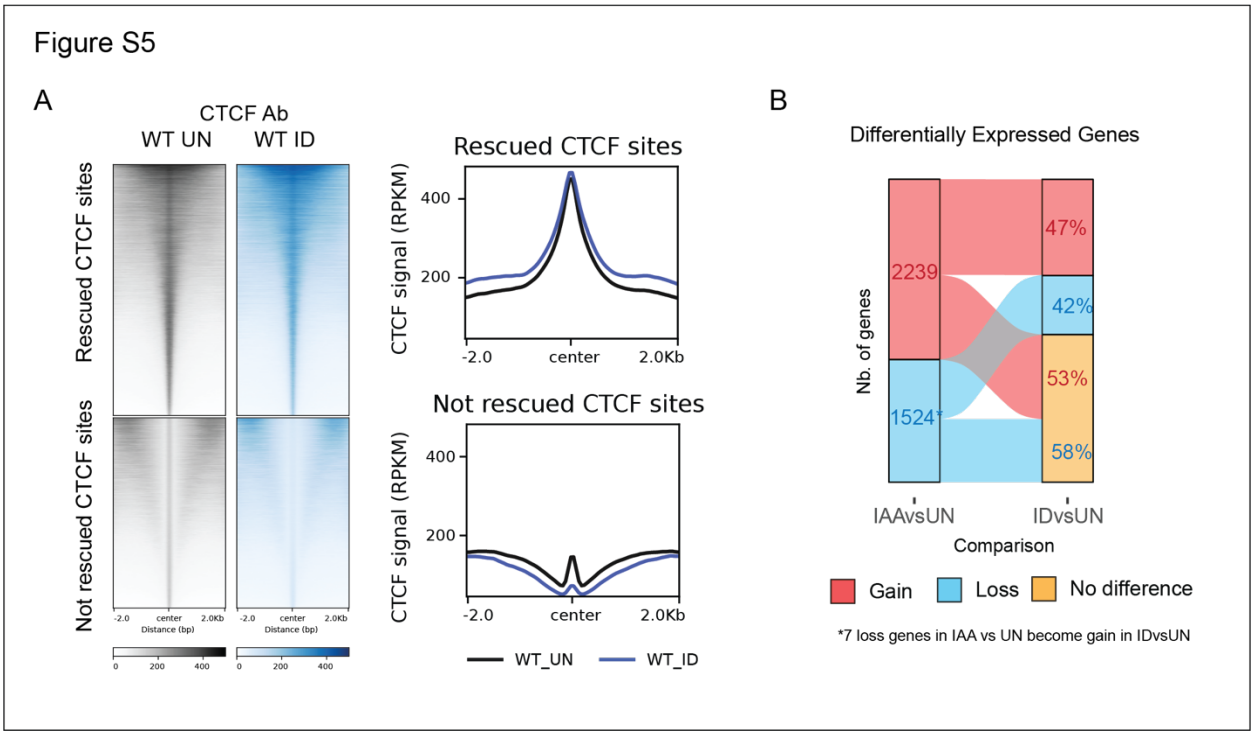
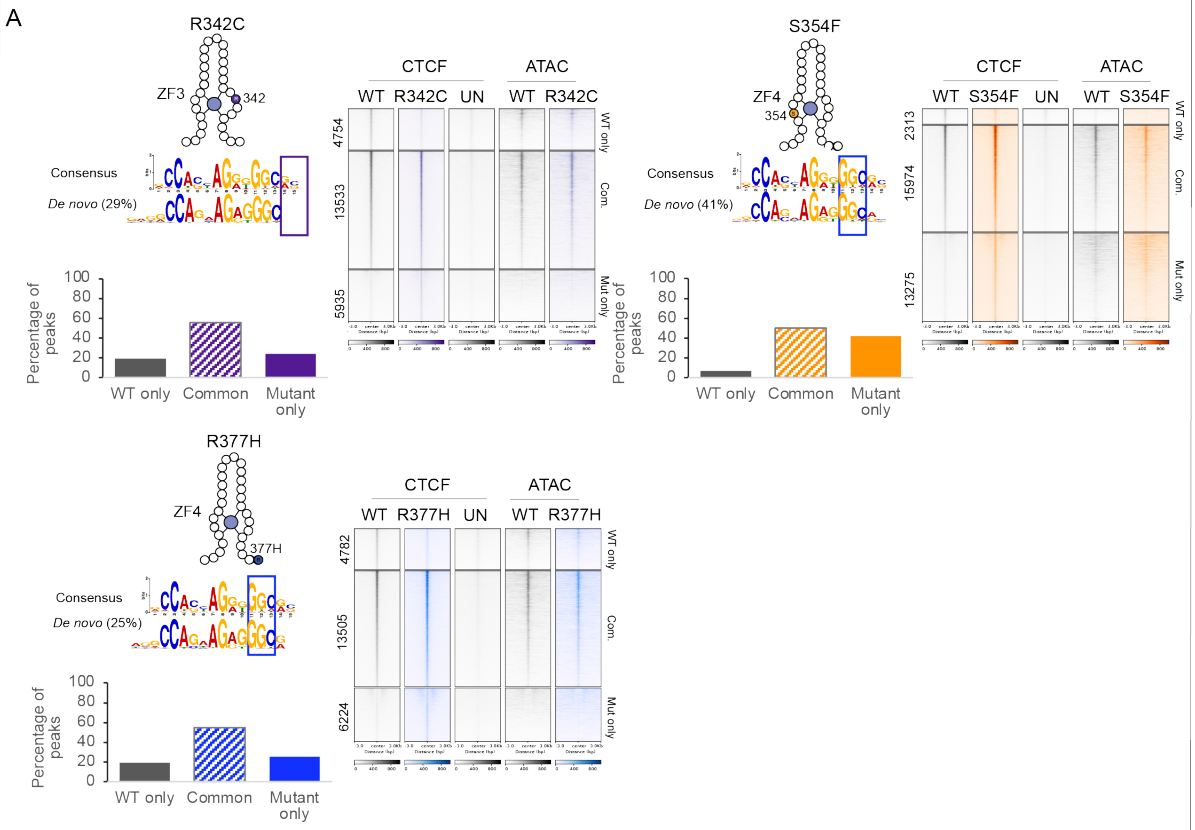


Figure S5. Rescue of CTCF binding and gene expression by the transgene WT CTCF. (**A**) Heatmap and profiles showing the rescued CTCF binding sites in WT-ID compared to WT-UN. (**B**) Alluvial plot showing the number of DEGs in IAA versus WT untreated cells and WT-ID versus WT untreated fells. The plot shows the percentage of genes that return to baseline in the WT-ID condition (no significant expression level changes). DEGs were defined as fold change>1.5 and adjusted p-value<0.05. This figure relates to **Figure 1.**

Figure S6

A


R342C
ZF3
342
Consensus
De novo (29%)
Percentage of peaks
WT only
Common
Mutant only
CTCF
ATAC
WT R342C UN
WT R342C
4754
13533
5935
S354F
ZF4
354
De novo (41%)
WT S354F UN
WT S354F
2313
15974
13275
R377H
377H
De novo (25%)
WT R377H UN
WT R377H
4782
13505
6224

B

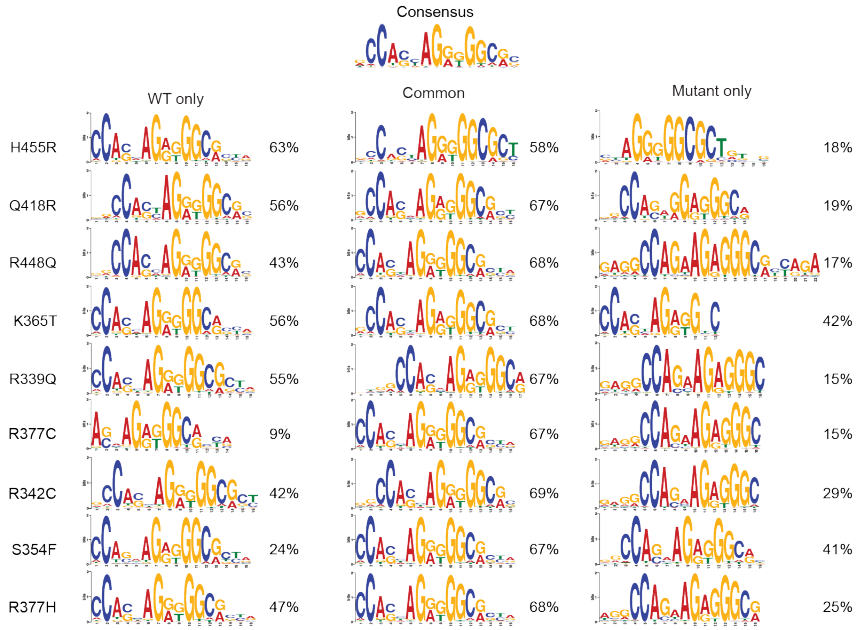

Consensus
WT only
Common
Mutant only
H455R 63% 58% 18%
Q418R 56% 67% 19%
R448Q 43% 68% 17%
K365T 56% 68% 42%
R339Q 55% 67% 15%
R377C 9% 67% 15%
R342C 42% 69% 29%
S354F 24% 67% 41%
R377H 47% 68% 25%

Figure S6: **CTCF mutations have unique chromatin binding profiles. (A**) Scheme showing the locations of the different CTCF mutations within the ZF. The consensus CTCF motif highlights the triplet to which each mutant zinc finger binds (top) as well as the motif identified by MEME for the *de novo* mutant only binding sites (bottom). Each graph shows the percentage of WT and mutant CTCF binding at WT only, common and mutant only sites and heatmaps show the profile of CTCF binding and ATAC-seq signal for WT only, common and mutant only sites. The remaining mutants are shown in **Figure 2**. (**B**) Most prevalent motifs found at WT only, common and mutant only sites. The motifs were identified using MEME de novo motif search. The percentages correspond to the percentage of peaks containing the motif. This figure relates to **Figure 2**.

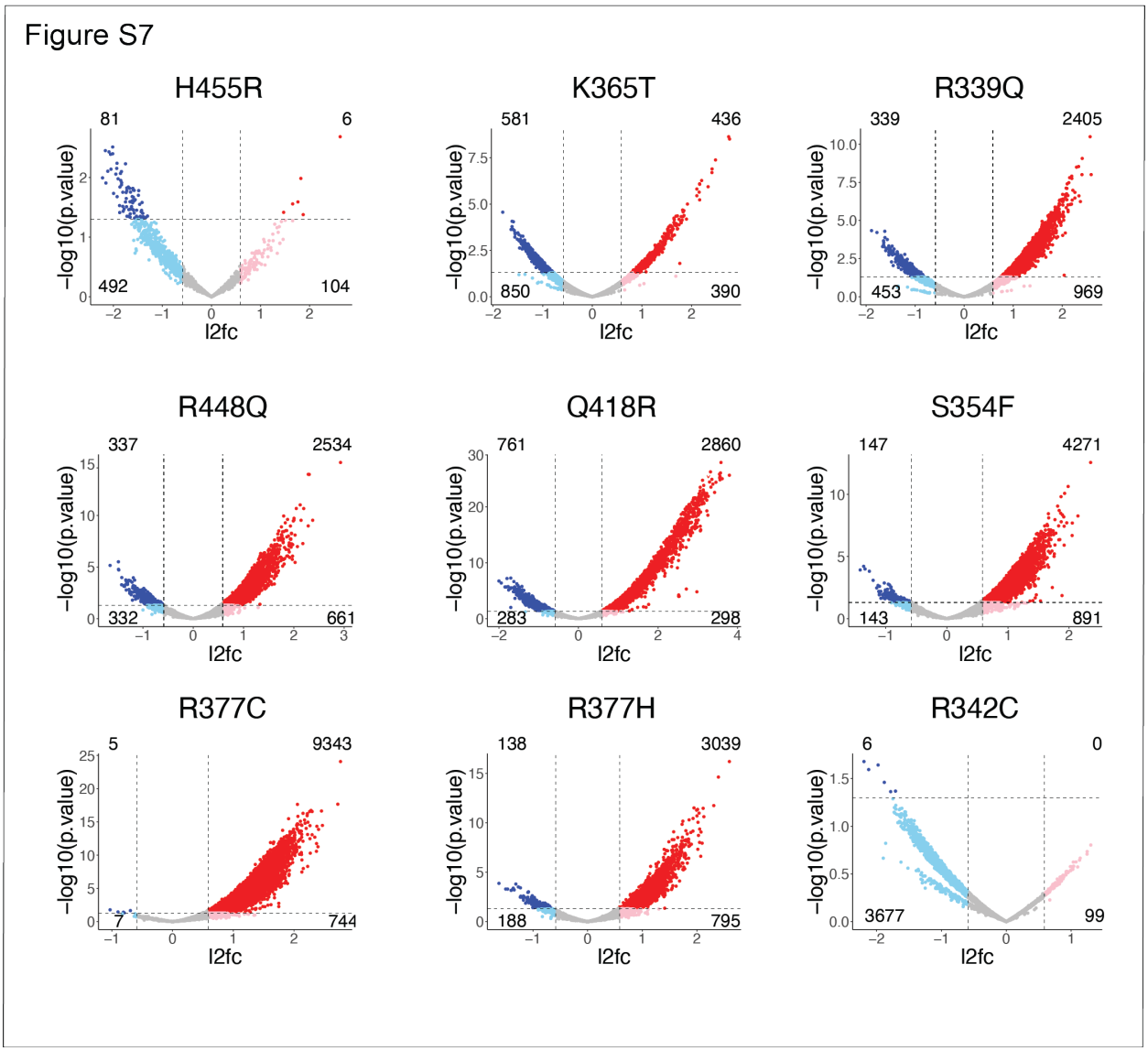


Figure S7. Differential CTCF binding in mutant versus WT-ID. Volcano plots showing significant differences in WT and mutant transgenic CTCF binding ($p<0.05$ and fold change > 1.5 (red) or fold change < 1.5 (blue). The subthreshold of differentially bound sites are shown in light blue and pink. The p-values were computed using DiffBind. This figure relates to **Figure 2.**

