



Published in final edited form as:

*Genomics*. 2022 May ; 114(3): 110350. doi:10.1016/j.ygeno.2022.110350.

## Cross-talk between enhancers, structural elements and activating transcription factors maintains the 3D architecture and expression of the *CFTR* gene

Shiyi Yin<sup>1</sup>,

Monali Nandy Mazumdar<sup>1</sup>,

Alekh Paranjapye<sup>1</sup>,

Ann Harris<sup>\*</sup>

Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH 44106, USA

### Abstract

Robust protocols to examine 3D chromatin structure have greatly advanced knowledge of gene regulatory mechanisms. Here we focus on the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, which provides a paradigm for validating models of gene regulation built upon genome-wide analysis. We examine the mechanisms by which multiple *cis*-regulatory elements (CREs) at the *CFTR* gene coordinate its expression in intestinal epithelial cells. Using CRISPR/Cas9 to remove CREs, individually and in tandem, followed by assays of gene expression and higher-order chromatin structure (4C-seq), we reveal the cross-talk and dependency of two cell-specific intronic enhancers. The results suggest a mechanism whereby the locus responds when CREs are lost, which may involve activating transcription factors such as FOXA2. Also, by removing the 5' topologically-associating domain (TAD) boundary, we illustrate its impact on *CFTR* gene expression and architecture. These data suggest a multi-layered regulatory hierarchy that is highly sensitive to perturbations.

### Keywords

*Cis*-regulatory elements; 3D chromatin architecture; Enhancers; Topologically associated domain (TAD); boundaries; Transcription factors; Cystic fibrosis transmembrane conductance; regulator (*CFTR*)

---

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

<sup>\*</sup>Corresponding author. ann.harris@case.edu (A. Harris).

<sup>1</sup>These authors contributed equally.

CRedit authorship contribution statement

**Shiyi Yin:** Conceptualization, Investigation, Validation, Formal analysis, Data curation. **Monali Nandy Mazumdar:** Investigation, Validation, Formal analysis, Writing – review & editing. **Alekh Paranjapye:** Investigation, Validation, Formal analysis, Data curation.

**Ann Harris:** Conceptualization, Formal analysis, Validation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2022.110350>.

## 1. Introduction

The direct interaction of cell-type specific enhancers with the gene promoter is known to be critical for expression of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene [1–3]. However, our understanding of the mechanisms underlying this interaction is incomplete: i) what drives the coordination of multiple *cis*-regulatory elements (CREs) in a cell, ii) what is the contribution of chromatin structural elements and of cell-selective transcription factors? To address these key questions we used a combination of CRISPR/Cas9 manipulation of CREs, siRNA-mediated depletion of TFs and examination of higher order chromatin structure and gene expression. We focus specifically on intronic enhancers in *CFTR* that are utilized in intestinal epithelial cells, and on the transcription factors (TFs) that bind to them.

Mutations in *CFTR* cause cystic fibrosis (CF) and several related phenotypes [4]. Current attempts to repair the gene in individuals with CF who lack CFTR protein, and thus are not responsive to current pharmacological therapies [5], are focused on gene editing or gene replacement. For both approaches a detailed knowledge of *CFTR* regulation is pertinent. The *CFTR* locus is organized within a topologically associating domain (TAD) flanked by sites of CCCTC-binding factor (CTCF) and cohesin occupancy at –80.1 kb upstream and +48.9 kb downstream of the gene (the TAD boundaries) [1,2,6–8]. Within this TAD, specific CREs play critical roles in the regulation of *CFTR* expression. We previously identified and characterized multiple cell-type specific CREs at the locus, which are summarized in Suppl. Table 1 according to genomic location (hg19) and common name according to Legacy and Standard (Refseq) nomenclature. These CREs include both structural and enhancer elements in addition to some sites with as yet uncharacterized functions. Critical among the structural elements are two sites at –20.9 kb [9] and +6.8 kb [10] which have classical enhancer-blocking insulator activity (insulators), that is the ability to prevent enhancer-promoter interactions in an in vitro assay [11], and that bind CTCF in a cell-type selective manner. An additional insulator element 3' to the gene at +15.6 kb does not bind CTCF [12,13].

Here we focus on enhancers associated with DNase I hypersensitive sites (DHS) in intron 1 (int1) and intron 11 (Legacy; int11) of *CFTR*, which were previously studied in detail individually [1,14–18]. We now address their coordinated mechanism of interaction in the same cell type. Int1 (at 185 + 10 kb, where 185 is the last base in *CFTR* exon 1) was shown to be an important *CFTR* enhancer in a transgenic mouse model, with the human *CFTR* gene on a yeast artificial chromosome (YAC). Deletion of int1 from the YAC reduced human *CFTR* expression by 60% [15]. Similarly, CRISPR/Cas9-mediated removal of the int11 enhancer from its endogenous location in Caco2 colon carcinoma cells reduced *CFTR* expression by ~80% [2]. Together with other CREs across the *CFTR* locus [19], both int1 and int11 sites recruit hepatocyte nuclear factor 1 (HNF1) [16], forkhead box A2 (FOXA2) [2,17,18,20] and caudal-type homeobox 2 (CDX2) [17,18,21] transcription factors in intestinal epithelial cells. Depletion of FOXA2 and CDX2 reduces *CFTR* expression in Caco2 cells [17,18] and *Cfr* expression is diminished in the small intestine of *Hnf1* –/– mice [19]. Of note, CDX2 is the most over-represented motif in open chromatin mapped by assay for transposase accessible chromatin (ATAC-seq) in human intestinal organoids

[22], a model for CFTR functional assays. Though most evident in cells of intestinal origin, int11 is also seen in a subset of human pancreatic adenocarcinoma cells and in human epididymis-derived epithelial cells [1,2].

Another common feature of the int1 and int11 CREs is their direct looping interaction with the *CFTR* promoter, as evidenced by several chromatin conformation capture protocols [1,2,6–8]. Due to the proximity of int1 to the promoter its interaction data are less robust than those of int11. The availability of detailed chromatin conformation capture data together with knowledge of key activating transcription factors for these sites enabled us to interrogate the mechanism underlying the cross-talk between int1, int11, structural features of the locus and the transcriptional network. To achieve this, we built upon data from Caco2 cell clones generated previously, from which the int11 CRE was removed by CRISPR/Cas9 non-homologous end joining (NHEJ) [2]. We generated clones in the same cell line from which the int1 element alone was removed, or both int1 and int11 sites were deleted within the same clone. Next we examined the impact of these deletions on higher order chromatin structure across the *CFTR* locus and on *CFTR* expression. The results suggest that the int1 element, though a weak enhancer, has a greater structural role at the locus and may facilitate the recruitment of the stronger int11 enhancer to the gene promoter. Loss of both enhancers activates other CREs at the locus. Also, using siRNA-mediated depletion of specific TFs we detail a pivotal role for these factors in the active locus 3D structure. Our results will inform mechanisms of regulation of other large genes with complex cell-type specific expression patterns.

## 2. Materials and methods

### 2.1. Cell culture

The Caco2 cell line [23] was purchased from ATCC and grown in DMEM (Dulbecco's Modified Eagle's medium) with 10% FBS (fetal bovine serum). For all experiments with Caco2 cells they were harvested 48 h post-confluence, a time at which *CFTR* expression is close to maximum levels [19].