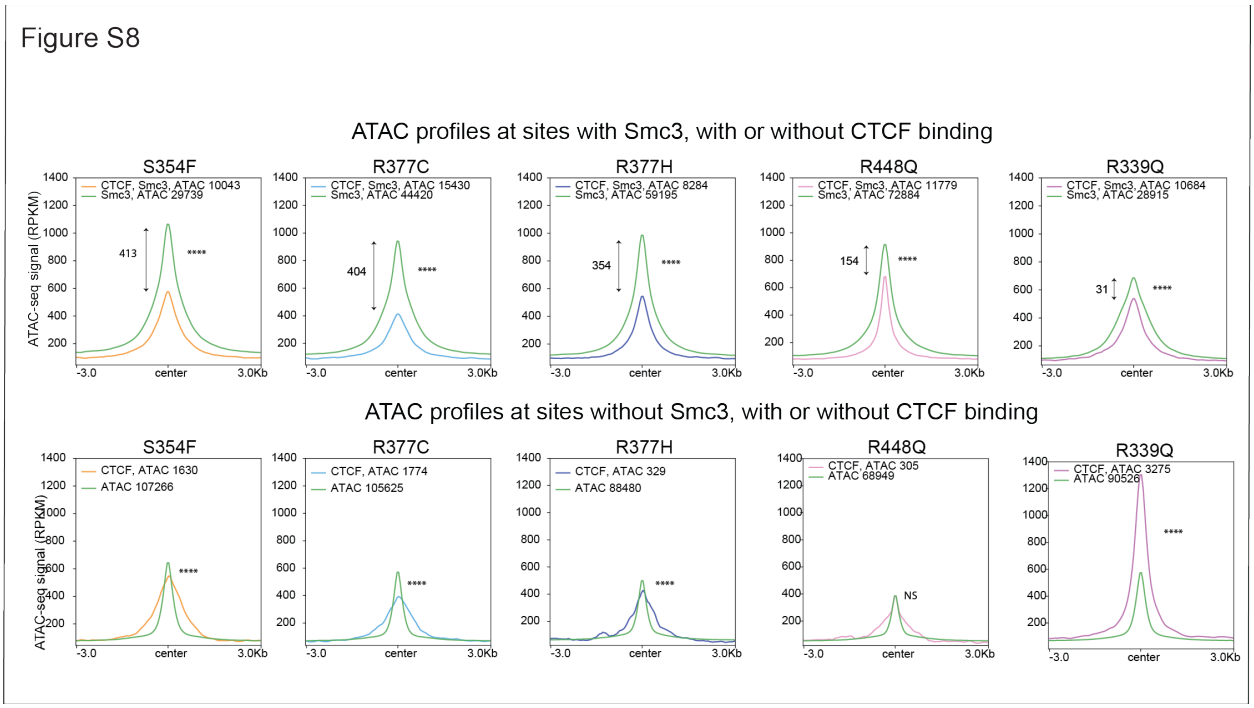


Figure S8: At CTCF WT and mutant sites, accessibility is more or less reduced in a cohesin-dependent manner**.** Profiles of ATAC-seq in WT and mutant CTCF expressing cells are shown for sites, with (color-coded for each mutant) or without (green) CTCF binding in the presence or absence of SMC3. The remaining mutants are shown in **Figure 2B**. Wilcoxon p-values were coded as follow: * $5x10^{-2}$-$5x10^{-3}$, ** $5x10^{-3}$-$5x10^{-4}$, *** $5x10^{-4}$-$5x10^{-5}$, **** $<5x10^{-5}$. This figure relates to **Figure 2.**

Figure S9

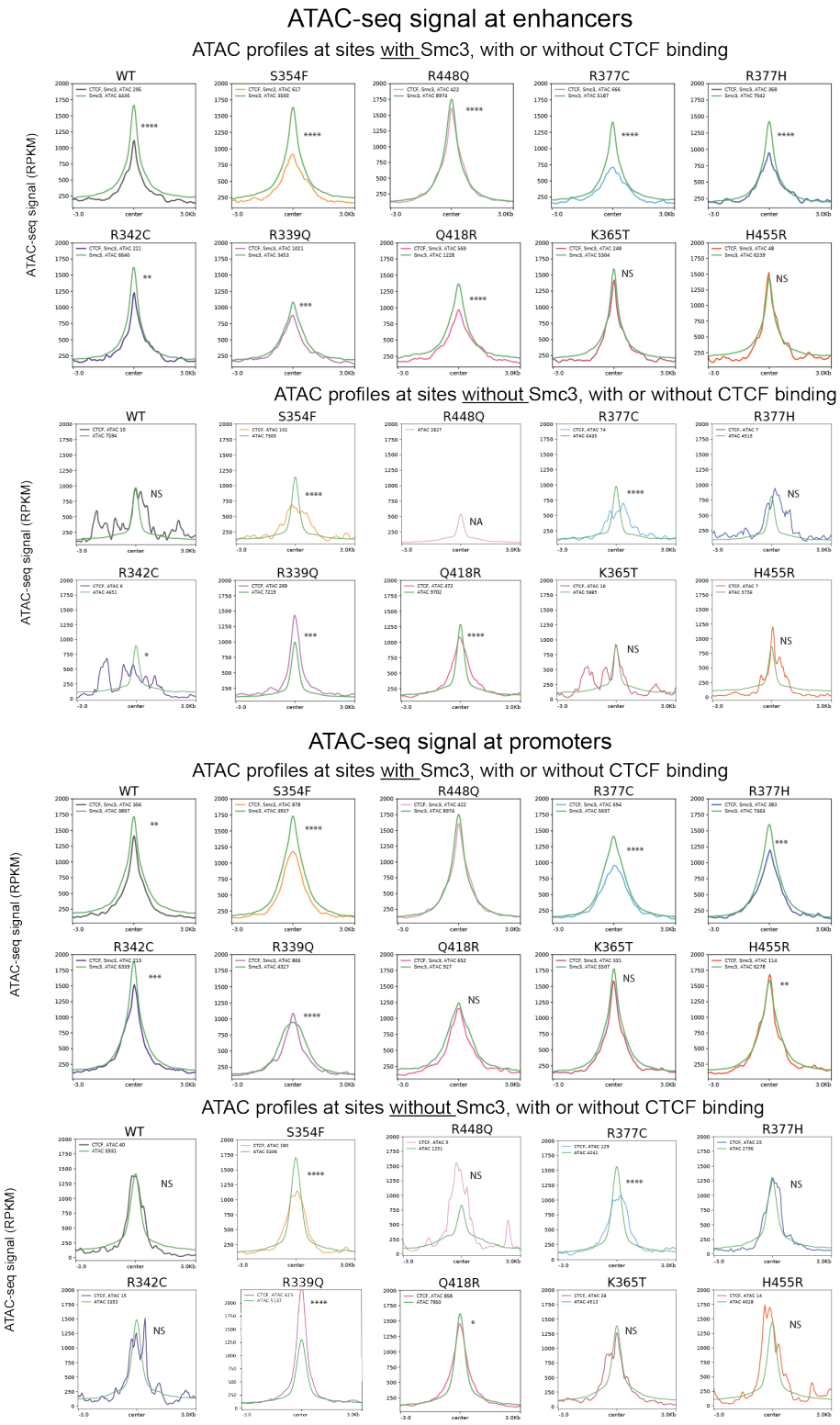

ATAC-seq signal at enhancers
ATAC profiles at sites with Smc3, with or without CTCF binding
ATAC profiles at sites without Smc3, with or without CTCF binding
ATAC-seq signal at promoters
ATAC profiles at sites with Smc3, with or without CTCF binding
ATAC profiles at sites without Smc3, with or without CTCF binding
ATAC-seq signal (RPKM)
WT
S354F
R448Q
R377C
R377H
R342C
R339Q
Q418R
K365T
H455R
NS
NA
-3.0
center
3.0Kb

Figure S9: At CTCF WT and mutant sites overlapping gene promoters and enhancers, accessibility is more or less reduced in a cohesin-dependent manner**.** Profiles of ATAC-seq in WT and mutant CTCF expressing cells are shown for sites, with (color-coded for each mutant) or without (green) CTCF binding in the presence or absence of SMC3 at gene enhancers (top) and gene promoters (bottom). In the absence of SMC3, for WT and most of the mutants, very few accessible CTCF peaks were observed at enhancers or promoters. Those data should be interpreted with caution. Wilcoxon p-values were coded as follow: * $5x10^{-2}$-$5x10^{-3}$, ** $5x10^{-3}$-$5x10^{-4}$, *** $5x10^{-4}$-$5x10^{-5}$, **** $<5x10^{-5}$. This figure relates to **Figure 2.**