### 2.2. CRISPR guide design, CRISPR/Cas9 transfection and screening

One pair of gRNAs flanking the DHS in intron 1, intron 11 region [2], –80.1 kb, and + 15.6 kb (Suppl. Table S3) were designed using the CRISPR Design Program (<http://crispr.mit.edu>). gBlocks from Integrated DNA Technologies (Iowa), were cloned into pSCB (StrataClone 240207) and sequenced to verify the correct sequence. Caco2 (WT) or Caco2 int1 cl1 (this work) or Caco2 int11 cl6 [2] cells were transfected with pMJ290 (wild-type Cas9 plasmid tagged with GFP) (Addgene, #42234) and the cloned gRNAs (for the particular deletion) in pSCB; using Lipofectamine 2000 (Life Technologies (LT), Carlsbad, CA). 48 h later, GFP positive cells single cells were isolated by fluorescence-activated cell sorting and seeded into 96-well plate for clonal expansion. Clones with homozygous deletions of each *cis*-regulatory element were confirmed by PCR of genomic DNA using primers flanking the gRNA PAM sites (specific reduction in size of product) and one flanking primer together with one located within the deletion (no product) and by Sanger sequencing. Primers are shown in Suppl. Table S2.

### 2.3. Reverse transcription quantitative PCR (RT-qPCR)

Total RNA from confluent cultures was extracted with TRIzol (LT) and cDNA prepared with the.

TaqMan reverse transcription kit (Invitrogen). *CFTR* mRNA levels were assayed in duplicate using a well characterized Taqman assay [19] (Suppl. Table S2) and normalized to beta 2 microglobulin ( $\beta$ 2M) as a housekeeping-gene control. All experiments were performed in triplicate.

### 2.4. siRNA-mediated depletion of transcription factors

For protein lysates and 4C-seq experiments upon knockdown,  $3 \times 10^6$  Caco2 cells were seeded in 10 cm dishes. After 24 h, they were forward transfected with NC siRNA-A (Santa Cruz sc-37007) or FOXA1 (sc-37930) and FOXA2 (sc-35569), CDX2 (sc-43680), or GATA6 (sc-37907) siRNAs at 200 pmol/dish using Lipofectamine RNAiMAX reagent. 96 h post-transfection, cells were harvested and pellets were prepared as per [24] for 4C-seq library preparation and analysis as explained below. Cell lysates were also collected simultaneously for validation of TF depletion by western blots. (Suppl. Table S4).

### 2.5. Western blot

To validate the knockdown of CDX2, 96 h post-transfection, lysates were collected using NET buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 1 mM EDTA) supplemented with Triton-X 100 and protease inhibitors (Sigma). Protein concentration was determined by Bradford assay. SDS sample loading buffer with  $\beta$ -mercaptoethanol was added and samples were boiled at 95 °C for 5 min and then resolved by SDS-PAGE, transferred to Immobilon membrane and western blots probed with antibodies specific for FOXA1 (Abcam 23738)/A2 (Millipore 07–633) CDX2 (Bethyl Laboratories #A300–692A or GATA6 (Cell signaling #D61E4) and  $\beta$ -tubulin (T4026, Sigma-Aldrich) with ECL detection.

### 2.6. Circular chromatin conformation capture and sequencing (4C-seq)

4C-seq libraries were generated from cultured cells as per [24] or [25]. All 4C experiments were done a minimum of twice on the same clonal cell line (technical replicate) and each deletion event was evaluated in at least two independent clonal lines (biological replicate). NlaIII and DpnII or Csp6I were used as the primary or secondary restriction enzymes respectively. Enzyme pairs and primer sequences used to generate 4C-seq libraries for each viewpoint are shown in Suppl. Table S2. The primers are marked with a unique barcode to enable multiplexing of libraries generated from same viewpoint on the same Hi-Seq 4000 or NextSeq 550 flow cell for sequencing. The sequencing data were processed using the 4Cseq pipe protocol [26] for the generation of domainograms and the data were quantified and mapped using the pipe4C processing pipeline [24]. For both pipelines, default parameters were used. Raw reads were aligned to the hg19 genome using Bowtie2 v4.8.3 and sorted using SAMtools v1.3. Bigwig subtraction tracks were generated using deepTools bigwigCompare [27] with default settings.

## 2.7. Statistics

Error bars in all graphs denote standard error of the mean (SEM). Statistical analysis used the Student's unpaired *t*-tests in Prism software (GraphPad).

## 3. Results

The genomic region encompassing the *CFTR* locus and the location of key CREs in intestinal epithelial cells are shown in Fig. 1A and Suppl. Table 1, which defines *CFTR* CREs in multiple cell types according to their coordinates on hg19 and hg38, with Legacy or RefSeq nomenclature.

### 3.1. Cross talk between intronic cis-regulatory elements of CFTR

**a) Generation of single and double CRE deletion clones in Caco2 cells—**In order to investigate the dependency of different CREs in one cell type we focused on two intestinal enhancers of *CFTR* located in intron 1 (185 + 10 kb; called DHS1/int1 here) and intron 11 (RefSeq intron 12, 1811 + 0.8 kb; called DHS11/int11 here). As shown in Fig. 1C, we first used CRISPR/Cas9 with guide RNAs (gRNAs) flanking each element to delete these two CREs independently in clonal cell lines. Caco2 cells were transfected with a pair of gRNAs to remove ~200 bp flanking the open chromatin peak at DHS1 (Fig. 1A,B Suppl. Table 1) to generate DHS1 deletion clones (del int1) (c11, c18). Homozygous deletions of DHS were validated by PCR followed by Sanger sequencing in 5 clones (Suppl. Fig. 1). Clones lacking ~1 kb flanking the open chromatin peak at the DHS11 element (del int11) were generated previously [2]. A second round of CRISPR/Cas9 editing was then performed on one del1 (c11) and one del11 (c16) clone to sequentially remove the DHS11 and DHS1 elements, respectively, thus generating del int1/ int11 double deletion clones (Fig. 1C). One double deletion clone was generated from del int1 c11 (D58) and two from del int11 c16 (T5, T14) and these were also validated by PCR and sequencing (Suppl. Fig. 1A).

**b) Loss of DHS1 and DHS11 CREs has additive effect in reducing CFTR expression—**The impact of CRE deletion on *CFTR* expression was measured by RT-qPCR in multiple clones and compared to both WT Caco2 cells and non-targeted WT clones generated in the same CRISPR/Cas9 NHEJ experiments (Fig. 1D). Deletion of the DHS1 or DHS11 CREs alone reduced *CFTR* expression to ~45% or 20% of WT levels, respectively. The DHS11 data confirm earlier results on the same clones [2]. In contrast, *CFTR* expression was almost undetectable in the del int1/int11 double deletion clones. The results are consistent with earlier luciferase reporter gene assays [1,14] showing that the DHS1 CRE had much lower enhancer activity than the DHS11 CRE and that these two elements acted cooperatively. Moreover our data suggest that these two CREs may contain enhancers with complementary functions. Of note neither the single nor double deletions altered splicing of adjacent exons of the *CFTR* transcript (Suppl. Fig. 1B).