Figure S10

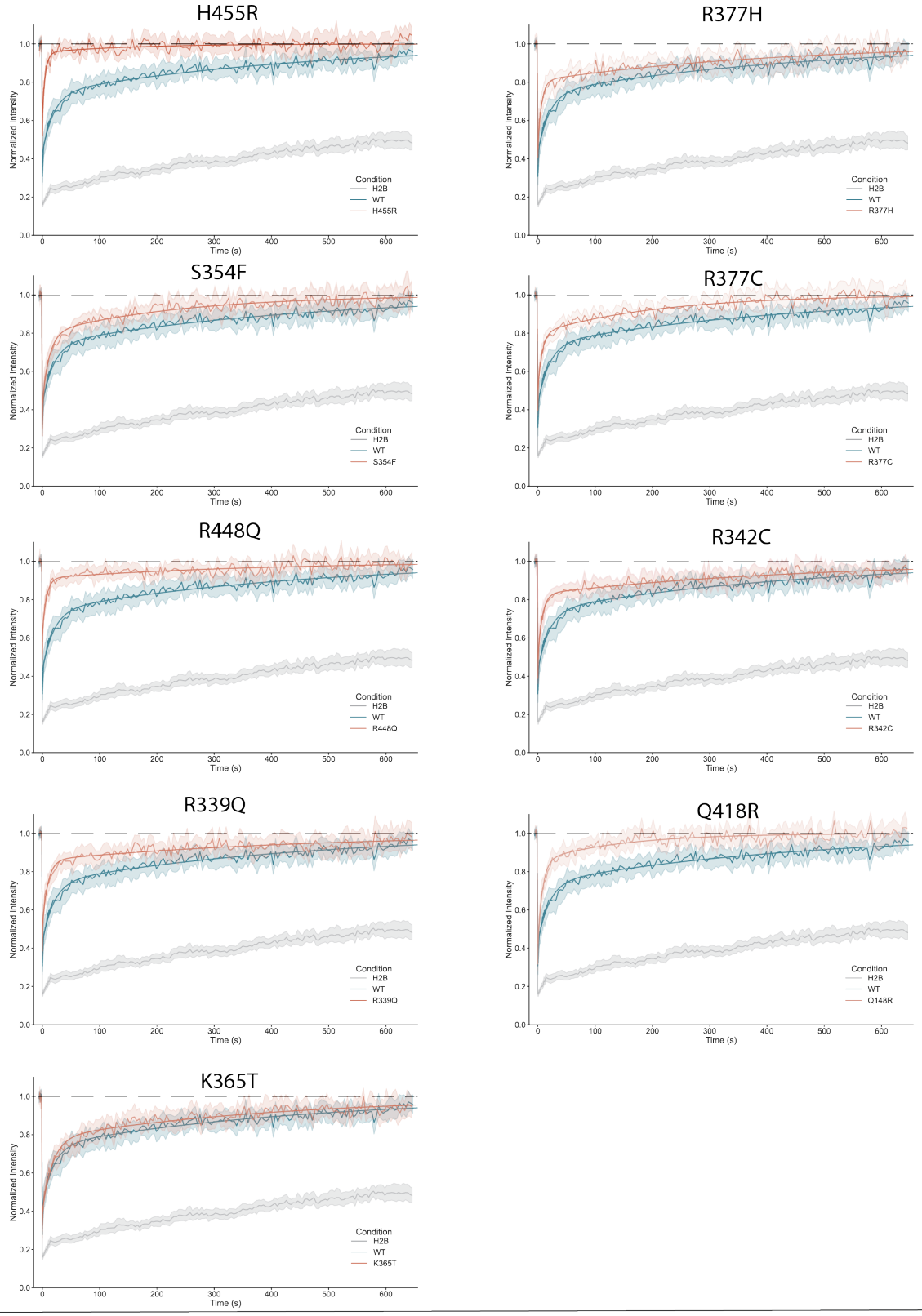
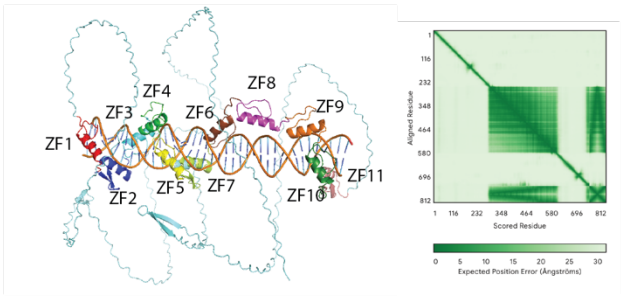

H455R
R377H
S354F
R377C
R448Q
R342C
R339Q
Q418R
K365T
Normalized Intensity
Time (s)
Condition
H2B
WT
H455R
R377H
S354F
R377C
R448Q
R342C
R339Q
Q148R
K365T

Figure S10. **Plots of FRAP dynamics for WT and mutant mRuby-CTCF for each pairwise comparison between mutant and WT**. The scatterplots show the average recovery across all movies analyzed, and the outlines give the 95% CI. The bold line shows the fitted model. This figure is related to **Figure 3**.

Figure S11

A

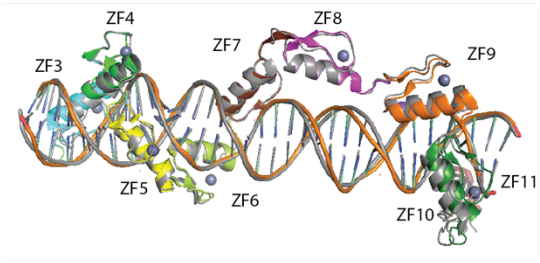
AlphaFold prediction CTCF, full protein



B

Experimental (8SSQ) versus alpha WT, ZF3-11

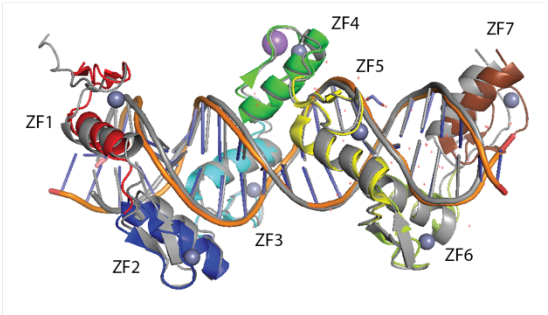
RSMD=1.2 Å



C

Experimental (8SSS) versus alpha WT, ZF1-7

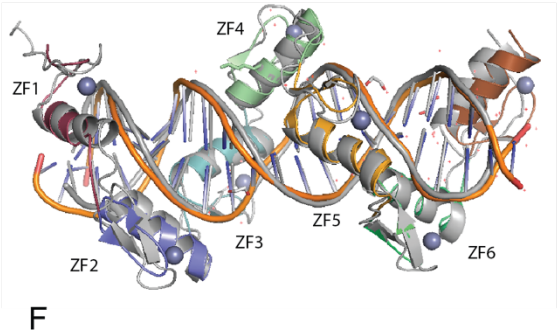
RSMD=1.5 Å



D

Experimental (8SST) versus alpha K365T, ZF1-7

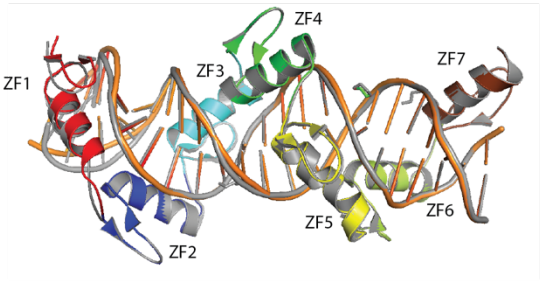
RSMD=1.4 Å



E

Experimental WT (8SSS) versus K365T (8SST), ZF1-7

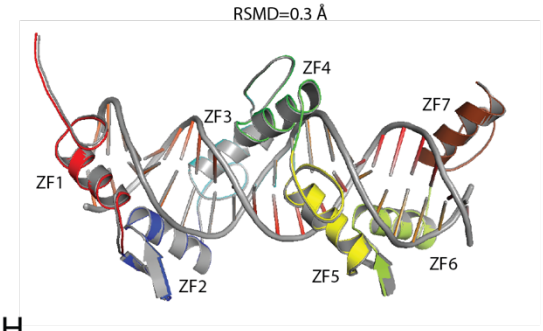
RSMD=0.3 Å



F

AlphaFold WT versus K365T, ZF1-7

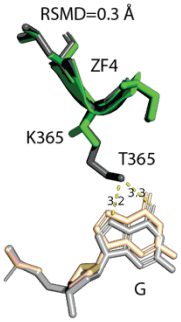
RSMD=0.3 Å



G

Experimental WT (8SSS) versus K365T (8SST), K365T

RSMD=0.3 Å



H

AlphaFold WT versus K365T, K365T

RSMD=0.2 Å

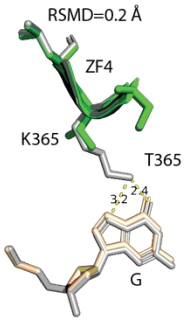


Figure S11: Comparison between the prediction the CTCF-DNA complex and the experimental structure. (**A-H**) Comparison of Alphafold3 prediction versus experimental crystal structure of WT and K365T-DNA complex. (**A**) Prediction of the full CTCF structure. On the right the pLDDT matrix shows a high confidence prediction of CTCF ZF1-11 and its positions relative to the DNA double helix. (**B**) Comparison between the experimental and predicted DNA-WT CTCF ZF3-11 complex. (**C**) Comparison between the experimental and predicted DNA-WT CTCF ZF1-7 structure. (**D**) Comparison between the experimental and predicted DNA-K365T mutant CTCF ZF1-7 structure. For **B-D**, the experimental structures are shown in grey and the prediction in color. Each ZF is color-coded. (**E**) Comparison of the DNA-CTCF complex between K365T mutant and WT from the experimental data and (**F**) between K365T mutant and WT from the AlphaFold3 prediction. For **E** and **F**, the WT structures are shown in grey and the mutant in color. (**G**) Comparison of the DNA-CTCF structure between the K365T mutant and WT showing the hydrogen bond between the mutated or WT residue and interacting DNA nucleotide from the experimental data and (**H**) from the prediction. The experimental crystal structures of the DNA-WT and DNA-K365T complexes were generated in Yang et al [29]. This figure is related to **Figure 3**.

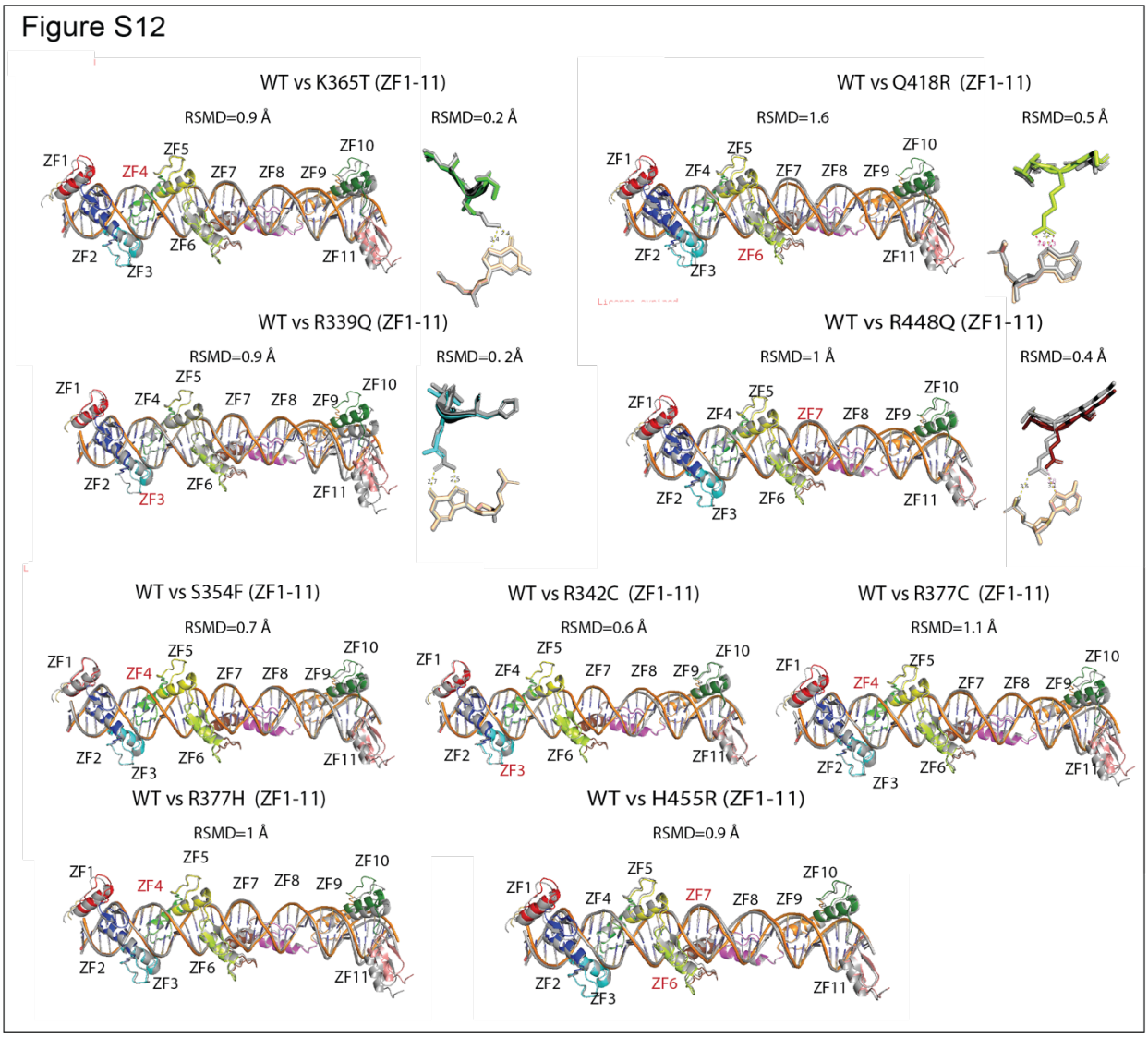
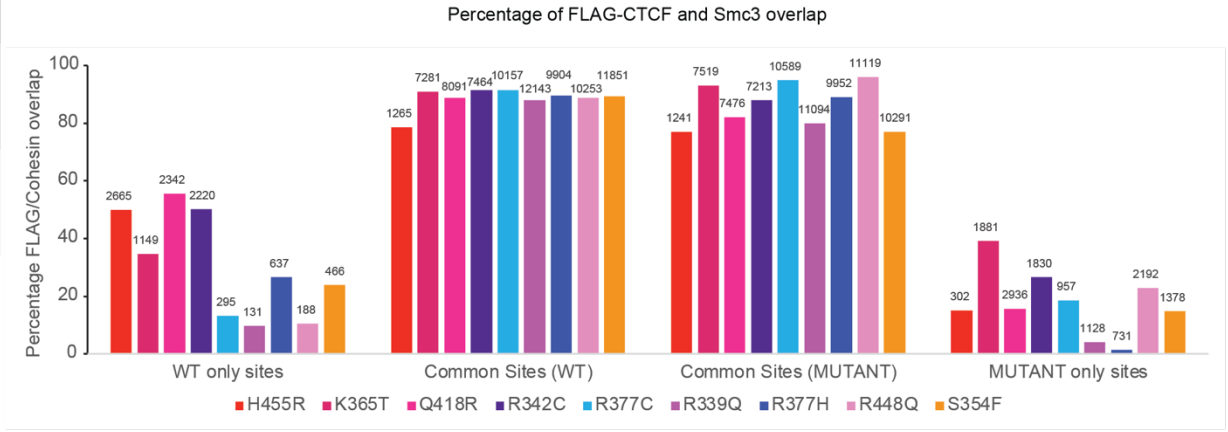


Figure S12: **Comparison of the predicted DNA-CTCF complexes between WT and CTCF mutants.** Comparison of the DNA-CTCF (ZF1-11) complexes between WT and each mutant CTCF predicted by AlphaFold3. For mutated residues making direct contact with the DNA, the Zoom-in shows the hydrogen bond between the mutated residue and interacting DNA nucleotide compared to WT. The WT structures are shown in grey and the mutant in color. Each ZF is color-coded. The ZFs bearing the mutations are labelled in red. This figure is related to **Figure 3**.