### 3.2. The DHS1 and DHS11 CREs contribute together to higher order chromatin structure

We showed previously that intronic enhancers coordinate epithelial- specific looping of the active *CFTR* locus [1]. Also, that the contribution of individual CREs to higher order chromatin structure and gene expression was site specific [2]. So we next used a

quantitative, chromatin conformation capture protocol (q4C-seq) to examine the impact of the DHS1 and DHS11 CRE deletions on locus architecture (Fig. 2). All deletion clones were evaluated by replicated 4C-seq, but data from only one representative clone of each deletion are illustrated in the figures and supplementary figures. The interaction profile of each 4C-seq viewpoint in WT Caco2 cells is plotted from a wiggle file at the top of the figure and below it the subtraction of the WT wiggle file from the specific deletion clone wiggle file. Above each panel genomic locations and the map of open chromatin (ATAC-seq) peaks in Caco2 cells are shown. Loss of DHS1 (del int1) is associated with a gain of interactions (Fig. 2A, red bars and arrows) between the *CFTR* promoter and a) upstream sequences including the  $-80.1$  kb 5' TAD boundary and sequences close to the  $-20.9$  kb site together with b) downstream sequences 5' to and including the intron 23 DHS, the  $+6.8$  kb site and beyond the  $+48.9$  kb TAD boundary. The  $-20.9$  kb and  $+6.8$  kb sites are key sites of cell-type-selective CTCF recruitment at the *CFTR* locus, and function as enhancer-blocking insulator elements [9,10,12,13]. Elsewhere, there was a general slight reduction in interactions across the 5' half of the locus, though not involving any specific well-characterized CREs except a site in intron 10 (DHS10c/int10c, blue arrow). The structural contribution of the DHS1 enhancer was further supported by using a viewpoint at intron 11 which showed a marked increase in interactions at the DHS4 (int4) site and a loss of interactions with regions between this site and DHS11 (int11), together with a gain of interactions 3' to DHS11 in the del int1 clone (Suppl Fig. 2, grey arrow, blue bar, and grey bar respectively). In contrast, loss of DHS11 (del int11) was associated with few gains in interactions and again a slight reduction in interactions across the locus, particularly of elements close to DHS1 (int1) and in intron 10 adjacent to the deletion (Fig. 2A, red bar), consistent with our earlier observations [2]. In the double deletion clone (del int1/int11) the most notable feature was the substantial loss of interactions between the promoter and the 5' half of the locus (between DHS1 and DHS11) (Fig. 2A, red bars) suggesting that these 2 DHS have dependent functions in maintaining locus architecture.

With a viewpoint 3' to the coding region of *CFTR* (DHS + 15.6, an enhancer-blocking insulator element that does not recruit CTCF), a more substantial loss of interactions is seen across the locus upon deletion of DHS1 or DHS11, as noted by the red bars in Fig. 2B. This loss extends across most of the locus in the del int1 clone but is mainly in the 3' half in the del int11 clone. In both clones, a loss of interactions is also seen beyond the  $-80.1$  kb TAD boundary (Fig. 2B, red arrow). Unexpectedly the del int11 clone also shows a reduced interaction between the DHS15.6 viewpoint and regions immediately 5' to the  $-20.9$  kb site (Fig. 2B, red arrow). Deletion of both sites concurrently causes a dramatic loss of interactions of DHS + 15.6 with almost the whole locus and the 5' TAD boundary ( $-80.1$  kb, red dotted line and red arrow respectively).

Next, we examined interactions across the TAD in the del int1, del int11 and double deletion clones using viewpoints at key CTCF-bound structural elements at the locus ( $-20.9$  kb and  $+48.9$  kb (Fig. 3), and  $-80.1$  kb (Suppl. Fig. 3). For the  $-20.9$  kb viewpoint, del int1, del int11 and del int1/int11 clones all showed an enhancement of interactions with other sites of CTCF recruitment at  $+6.8$  kb and 3' to the  $+48.9$  kb 3' TAD boundary (marked by red bar and arrow in Fig. 3A). An increase in interactions was also seen between the viewpoint and the gene promoter and also the middle of the locus (red arrow and bar in Fig. 3A)

in the del int1 clone. In the del11 clone the -20.9 kb viewpoint also gained interactions with the promoter and an extended region across introns 1–3, together with a site in intron 4 DHS4/int4 (Fig. 3A red bar and arrow, respectively). DHS4 was previously noted as an important site of interactions at the locus [2]. Of note it is also seen as a site of H3K4me3 deposition during epithelial differentiation [28]. Only in the double deletion clone was a loss of interactions evident between the -20.9 viewpoint and the whole of the locus (dashed red line in Fig. 3A). A similar increase in interactions with CTCF sites (-20.9 kb and +6.8 kb) in the del int1, del int11 and del int1/11 clones was evident using the +48.9 kb TAD boundary as a viewpoint (red arrows in Fig. 3B), together with enhanced interactions with the 5' TAD boundary (-80.1 kb). Again a loss of interactions was evident across the whole locus in the double deletion clone. Only minor alterations in interaction profiles were seen with the -80.1 kb 5' TAD boundary viewpoint (Suppl. Fig. 3), but these were consistent with the other CTCF-bound viewpoints in that the del int1 and del int1/int11 clone showed increased interactions with the -20.9 kb and +6.8 kb sites (Suppl. Fig. 3, red arrows). Overall, these data demonstrate a key role for structural elements at the *CFTR* locus in responding to the loss of specific enhancers by altering the higher order structure of the locus, possibly enabling the recruitment of other CREs. These structural elements include those binding CTCF but also others, such as the +15.6 kb element which recruits the cohesin complex but not CTCF [29]. However, our data on the del int1/int11 double deletion clones suggests that this response is not able to maintain gene expression when more than one CRE, (one with weak enhancer function together with a structural role and the other with strong enhancer function), are both absent.

### 3.3. The structural contribution of sites recruiting CTCF and cohesin complex to the locus

To investigate further the role of CTCF and cohesin recruitment at the *CFTR* locus in interpreting the impact of the DHS1 and DHS11 enhancer deletions, we generated clones lacking the -80.1 kb CTCF site at the TAD boundary or the +15.6 kb cohesin (RAD21) site, by CRISPR/Cas9 NHEJ. The precise deletions are shown in Suppl. Fig. 1C. Gene expression analysis by RT-qPCR showed that loss of the -80.1 kb site resulted in a 31% reduction in *CFTR* expression (Suppl. Fig. 1C left), while deletion of the +15.6 kb site was associated with a 68% reduction (Suppl. Fig. 1C right). Upon deletion of the -80.1 kb site, the promoter gains 5' interactions towards and beyond the TAD boundary (Fig. 4A, grey arrows and bar), particularly showing increased interaction with an enhancer at -35 kb [3,30] which is not normally evident in intestinal cell lines. Concurrently a loss of promoter interactions within intron 3, also CREs at the center of the locus (blue bars) and 5' to a DHS in intron 23 (blue arrow) are seen. Thus removal of the 5' TAD boundary causes an extensive change in looping of CREs to the gene promoter. Upon deletion of the -80.1 kb site, the -20.9 kb viewpoint also gains interactions with an extended 5' region, (Fig. 4B, grey bar) which may include alternative CTCF binding sites, though none are evident in ENCODE data from other cell types [31]. Enhanced interaction near DHS23/int23 and at the +6.8 kb CTCF-binding insulator are also evident (Fig. 4B, grey arrows). Losses of interaction with the -20.9 kb viewpoint are evident in the 5' part and the middle of the gene (blue bars) together with the regions adjacent the +15.6 kb cohesin binding site and the 3'

TAD boundary at +48.9 kb (blue arrows). These altered interactions may reflect structural changes at the locus that partially compensate for loss of the -80.1 kb site.

In contrast, deletion of the +15.6 kb site of cohesin occupancy was associated with a loss of interactions between the TAD boundaries, as seen from the -80.1 kb viewpoint, but also a reduction at multiple interacting sites across the locus including on the 5' side of -20.9 (Fig. 4C, blue arrows). Together these data support a mechanism whereby multiple structural elements interact coordinately across the locus, so that loss of one of them can cause recruitment of other sites.

### 3.4. Activating transcription factors are required for normal CRE interactions and promoter recruitment

We previously identified several transcription factors (TF) with a pivotal role in activating *CFTR* expression in intestinal epithelial cells [17–19,22]. These include the pioneer factor Forkhead box A2 (FOXA2) and also caudal-type homeobox 2 (CDX2). In earlier work we showed by quantitative chromatin conformation capture (q3C, [32]) that siRNA-mediated depletion of FOXA1/FOXA2 substantially reduced the interaction frequency between the *CFTR* promoter, the middle of the locus (introns 10 and 11) and the 3' insulator region (+6.8 kb to +15.6 kb) [18]. Here we followed up on these observations to investigate the role of TFs in the higher order chromatin structure across the whole TAD using 4C-seq and multiple viewpoints. FOXA1/A2, CDX2 and GATA6 are all recruited to the DHS11 enhancer element among other CREs, as shown by ChIP-seq [20,21,33]. Following siRNA-mediated depletion of each factor and validation of effective depletion by RT-qPCR (Suppl. Fig. 4) and western blot, 4C-seq was performed to determine whether loss of the TF impacted higher order chromatin structure at the *CFTR* locus. In Fig. 5 the effect of depleting FOXA1/FOXA2 on interactions across the locus is shown from viewpoints at -80.1 kb and +15.6 kb, and from DHS11 in Suppl. Fig. 5. The most notable feature of these domainograms and wiggle subtraction files is the loss of interaction of multiple sites with all 3 viewpoints (shown by dotted red lines under the subtraction tracks). The loss of interactions with the -80.1 kb viewpoint is primarily between DHS4 (red arrowhead), DHS11 (int11) and DHS23 (blue arrowhead) (Fig. 5A), while the +15.6 kb viewpoints loses interactions from the promoter (red arrowhead) through to an element in intron 18 (legacy, blue arrowhead) (Fig. 5B). For the DHS11 viewpoint the major loss of interactions is between this site and the adjacent regions of intron 10 (Suppl. Fig. 5, blue bar), which also contain FOXA2 binding sites, at DHS4 (Suppl. Fig. 5, red arrow) and at the 3' end of the locus including the +48.9 kb TAD boundary (Suppl. Fig. 5, grey bar and arrow).

Depletion of CDX2 by siRNA also caused a reduction in *CFTR* expression (Suppl. Fig. 4), however, the impact on the higher order chromatin structure at the locus was less dramatic than seen upon FOXA1/A2 depletion. With a viewpoint at -80.1 kb, in addition to loss of interaction with sites proximal to the -20.9 kb site, (Fig. 6A, blue bar) and a number of other regions of the gene, not associated with known CREs, a marked gain in interactions was seen with the 6.8 kb downstream CTCF-binding insulator and the +15.6 kb insulator (Fig. 6A, red bar and arrow, respectively). With a viewpoint at the promoter, loss of CDX2 was accompanied by a marked decrease in interactions with CDX2 binding sites at intron 10ab

and adjacent to +15.6 kb (Fig. 6B, red arrows), together with a minor gain in interactions in the 3' part of the gene. These results suggest that loss of recruitment of specific (CDX2 bound) CREs to the gene promoter when CDX2 is lost is accompanied by a strengthening of interactions between CTCF- bound structural elements. This mechanism may partially close the locus and reduce accessibility to other activating factors.

Finally, since a robust peak of GATA6 occupancy was evident at DHS11 in Caco2 cells (Fig. 1A), the impact of siRNA-mediated depletion of GATA6 on looping at the locus was examined. Though CFTR protein decreases substantially (Suppl. Fig. 4c), no specific changes in interactions were seen with viewpoints at the promoter or DHS11 (data not shown) suggesting additional (possibly indirect) mechanisms of action of this factor.

#### 4. Discussion

The coordinated action of multiple *cis*-regulatory elements that act upon one gene promoter is rarely fully elucidated. Here we used the *CFTR* locus to address the mechanisms underlying this synchronization since its regulatory complexity is well studied (reviewed in [34,35] and the CREs controlling its expression appear to be housed within a single ~316 kb TAD. Our goal was to focus on two elements that we showed previously to cooperate in luciferase reporter gene assays and determine a) how they cooperate when separated by ~100 kb of genomic DNA in the endogenous locus and b) to what extent their interaction depends upon i) the structure of the TAD itself and ii) the recruitment of activating transcription factors.

Our results suggest that for this locus, two intronic enhancers that are required for the high levels of *CFTR* expression seen in intestinal epithelial cells, both of which are associated with enrichment of active histones (H3K27Ac), appear to have different primary functions when working together in the genomic context. Of note both enhancers correspond to peaks of open chromatin in intestinal organoids [22] as well as shown here in Caco2 cells. We have not observed a cell line that exhibits an active DHS1 in the absence of open chromatin at DHS11 and predict that these 2 sites are active in the same cell. This prediction is supported by preliminary scATAC-seq data in a pancreatic adenocarcinoma cell line (unpublished). In earlier work in the 16HBE14o- airway epithelial cell line, which unlike the majority of secretory cells in the airway epithelium expresses high levels of *CFTR*, we showed that deletion of either the -44 kb or -35 kb upstream enhancer reduced *CFTR* expression to less than 5% of WT cells [3]. These data suggest that these two CREs work independently, consistent with their different roles in the immune response (-35 kb DHS, [30]) or oxidative stress (-44 kb DHS, [36]). In contrast, we show here that the DHS1 and DHS11 intronic enhancers apparently work together. Removal of both DHS1 and DHS11 in the same cell has a dramatic effect on *CFTR* expression and locus architecture. Not only is *CFTR* expression abolished, interactions between the promoter and sites within the gene are all substantially reduced. Concurrently, interactions between the TAD boundaries (-80.1 kb and + 48.9 kb) and sites across the gene are also diminished, consistent with the involvement of these structural elements in establishing and maintaining gene expression. A similar phenomenon is observed upon loss of the +15.6 kb insulator element, which apparently recruits cohesin but not CTCF [10,29]. However, the loss of DHS1 and DHS11 individually

has different effects, with the former primarily increasing interactions between the promoter and structural elements (−20.9 kb and + 6.8 kb) while the latter mainly causes a reduction of interaction between the promoter and other DHS in the middle of the locus. Viewpoints at the 3' TAD boundary or the −20.9 kb structural elements suggest that the locus responds to perturbations caused by loss of either DHS1 or DHS11 by increasing interactions between these sites. In some cases the loss of DHS11 may also activate interactions with a site of open chromatin in a distal intron of the gene (DHS23). Together our results suggest that though DHS1 has weak enhancer activity in Caco2 cells and carries the H3K27Ac mark, its most important role may be in spooling the regulatory elements in the middle of the locus (~ 100 kb distal, DHS10a,b and DHS11), which recruit activating transcription factors, into close proximity with the gene promoter. The peak of open chromatin in intron 4 (DHS4) may also contribute to this spooling mechanism.