Figure S13

A



B

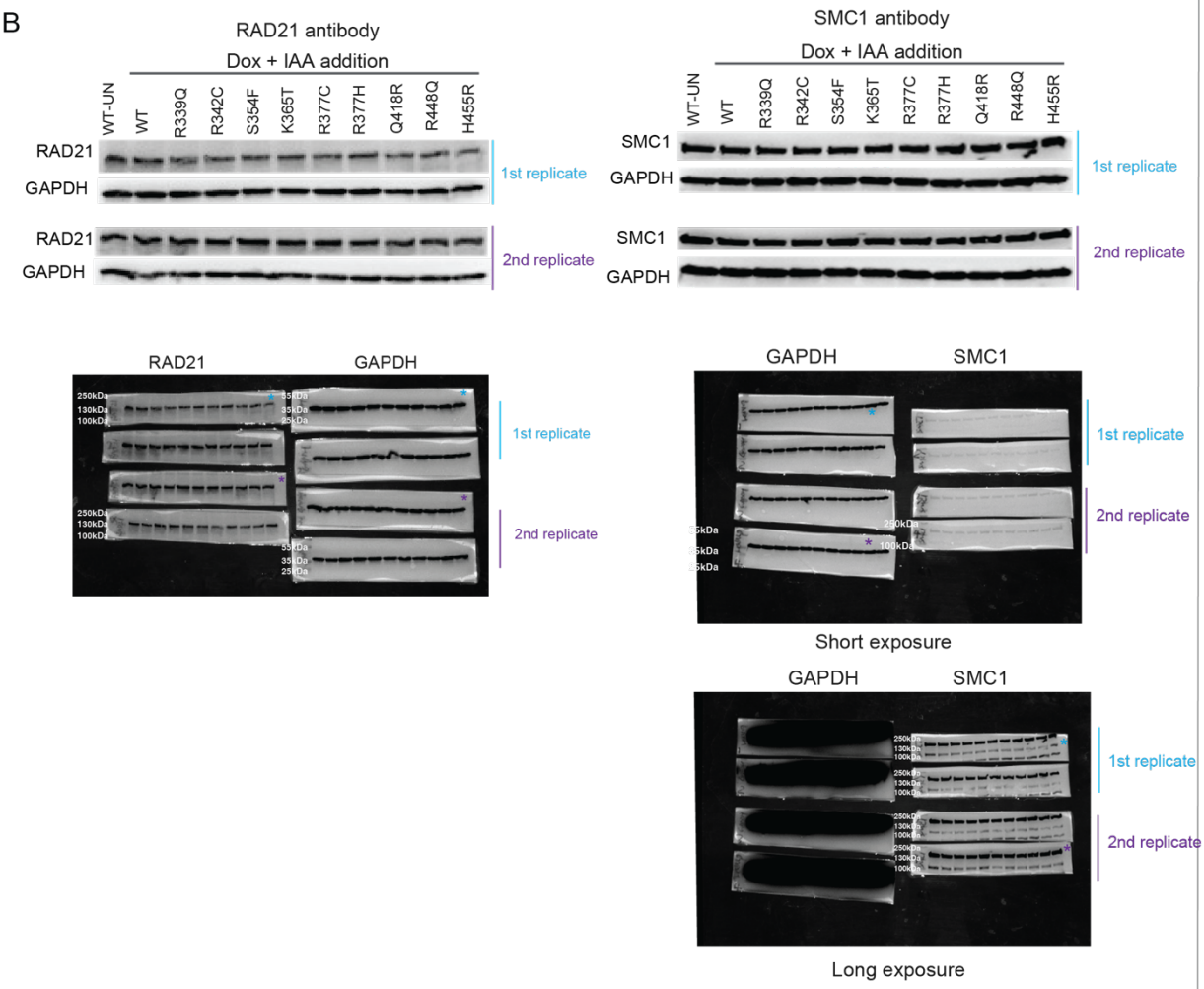


Figure S13: The ability of the CTCF mutants to block cohesin is mostly retained at common sites. Each mutation uniquely. (A) Bar graphs showing the percentage of CTCF-cohesin overlap for each mutant at WT only, common and mutant only sites. For common sites, the CTCF peaks were overlapped with either the WT or mutant SMC3 peaks. (**B**) Western blot showing the protein levels of 2 components of the cohesin complex (SMC1 and RAD21) in WT untreated (UN), IAA (CTCF degraded) and WT and CTCF mutant ID treated cells (cells expressing WT and mutant transgenes in the absence of endogenous CTCF). The uncropped pictures of the WB gels are shown below. This figure is related to **Figure 3**.

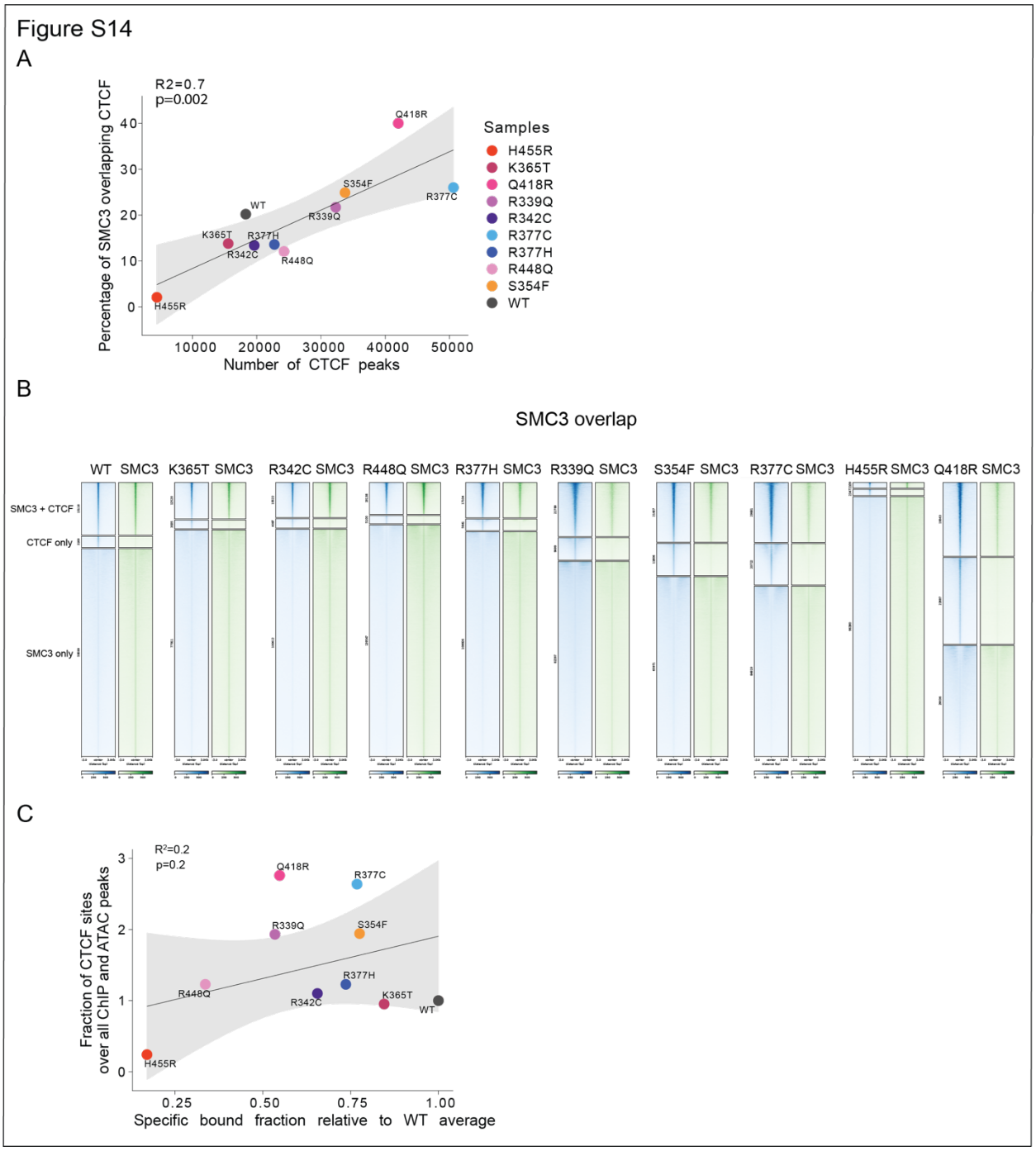
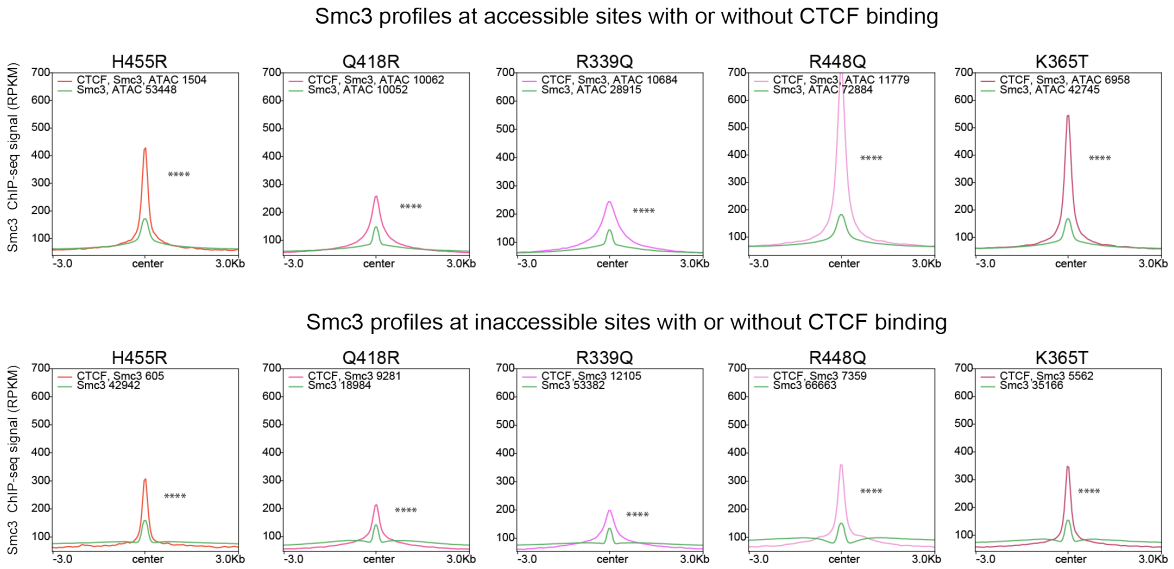


Figure S14: The percentage of SMC3 overlapping CTCF and the overall fraction of bound CTCF sites are less informative . (A) Correlation between the percentage of SMC3 overlapping CTCF peaks and the number of CTCF peaks. (**B**) Heatmaps of CTCF (blue) and SMC3 (green) in CTCF WT and mutant expressing cells at CTCF only, CTCF and SMC3, and SMC3 only sites. (**C**) Non-significant correlation between the chromatin bound mutant versus WT fraction detected by FRAP and the fraction of all CTCF binding sites relative to all FLAG and SMC3 ChIP-seq and ATAC-seq peaks. This figure is related to **Figure 3**.