The role of the TAD boundaries in maintaining normal regulatory mechanisms for the *CFTR* locus was partially addressed by experiments in which the 5' TAD boundary (−80.1 kb) was removed. The results showed that despite little impact on *CFTR* expression, this deletion caused a marked loss of interactions between the promoter and multiple sites in the middle of the locus, particularly downstream of DHS11 and at DHS23. Concurrently, increased interactions were evident between the promoter, new sites 5' to the deleted TAD boundary, (presumably alternative CTCF sites) and also at the −35 kb DHS that is not normally seen in Caco2 cells. Loss of the 5' TAD boundary also reduced interactions between the −20.9 kb insulator and the promoter, sites in introns 1–3 and the middle of the locus, but also increased interactions with the +6.8 kb 3' insulator. We showed previously that removal of the −20.9 kb element had little impact on *CFTR* expression but resulted in a major reorganization of locus architecture associated with enhanced recruitment of CTCF to other structural elements including the TAD boundaries (−80.1 kb and + 48.9 kb), a CTCF site at the 5' end of intron 1 and the +6.8 kb insulator [2]. Our current observations upon removal of the 5' TAD boundary are thus consistent with this inherent response mechanism of the locus to structural perturbations.

Finally, addressing the importance of activating transcription factor recruitment to the higher order structure of the locus, we depleted FOXA1/A2, CDX2 and GATA6, all of which occupy one or more enhancer sites in the gene. In earlier work we showed that loss of the DHS11 enhancer had a broader effect on FOXA2 recruitment as this dropped significantly at the nearby DHS10a,b, though not at more distal FOXA2 binding sites [2]. Loss of FOXA2 was also associated with a decrease in looping interactions between the *CFTR* promoter, the middle of the locus (particularly DHS10a,b) and the 3' insulators [18]. Our 4C-seq data suggest that depletion of FOXA2 has a substantial effect on looping of CREs across the locus to the 5' TAD boundary. A loss of interactions is noted at known CREs in intron 4 (DHS4) in intron 11 (DHS11) and intron23 (DHS23) in addition to other sites of open chromatin in the 3' half of the gene. In parallel, the DHS +15.6 kb insulator shows reduced interactions with the *CFTR* promoter and multiple other sites across the locus. This may be consistent with the pivotal role of FOXA2 as a pioneer factor acting to open chromatin at the *CFTR* locus [17,18]. Of note, Caco2 clones carrying a mutation in a single FOXA2 binding site in the DHS11 enhancer core did not exhibit altered *CFTR* expression, consistent with a combined role for multiple FOXA2 recruitment sites (data not shown). In contrast,

depletion of CDX2, though also a critical factor in intestinal epithelial cell function, has a more limited effect with loss of interaction between the promoter and the known sites of CDX2 occupancy at DHS10a,b and DHS +15.6 kb. As for other perturbations of locus architecture, loss of CDX2 is accompanied by an increase in interactions of the +6.8 kb insulator with the 5' TAD boundary. Loss of GATA6, though associated with an increase in CFTR expression may not have a key role in locus architecture, though it is accompanied by a general increase in promoter interactions.

Integrating the results presented here we can build a more advanced model for cell-specific expression of *CFTR* in intestinal epithelial cells and in other cells in the pancreas and epididymis that share the same intronic enhancers (eg in DHS1, DHS11 and DHS23; Fig. 7). When considering the location of these 3 CREs across the gene it is notable that DHS11 is almost equidistant between the promoter and the +6.8 kb 3' insulator. In contrast DHS1 is about 10 kb from the promoter and DHS23 about 8 kb from the 3' insulator, so the precise genomic distance over which the CREs need to communicate may be relevant. The strongest enhancer for intestinal expression of *CFTR* appears to be the CRE at DHS11 and its activation is associated with looping to the gene promoter. However, the data we show here suggests that this site cooperates with the much weaker enhancer at DHS1, which has a more dominant structural role in maintaining higher order architecture of the active locus. The cross-talk between the CREs is shown to be more extensive when either DHS1 or both DHS1 and DHS11 are lost, and the CRE at DHS23 then shows enhanced interactions with the gene promoter. Two additional structural elements appear critical to the looping mechanism: a) the CTCF-binding insulator at -20.9 kb which has an overarching role in supporting the 3D looped structure through direct interactions with other CTCF binding sites in introns 1, 2, 10 (DHS10c) and + 6.8 kb; b) the cohesin-binding insulator at +15.6 kb, which may have a key role in redistributing cohesin rings across the looped locus. Finally, considering the cell-specific transcription factors that are required for initiating and maintaining *CFTR* expression, these may be the primary drivers for establishing the CRE-to-promoter loops. The concept of transcription factories, with high concentration of TFs in discrete nuclear compartments was first suggested many decades ago [37] and has been reinforced by more recent high-resolution analysis (reviewed in [38,39]. If FOXA2, CDX2 and GATA6, among other activating TFs that are recruited to *CFTR* CREs, are physically concentrated in a discrete nuclear compartment this could draw the CREs into looped structures, which could then be stabilized by cohesin rings. Simultaneously CTCF could support the higher order structure of the active locus and tether it to the TAD boundaries. This refined model for *CFTR* expression in intestinal epithelial cells is consistent with loop extrusion models of chromosomal domain formation [40–43]. Of note in the context of gene editing approaches to correct or circumvent the disease-associated errors in the *CFTR* gene, it may be imperative to consider the complex higher order interactions of structural and enhancer elements at the locus that are required for normal gene expression.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank Dr. Pieter Faber and the University of Chicago Genomics Core and the Case Western Reserve Genomics Core for sequencing; also Alex Ge, Abigail Lindsey, Drs. Kay-Marie Lamar and Shih-Hsing Leir for their contributions.

## Funding

This work was supported by the National Institutes of Health [R01 HL094585, R01 HD068901 (AH) and T32 GM008056 (AP)] and the Cystic Fibrosis Foundation [Harris 16G0].

## Data availability

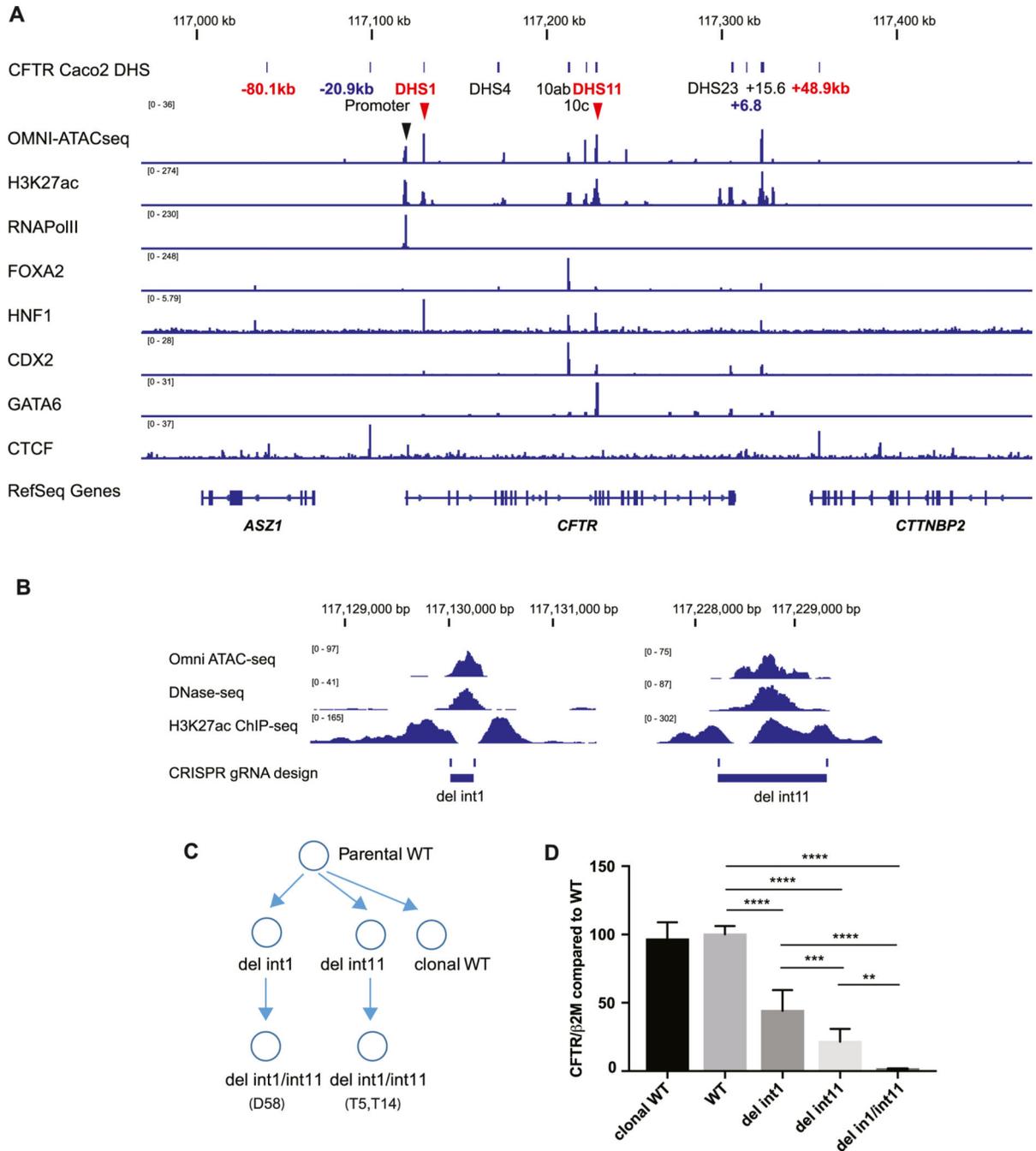
4C-seq data generated in this study are available at NCBI GEO accession GSE186272

## References

- [1]. Ott CJ, Blackledge NP, Kerschner JL, Leir SH, Crawford GE, Cotton CU, Harris A, Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 19934–19939.
- [2]. Yang R, Kerschner JL, Gosalia N, Neems D, Gorsic LK, Safi A, Crawford GE, Kosak ST, Leir SH, Harris A, Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus, *Nucleic Acids Res.* 44 (2016) 3082–3094. [PubMed: 26673704]
- [3]. NandyMazumdar M, Yin S, Paranjapye A, Kerschner JL, Swahn H, Ge A, Leir SH, Harris A, Looping of upstream cis-regulatory elements is required for CFTR expression in human airway epithelial cells, *Nucleic Acids Res.* 48 (2020) 3513–3524. [PubMed: 32095812]
- [4]. Bombieri C, Seia M, Castellani C, Genotypes and phenotypes in cystic fibrosis and cystic fibrosis transmembrane regulator-related disorders, *Semin Respir Crit Care Med* 36 (2015) 180–193. [PubMed: 25826586]
- [5]. Veit G, Roldan A, Hancock MA, Da Fonte DF, Xu H, Hussein M, Frenkiel S, Matouk E, Velkov T, Lukacs GL, Allosteric folding correction of F508del and rare CFTR mutants by elxacaftor-tezacaftor-ivacaftor (Trikafta) combination, *JCI Insight* 5 (2020).
- [6]. Gheldof N, Smith EM, Tabuchi TM, Koch CM, Dunham I, Stamatoyannopoulos JA, Dekker J, Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene, *Nucleic Acids Res.* 38 (2010) 4325–4336. [PubMed: 20360044]
- [7]. Moisan S, Berlivet S, Ka C, Gac GL, Dostie J, Ferec C, Analysis of long-range interactions in primary human cells identifies cooperative CFTR regulatory elements, *Nucleic Acids Res.* 44 (2015) 2564–2576. [PubMed: 26615198]
- [8]. Smith EM, Lajoie BR, Jain G, Dekker J, Invariant TAD boundaries constrain cell-type-specific looping interactions between promoters and distal elements around the CFTR locus, *Am. J. Hum. Genet.* 98 (2016) 185–201. [PubMed: 26748519]
- [9]. Nuthall H, Vassaux G, Huxley C, Harris A, Analysis of a DNase I hypersensitive site located –20.9 kb upstream of the CFTR gene, *Eur J Bioch* 266 (1999) 431–443.
- [10]. Blackledge NP, Ott CJ, Gillen AE, Harris A, An insulator element 3' to the CFTR gene binds CTCF and reveals an active chromatin hub in primary cells, *Nucleic Acids Res.* 37 (2009) 1086–1094. [PubMed: 19129223]
- [11]. Bell AC, West AG, Felsenfeld G, The protein CTCF is required for the enhancer blocking activity of vertebrate insulators, *Cell* 98 (1999) 387–396. [PubMed: 10458613]
- [12]. Blackledge NP, Carter EJ, Evans JR, Lawson V, Rowntree RK, Harris A, CTCF mediates insulator function at the CFTR locus, *Biochem. J.* 408 (2007) 267–275. [PubMed: 17696881]
- [13]. Nuthall HN, Moulin DS, Huxley C, Harris A, Analysis of DNase I hypersensitive sites at the 3' end of the cystic fibrosis transmembrane conductance regulator gene, *Biochem. J.* 341 (1999) 601–611. [PubMed: 10417323]