Figure S15

A



B

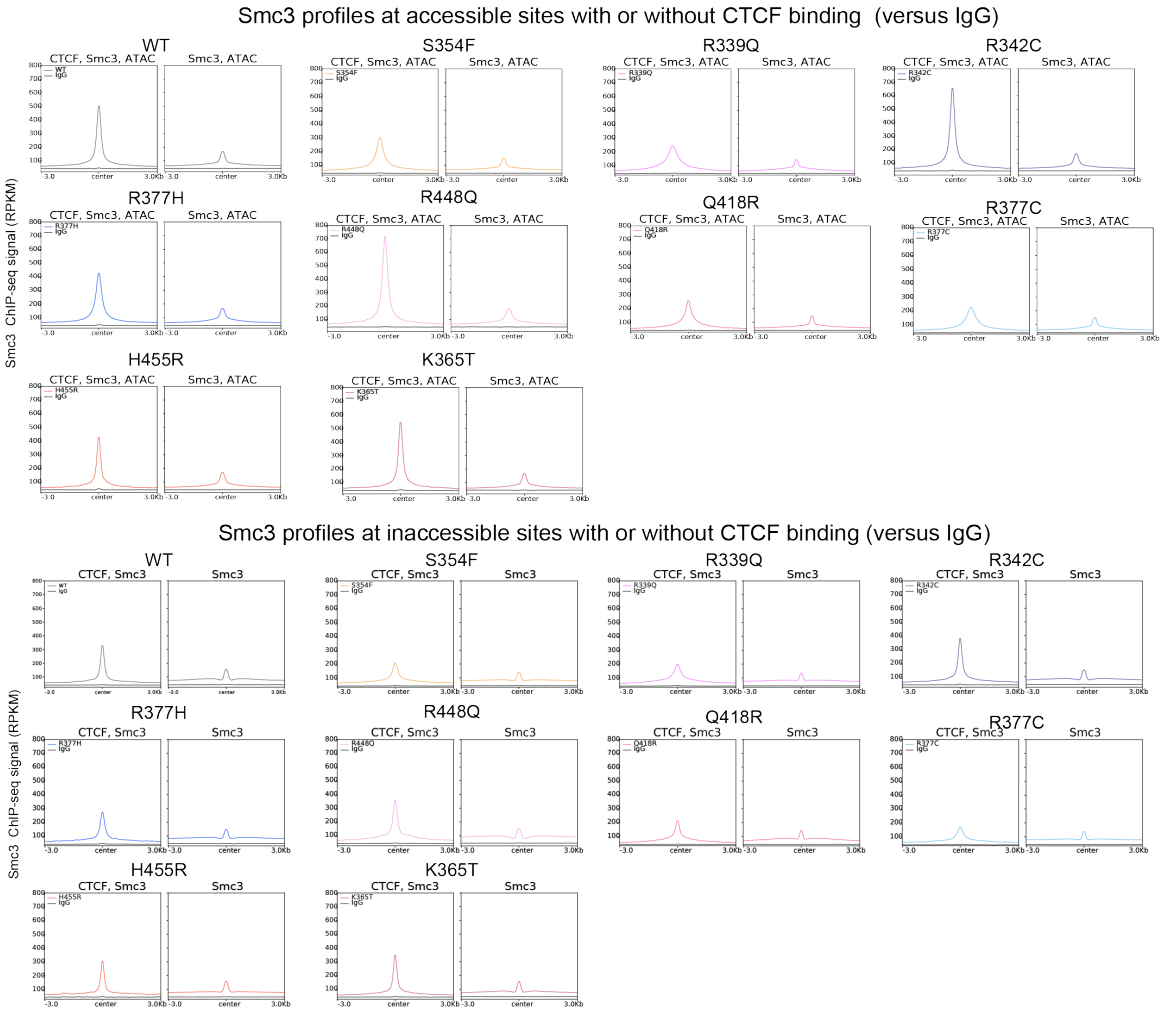


Figure S15: **The relationship between CTCF binding, cohesin overlap and accessibility. (A)** Profiles of SMC3 binding in WT and mutant CTCF expressing cells is shown for accessible (top) and inaccessible (bottom) sites with (color-coded for each mutant) or without (green) CTCF binding. The remaining mutants are shown in **Figure 4C**. Wilcoxon p-values were coded as follow: * $5x10^{-2}$-$5x10^{-3}$, ** $5x10^{-3}$-$5x10^{-4}$, *** $5x10^{-4}$-$5x10^{-5}$, **** $<5x10^{-5}$. (**B**) Profiles of SMC3 binding in WT and mutant CTCF expressing cells compared to IgG at accessible (top) and inaccessible (bottom) sites with or without CTCF (right and left profiles, respectively). Wilcoxon p-values were coded as follow: * $5x10^{-2}$-$5x10^{-3}$, ** $5x10^{-3}$-$5x10^{-4}$, *** $5x10^{-4}$-$5x10^{-5}$, **** $<5x10^{-5}$. This figure relates to **Figure 4**.

Figure S16

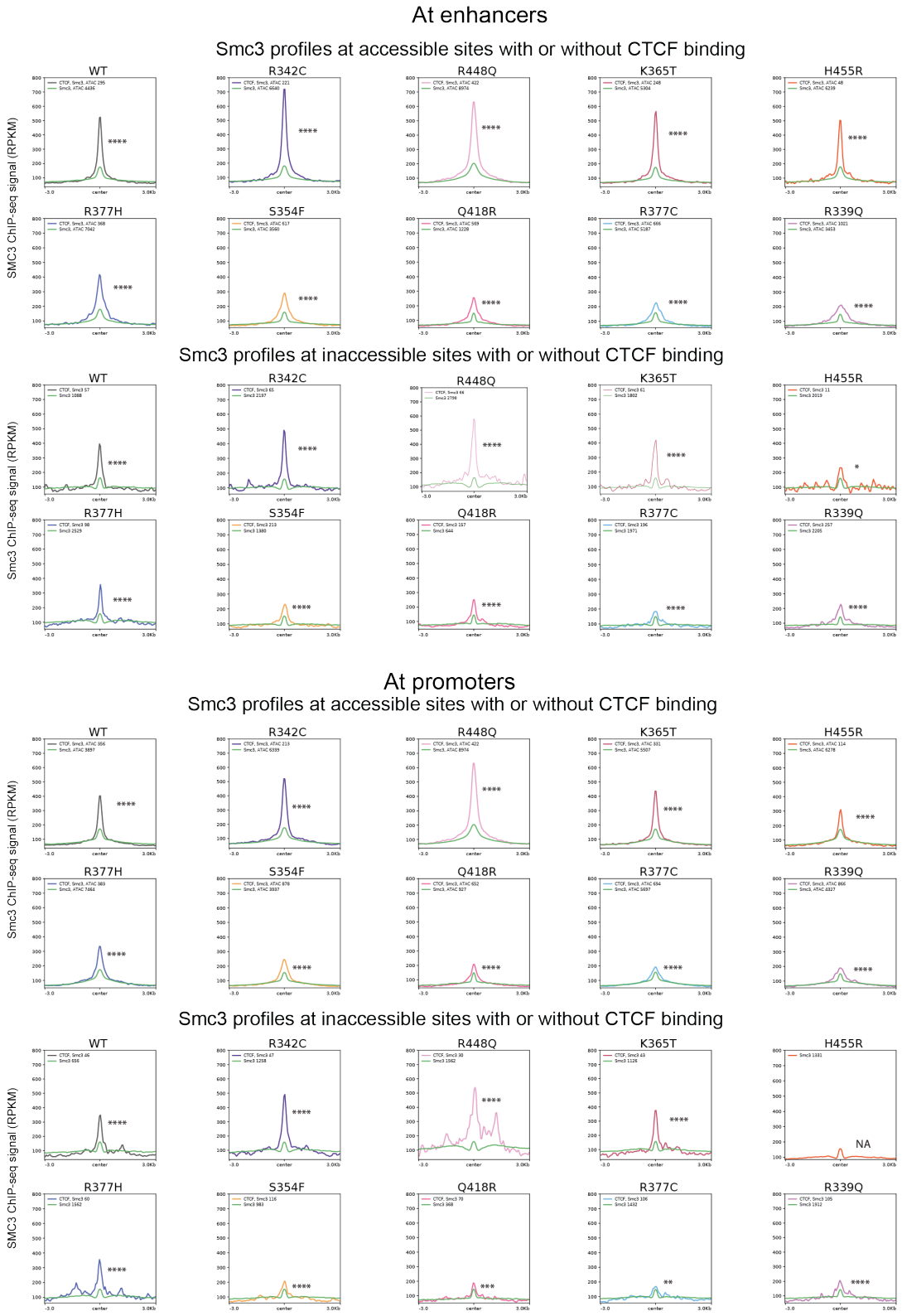
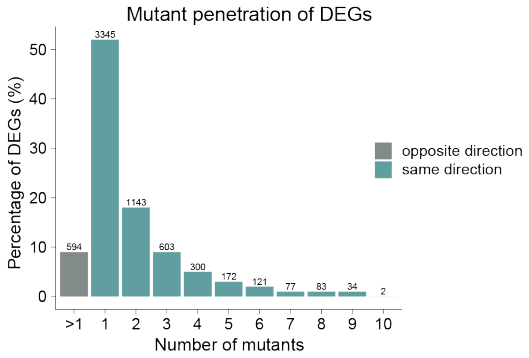

At enhancers
Smc3 profiles at accessible sites with or without CTCF binding
Smc3 profiles at inaccessible sites with or without CTCF binding
At promoters
Smc3 profiles at accessible sites with or without CTCF binding
Smc3 profiles at inaccessible sites with or without CTCF binding
SMC3 ChIP-seq signal (RPKM)
Smc3 ChIP-seq signal (RPKM)
Smc3 ChIP-seq signal (RPKM)
SMC3 ChIP-seq signal (RPKM)
WT
R342C
R448Q
K365T
H455R
R377H
S354F
Q418R
R377C
R339Q
NA

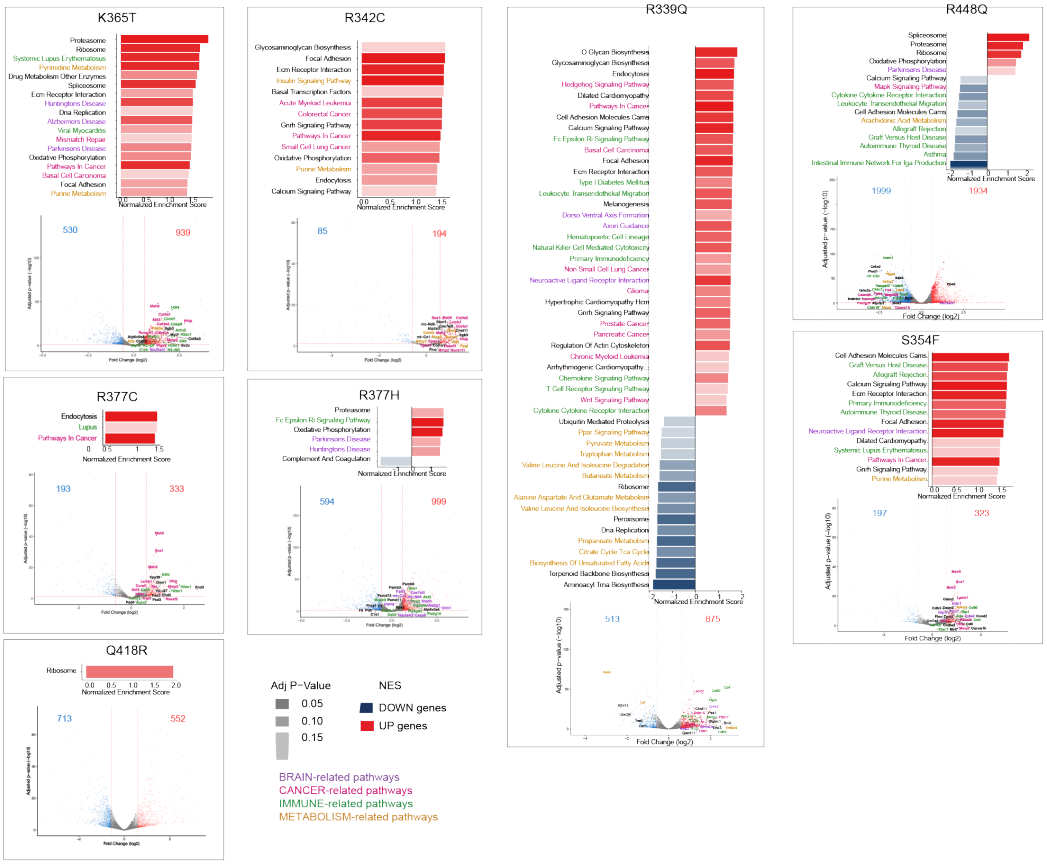
Figure S16: **The relationship between CTCF binding, cohesin overlap and accessibility at enhancers and promoters.** Profiles of SMC3 binding in WT and mutant CTCF expressing cells is shown for accessible and inaccessible CTCF binding at enhancers (top) and gene promoters (bottom) with (color-coded for each mutant) or without (green). Wilcoxon p-values were coded as follow: * $5x10^{-2}$-$5x10^{-3}$, ** $5x10^{-3}$-$5x10^{-4}$, *** $5x10^{-4}$-$5x10^{-5}$, **** $<5x10^{-5}$. This figure relates to **Figure 4**.