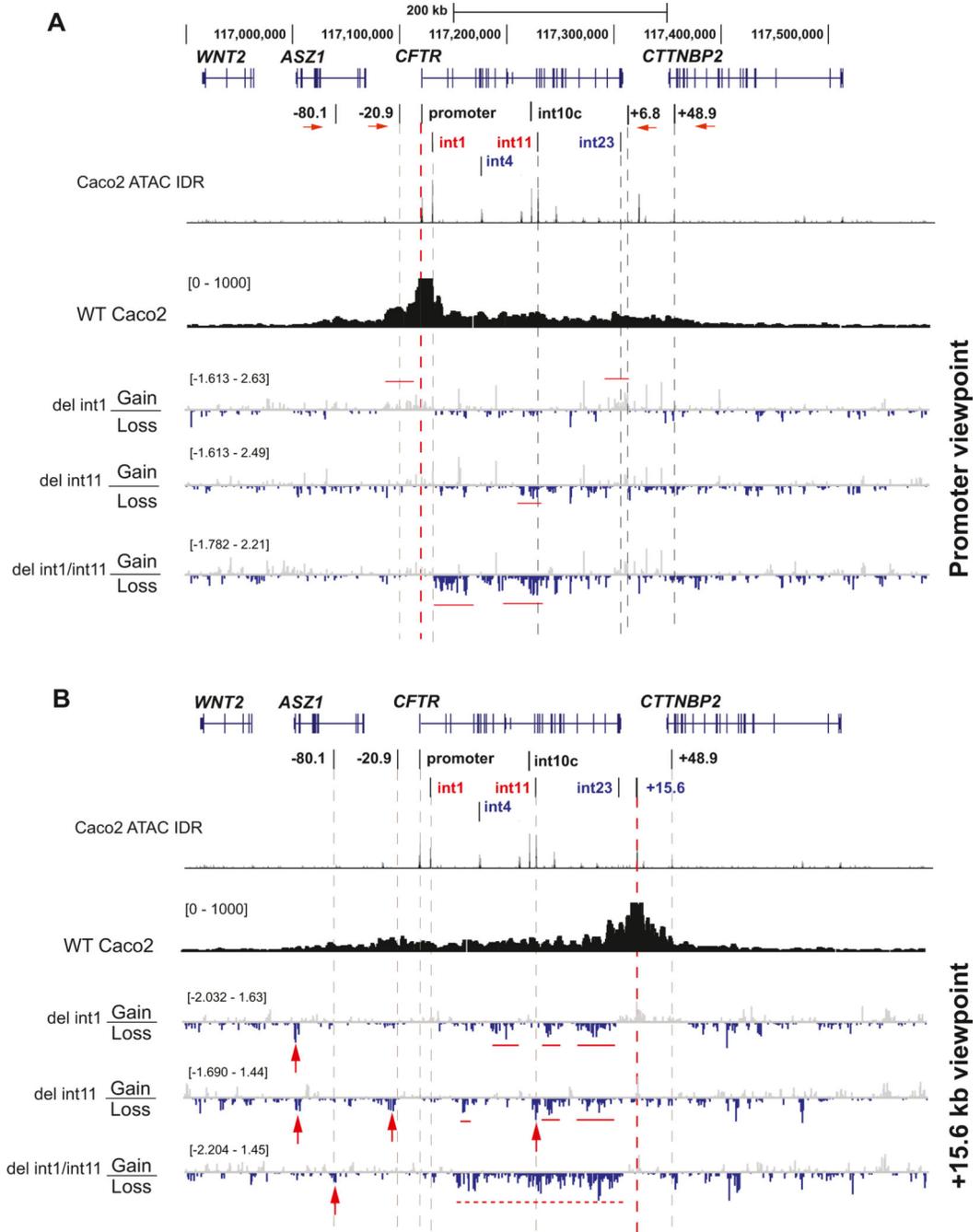
- [14]. Smith AN, Barth ML, McDowell TL, Moulin DS, Nuthall HN, Hollingsworth MAAH, A regulatory element in intron 1 of the cystic fibrosis transmembrane conductance regulator gene, *J. Biol. Chem.* 271 (1996) 9947–9954. [PubMed: 8626632]
- [15]. Rowntree R, Vassaux G, McDowell TL, Howe S, McGuigan A, Phylactides M, Huxley C, Harris A, An element in intron 1 of the CFTR gene augments intestinal expression in vivo, *Hum. Mol. Genet.* 11 (2001) 1455–1464.
- [16]. Ott CJ, Suszko M, Blackledge NP, Wright JE, Crawford GE, Harris A, A complex intronic enhancer regulates expression of the CFTR gene by direct interaction with the promoter, *J. Cell. Mol. Med.* 13 (2009) 680–692. [PubMed: 19449463]
- [17]. Kerschner JL, Harris A, Transcriptional networks driving enhancer function in the CFTR gene, *Biochem. J.* 446 (2012) 203–212. [PubMed: 22671145]
- [18]. Kerschner JL, Gosalia N, Leir SH, Harris A, Chromatin remodeling mediated by the FOXA1/A2 transcription factors activates CFTR expression in intestinal epithelial cells, *Epigenetics* 9 (2014) 557–565. [PubMed: 24440874]
- [19]. Mouchel N, Henstra SA, McCarthy VA, Williams SH, Phylactides M, Harris A, HNF1alpha is involved in tissue-specific regulation of CFTR gene expression, *Biochem. J.* 378 (2004) 909–918. [PubMed: 14656222]
- [20]. Gosalia N, Yang R, Kerschner JL, Harris A, FOXA2 regulates a network of genes involved in critical functions of human intestinal epithelial cells, *Physiol. Genomics* 47 (2015) 290–297. [PubMed: 25921584]
- [21]. Verzi MP, Shin H, He HH, Sulahian R, Meyer CA, Montgomery RK, Fleet JC, Brown M, Liu XS, Shivdasani RA, Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2, *Dev. Cell* 19 (2010) 713–726. [PubMed: 21074721]
- [22]. Yin S, Ray G, Kerschner JL, Hao S, Perez A, Drumm ML, Browne JA, Leir SH, Longworth M, Harris A, Functional genomics analysis of human colon organoids identifies key transcription factors, *Physiol. Genomics* 52 (2020) 234–244. [PubMed: 32390556]
- [23]. Fogh J, Wright WC, Loveless JD, Absence of HeLa cell contamination in 169 cell lines derived from human tumors, *J. Natl. Cancer Inst.* 58 (1977) 209–214. [PubMed: 833871]
- [24]. Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, de Laat W, 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis, *Methods* 170 (2020) 17–32. [PubMed: 31351925]
- [25]. Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W, Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation, *Methods* 58 (2012) 221–230. [PubMed: 22609568]
- [26]. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, et al. , Robust 4C-seq data analysis to screen for regulatory DNA interactions, *Nat. Methods* 9 (2012) 969–972. [PubMed: 22961246]
- [27]. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T, deepTools2: a next generation web server for deep-sequencing data analysis, *Nucleic Acids Res.* 44 (2016) W160–W165.
- [28]. Wang A, Yue F, Li Y, Xie R, Harper T, Patel NA, Muth K, Palmer J, Qiu Y, Wang J, et al. , Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates, *Cell Stem Cell* 16 (2015) 386–399. [PubMed: 25842977]
- [29]. Gosalia N, Neems D, Kerschner JL, Kosak ST, Harris A, Architectural proteins CTCF and cohesin have distinct roles in modulating the higher order structure and expression of the CFTR locus, *Nucleic Acids Res.* 42 (2014) 9612–9622. [PubMed: 25081205]
- [30]. Zhang Z, Leir SH, Harris A, Immune mediators regulate CFTR expression through a bifunctional airway-selective enhancer, *Mol. Cell. Biol.* 33 (2013) 2843–2853. [PubMed: 23689137]
- [31]. Consortium, E.P, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, An integrated encyclopedia of DNA elements in the human genome, *Nature* 489 (2012) 57–74. [PubMed: 22955616]