Figure S17

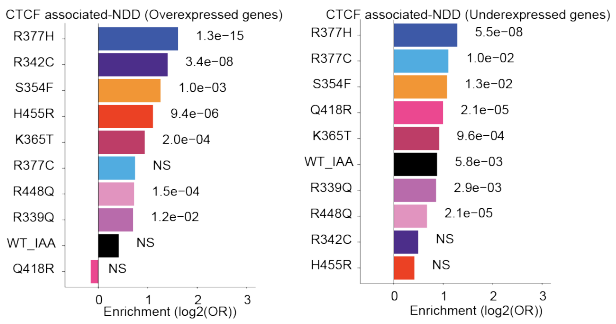
A



B



C



D

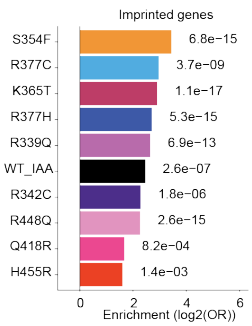


Figure S17: **CTCF mutations alter gene expression.** (**A**) Bar graph showing the percentage of DEGs found in 1 to 10 mutants, including IAA. The list of genes is reported in **Table S2**. (**B**) Gene set enrichment analysis of DEGs. All other mutants are shown in **Figure 5B**. For each condition, bar graphs show significantly enriched KEGG pathways. The volcano plots below highlight the DEGs belonging to these enriched pathways with genes outside of these pathways shown in black. Brain related pathways are shown in purple, cancer in pink, immune in green and metabolism in mustard. (**C**) Enrichment of the DEGs in over- and under-expressed genes reported in CTCF associated human neuro-developmental disorders [S4]**.** (**D**) Enrichment of the DEGs in imprinted genes reported in Geneimprint (https://www.geneimprint.com). This figure relates to **Figure 5**.

Figure S18

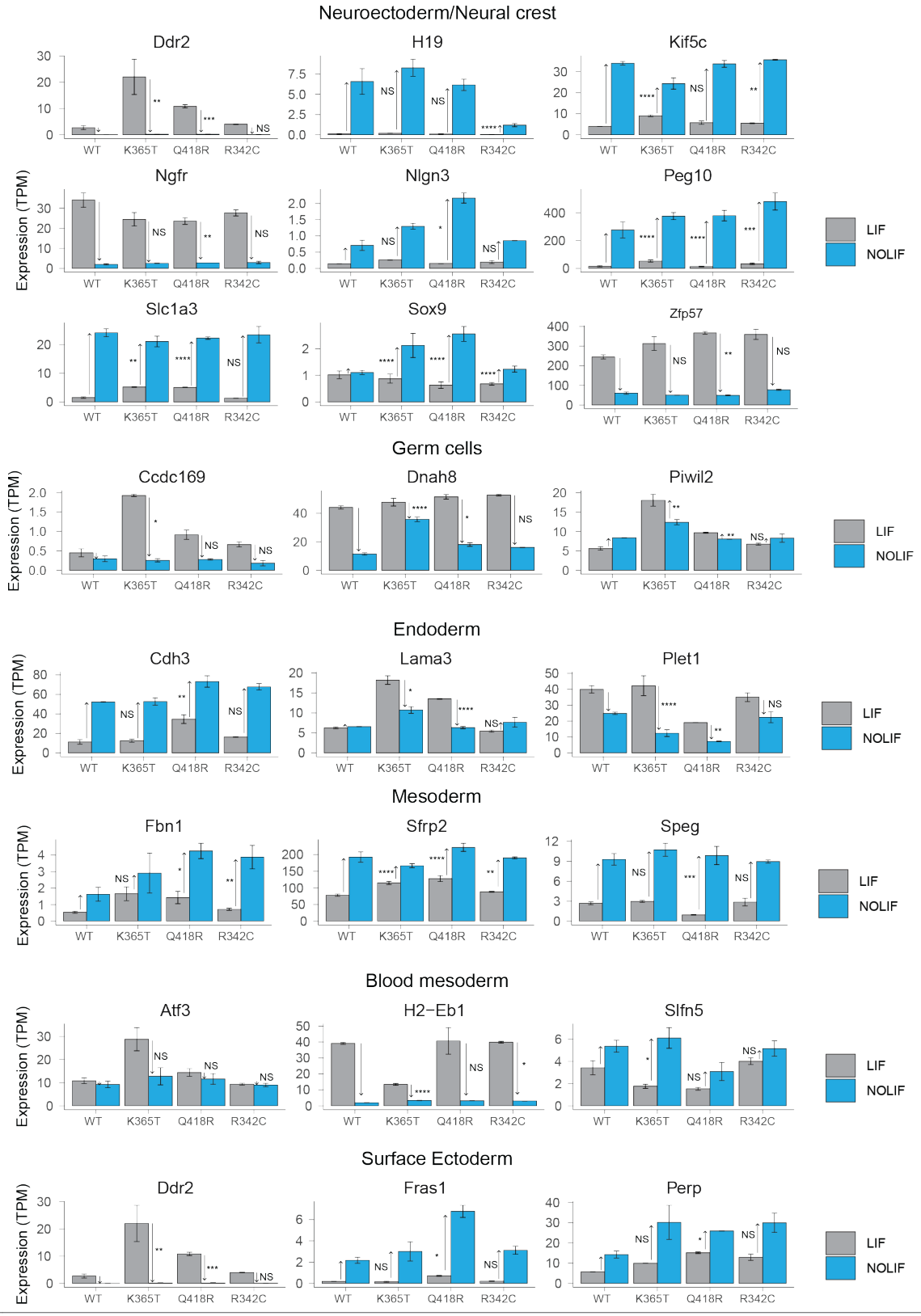

Neuroectoderm/Neural crest
Ddr2
H19
Kif5c
Ngfr
Nlgn3
Peg10
Slc1a3
Sox9
Zfp57
Germ cells
Ccdc169
Dnah8
Piwil2
Endoderm
Cdh3
Lama3
Plet1
Mesoderm
Fbn1
Sfrp2
Speg
Blood mesoderm
Atf3
H2−Eb1
Slfn5
Surface Ectoderm
Ddr2
Fras1
Perp
Expression (TPM)
WT
K365T
Q418R
R342C
LIF
NOLIF

Figure S18: **Examples of genes implicated in altered differentiation trajectories in WT versus CTCF mutant expressing cells.** Bar graphs showing the comparison of WT versus mutant CTCF expression levels (TPM) in LIF and no LIF medium. The p-values assessing the significance of the longitudinal differences between mutant and WT were computed using linear regressions at the gene level. p-values are coded as followed *: $5x10^{-2}$-$5x10^{-3}$, **$5x10^{-3}$-$5x10^{-4}$, *** $5x10^{-4}$-$5x10^{-5}$, ****<$5x10^{-5}$. This figure relates to **Figure 5**.

Figure S19

A

Log2 Fold chang bindinding score (Mutant vs WT)

TF	Motif	IAA	H455R	Q418R	K365T	R342C	S354F	R448Q	R339Q	R377C	R377H
MBD2		0.1	0.2	0.2	0.2	0.3	0.5	0.5	0.4	0.4	0.3
CTCFL		-0.2	-0.1	0.0	0.2	0.1	0.0	0.5	0.0	0.2	0.1
NRF1		0.2	0.2	0.1	0.3	0.2	0.3	0.5	0.3	0.3	0.3
NFYB		0.2	0.0	0.2	0.3	0.1	0.1	0.4	0.1	0.1	0.5
CTCF		-0.2	-0.2	0.0	0.2	0.1	0.0	0.4	0.0	0.1	0.0
NFYC		0.2	0.0	0.2	0.3	0.1	0.1	0.4	0.1	0.1	0.4
KAISO		0.1	0.2	0.2	0.3	0.3	0.3	0.4	0.3	0.3	0.3
NFYA		0.2	0.0	0.2	0.3	0.1	0.1	0.4	0.1	0.1	0.4
E2F2		0.1	0.2	0.1	0.2	0.3	0.3	0.4	0.3	0.3	0.2
MECP2		0.1	0.1	0.1	0.2	0.2	0.3	0.4	0.2	0.3	0.2
SP2		0.1	0.2	0.1	0.2	0.3	0.3	0.4	0.3	0.3	0.2
PBX3		0.2	0.0	0.2	0.2	0.1	0.0	0.3	0.1	0.1	0.3
E2F1		0.1	0.2	0.1	0.2	0.2	0.3	0.3	0.3	0.3	0.2
AP2B		0.1	0.2	0.1	0.1	0.2	0.3	0.3	0.2	0.2	0.2
E2F4		0.2	0.2	0.2	0.3	0.2	0.3	0.3	0.2	0.3	0.2

B

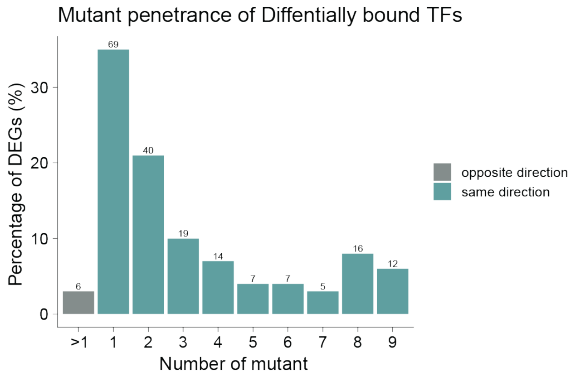


Figure S19: **CTCF mutations alter TF binding. (A)** Top 15 predicted differentially bound TFs. The TFs were sorted by the maximum absolute log2 fold change of binding score between mutant versus WT observed across mutants. The total number of sites and p-value for these TFs, as well as the full list of differentially bound TFs for each mutant are reported in **Table S4A**. (**B**) Bar graph showing the percentage of differentially bound TFs found in 1 to 9 mutants including IAA (none were found in all the conditions). The list of significant differentially bound TFs is reported in **Table S4A**. This figure relates to **Figure 5**.

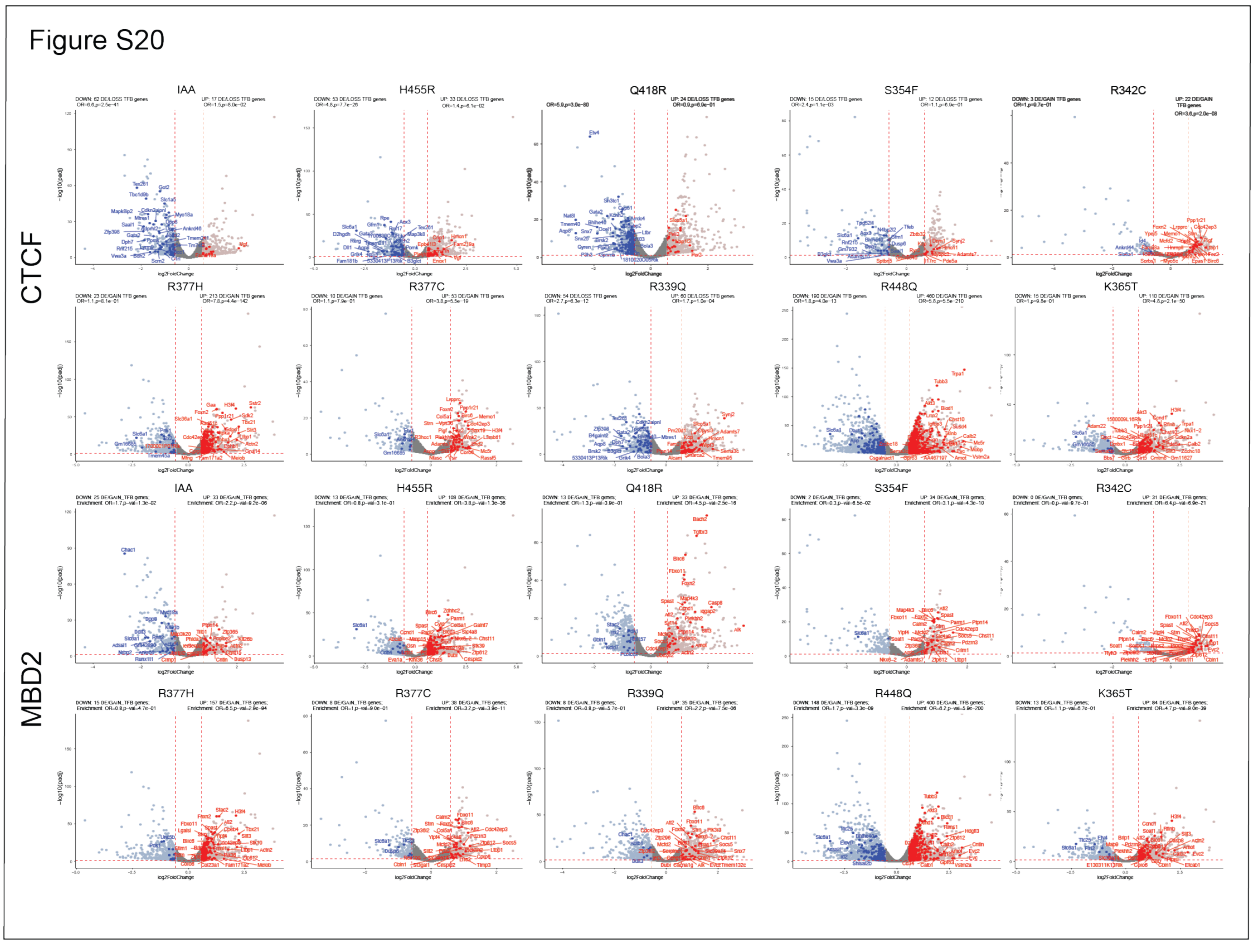
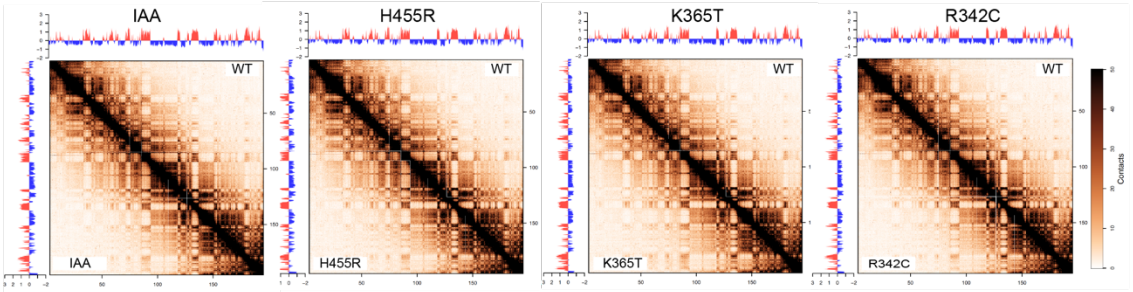


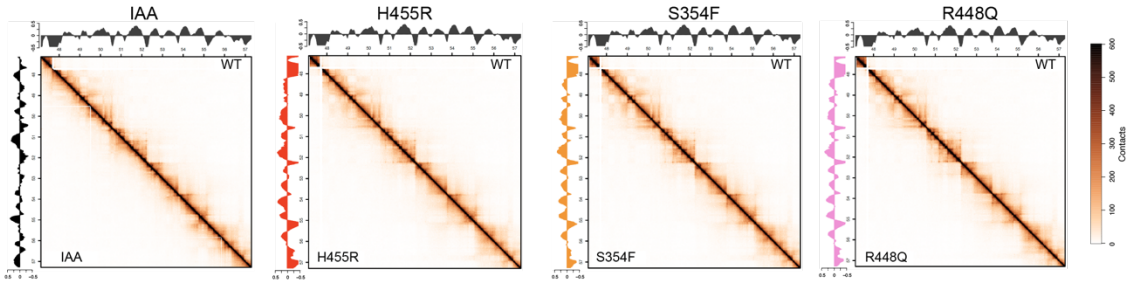
Figure S20: **The DEGs are enriched in the target genes of differentially bound TF.** Volcano plots show examples of two differentially bound TFS (CTCF, MBD2) overlapping the promoters of DEGs in CTCF mutant expressing cells (ID). The remaining mutant comparisons are shown in **Figure 5E**. The enrichment of the TF target genes among the up- and down-regulated genes are reported on top on the volcanos (Odds Ratios (ORs), and p-values). Enrichment for all the differentially bound TFs and their differentially expressed target genes are reported in **Table S4B** and **S4C**. This figure relates to **Figure 5**.

Figure S21

A



B



C

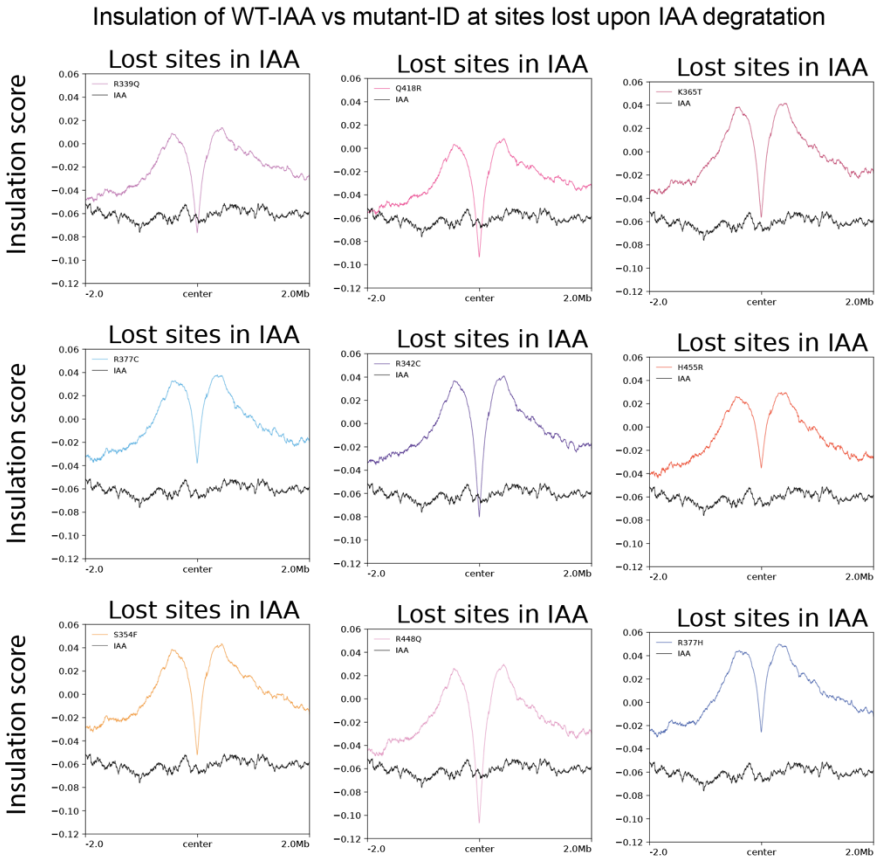
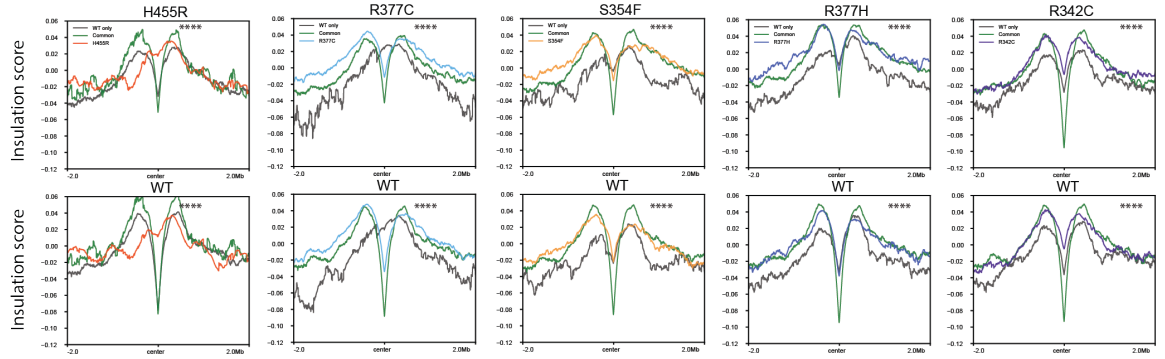


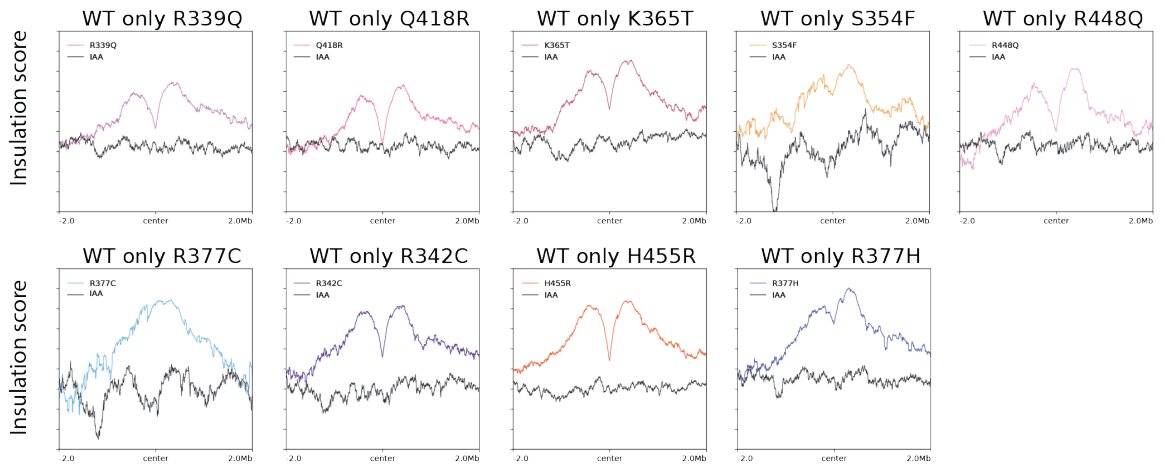
Figure S21: **CTCF mutations alter insulation and chromatin interactivity.** (**A**) Absolute Hi-C matrices of chromosome 1 using a 100kb resolution for IAA and 3 mutants (bottom left) versus WT (top right) showing the minimal changes in chromatin compartmentalization. The compartment scores are shown on the left for the mutant and top for WT. (**B**) Absolute Hi-C matrix for IAA and the mutant shown in Figure **6E** compared to WT. The insulation scores are shown on the left for the mutant and top for WT. (**C**) Profiles comparing the insulation score in WT-IAA versus mutant-ID at lost CTCF sites upon IAA degradation. This figure relates to **Figure 6**.

Figure S22

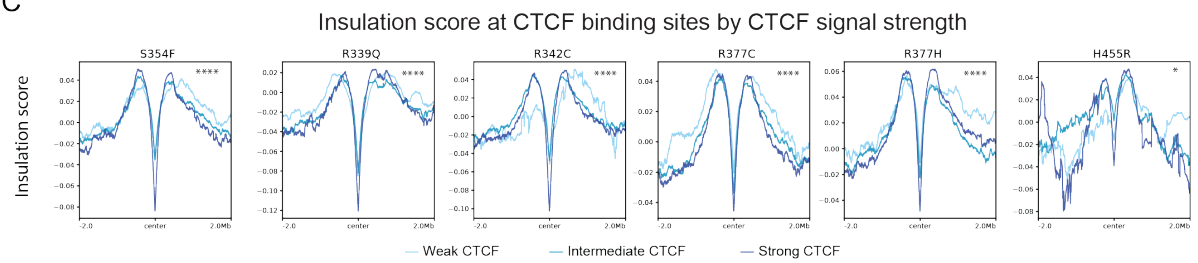
A



B



C



D

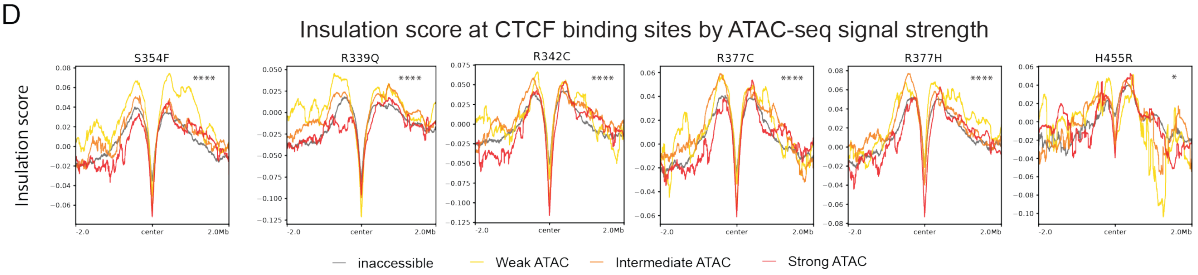


Figure S22: **Insulation is a function of CTCF and ATAC-seq signal strength .** (**A**) Profiles show the aggregated insulation score at WT only, common and mutant only binding sites for four mutants. The profiles for the remaining mutants are shown in **Figure 6F.** (**B**) Profiles comparing the insulation score in mutant-ID versus IAA conditions at WT only sites. (**C**) Profiles show the aggregated insulation score at CTCF binding sites with weak, intermediate and strong CTCF signals for different mutants. The profiles for the remaining mutants are shown in **Figure 6G.** (**D**) Profiles show the aggregated insulation score at CTCF binding sites with no, weak, intermediate and strong ATAC signals for different mutants. The profiles for the remaining mutants are shown in **Figure 6H.** For (**C**) and (**D**)**,** Kruskal-Wallis p-values were coded as follow: * $5x10^{-2}$-$5x10^{-3}$, ** $5x10^{-3}$-$5x10^{-4}$, *** $5x10^{-4}$-$5x10^{-5}$, **** $<5x10^{-5}$. This figure relates to **Figure 6.**

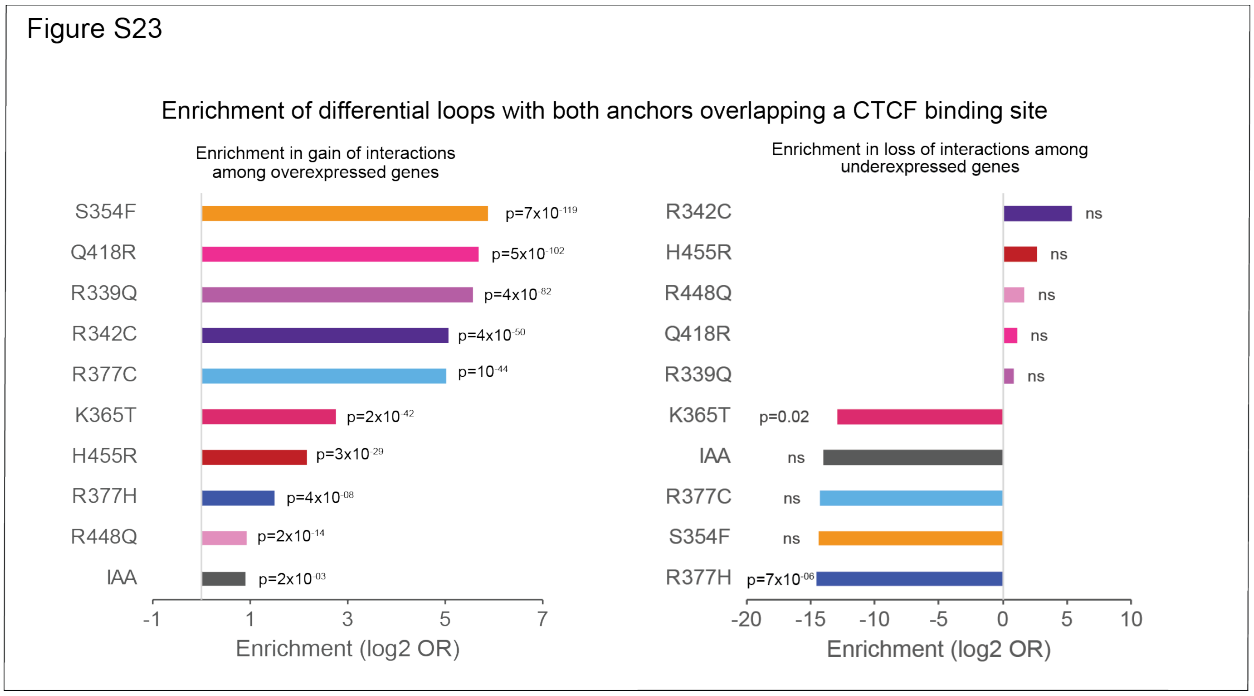
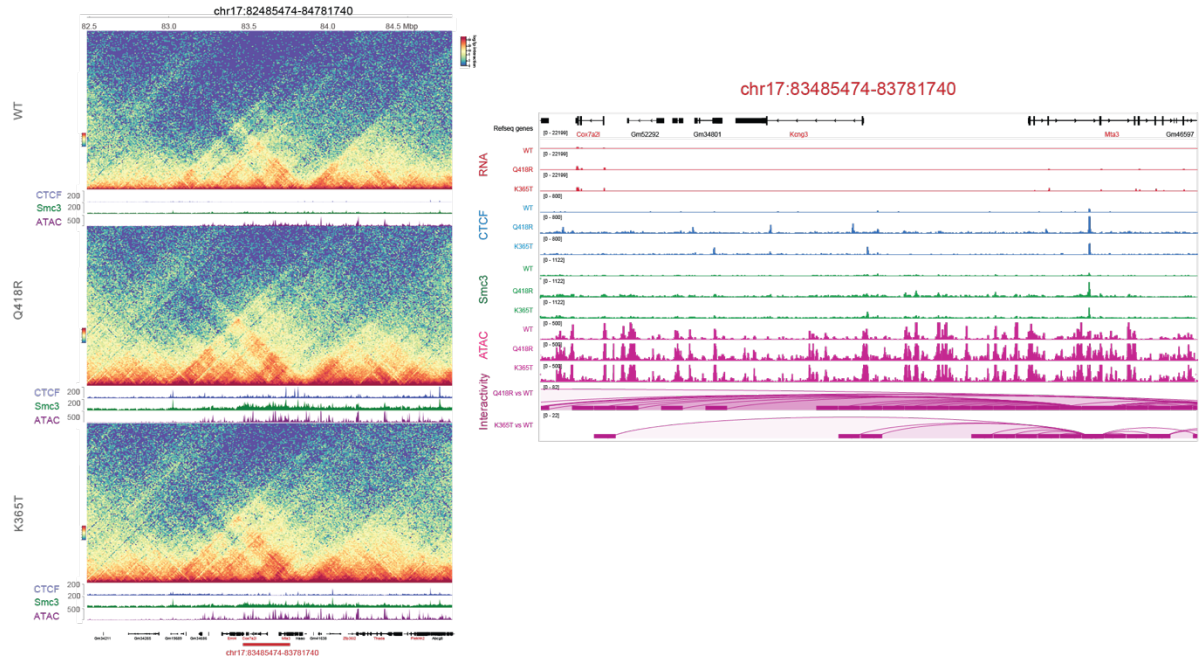


Figure S23: Differentially expressed genes are enriched in gain of loops with a CTCF binding site at both anchors. Bar graphs showing the enrichment (calculated over WT ID) of over-expressed (top) or under-expressed (bottom) genes in gained (top) or lost (bottom) loops with both anchors overlapping a CTCF binding sites and at least one anchor overlapping the promoter of DEGs in IAA and CTCF mutants. This figure relates to **Figure 7.**

Figure S24

A



B

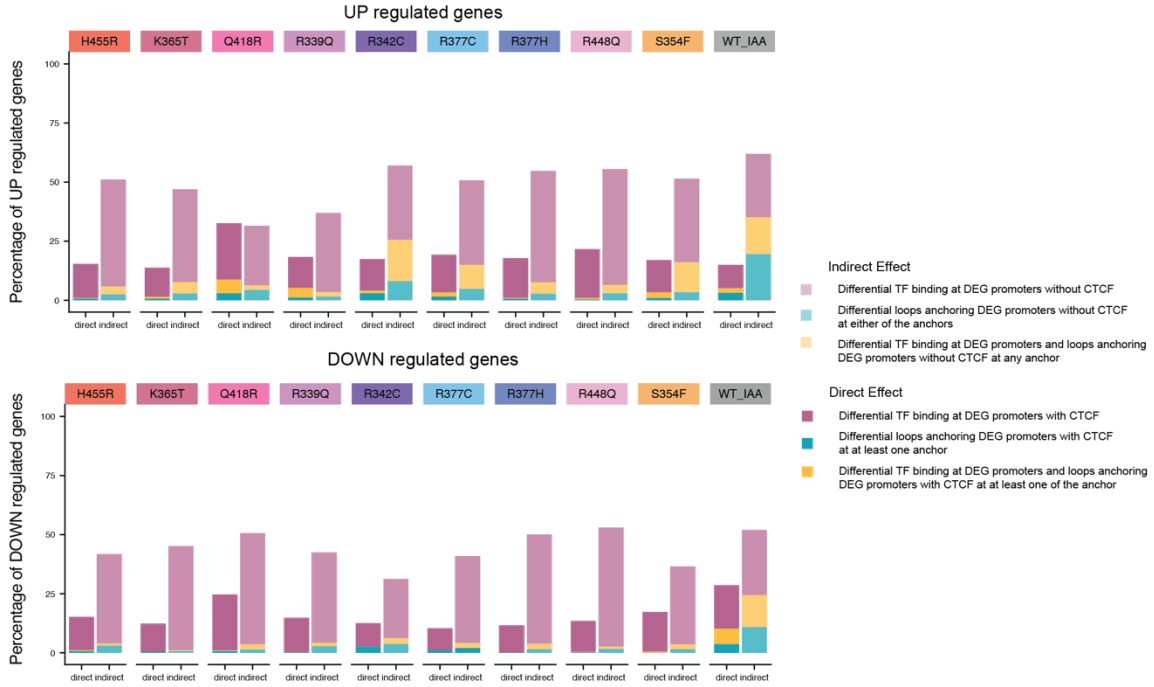


Figure S24: Indirect effects resulting from changes in TF binding at gene promoters account for the majority of CTCF mutation-mediated changes in expression. (**A**) Example of a locus on chromosome 7 (red rectangles) with a direct effect of gain in CTCF binding and chromatin interactivity. The left panel shows the Hi-C interaction matrices in WT, Q418R and K365T with both gain of intra- and inter-TAD interactions in the mutant compared to WT. The left panels show the zoom-in tracks of CTCF (blue) SMC3 (green), ATAC (red) and significant differential chromatin loops (purple) with one anchor overlapping the differential CTCF binding sites. Overexpressed genes are highlighted in red. Of note, only gained interactions were detected at these loci. This analysis was performed at 10 kb resolution. (**B**) Bar graphs showing the percentage of UP and DOWN DEGs associated with differential TF/CTCF binding, differential loops or both, stratified by direct and indirect effect. A direct effect was defined as predicted differential CTCF binding at the gene promoter and/or differential loops anchoring the DEG promoters with a CTCF peak at either anchor. This figure relates to **Figure 7.**

REFERENCES

S1. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov *2*, 401-404. 10.1158/2159-8290.CD-12-0095.

S2. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res *47*, D941-D947. 10.1093/nar/gky1015.

S3. Price, E., Fedida, L.M., Pugacheva, E.M., Ji, Y.J., Loukinov, D., and Lobanenkov, V.V. (2023). An updated catalog of CTCF variants associated with neurodevelopmental disorder phenotypes. Front Mol Neurosci *16*, 1185796. 10.3389/fnmol.2023.1185796.

S4. Konrad, E.D.H., Nardini, N., Caliebe, A., Nagel, I., Young, D., Horvath, G., Santoro, S.L., Shuss, C., Ziegler, A., Bonneau, D., et al. (2019). CTCF variants in 39 individuals with a variable neurodevelopmental disorder broaden the mutational and clinical spectrum. Genet Med *21*, 2723-2733. 10.1038/s41436-019-0585-z.