- [32]. Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T, Quantitative analysis of chromosome conformation capture assays (3C-qPCR), *Nat. Protoc.* 2 (2007) 1722–1733. [PubMed: 17641637]
- [33]. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al. , Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature* 447 (2007) 799–816. [PubMed: 17571346]
- [34]. Gosalia N, Harris A, Chromatin dynamics in the regulation of CFTR expression, *Genes (Basel)* 6 (2015) 543–558. [PubMed: 26184320]
- [35]. Swahn H, Harris A, Cell-selective regulation of CFTR gene expression: relevance to gene editing therapeutics, *Genes (Basel)* 10 (2019).
- [36]. Zhang Z, Leir SH, Harris A, Oxidative stress regulates CFTR gene expression in human airway epithelial cells through a distal antioxidant response element, *Am. J. Respir. Cell Mol. Biol.* 52 (2014) 387–396.
- [37]. Iborra FJ, Pombo A, Jackson DA, Cook PR, Active RNA polymerases are localized within discrete transcription ‘factories’ in human nuclei, *J. Cell Sci.* 109 (1996) 1427–1436. [PubMed: 8799830]
- [38]. Rieder D, Trajanoski Z, McNally JG, Transcription factories, *Front Genet* 3 (2012) 221. [PubMed: 23109938]
- [39]. Wollman AJ, Shashkova S, Hedlund EG, Friemann R, Hohmann S, Leake MC, Transcription factor clusters regulate genes in eukaryotic cells, *Elife* 6 (2017).
- [40]. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA, Formation of chromosomal domains by loop extrusion, *Cell Rep.* 15 (2016) 2038–2049. [PubMed: 27210764]
- [41]. Gassler J, Brandao HB, Imakaev M, Flyamer IM, Ladstatter S, Bickmore WA, Peters JM, Mirny LA, Tachibana K, A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture, *EMBO J.* 36 (2017) 3600–3618. [PubMed: 29217590]
- [42]. Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H, Human cohesin compacts DNA by loop extrusion, *Science* 366 (2019) 1345–1349. [PubMed: 31780627]
- [43]. Banigan EJ, van den Berg AA, Brandao HB, Marko JF, Mirny LA, Chromosome organization by one-sided and two-sided loop extrusion, *Elife* 9 (2020).



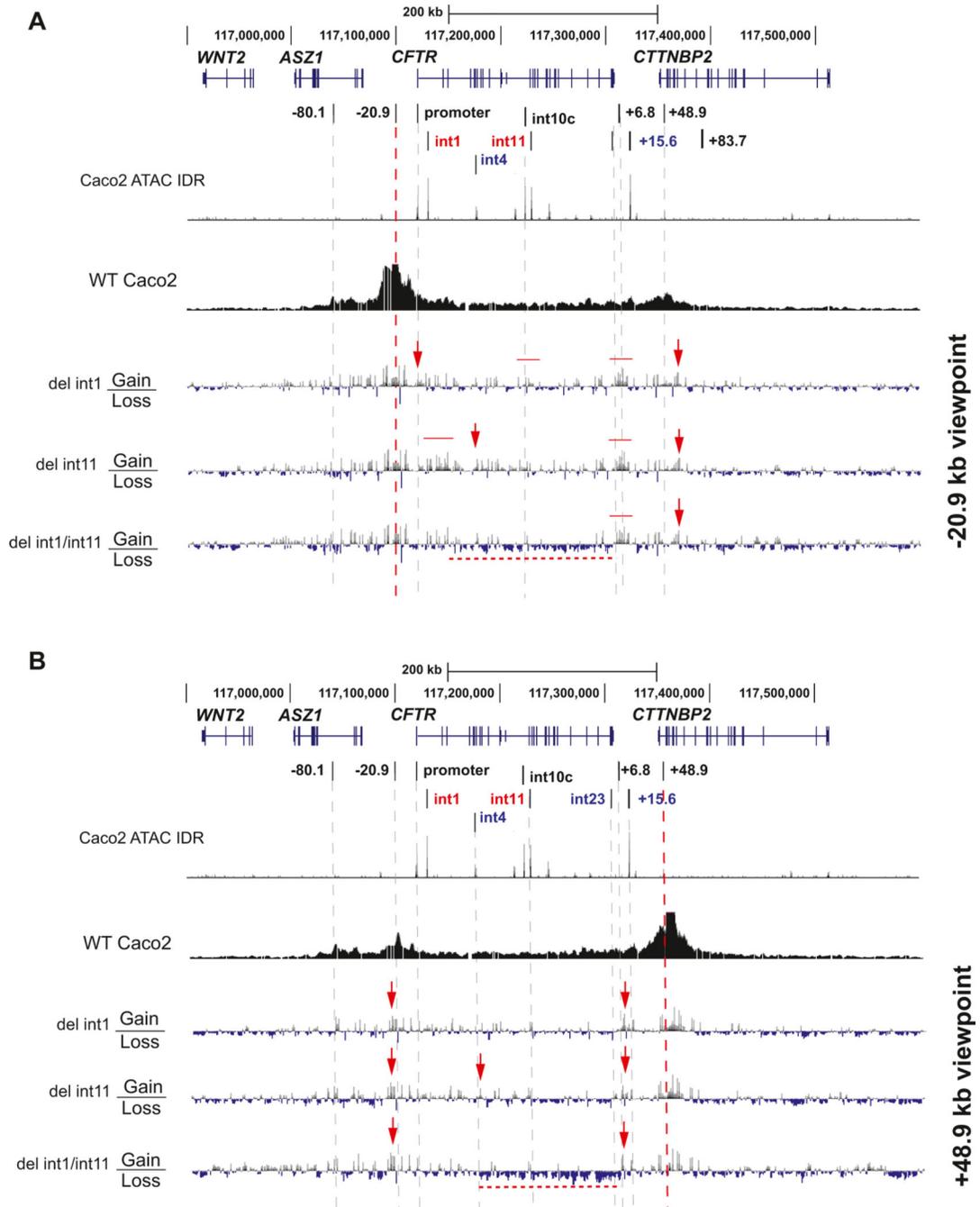
**Fig. 1.** Genomic location and epigenetic profile of key intestinal CREs at the *CFTR* locus and experimental plan for their removal. A. IGV browser view of data from Caco2 cells: DNase I hypersensitive sites (DHS) locations are listed in Suppl. Table 1; the TAD boundaries at -80.1 kb and + 48.9 kb are shown in red, the insulator elements at -20.9 kb and + 6.8 kb are shown in blue, enhancers at DHS1 (int1) and DHS11 (int11) are highlighted by red arrows; open chromatin data generated by Omni-ATACseq are from GSE140456; ChIP-seq data are sourced as follows: for H3K27ac and RNAPolIII, GSE132807; FOXA2,

GSE66218; HNF1, GSE67740; CDX2 and GATA6 GSE23436; CTCF, GSE30263. B. IGV browser view of CRISPR/Cas9 design for deletion of DHS1 and DHS11 CREs. Caco2 DNase-seq data are from GSE29692, other tracks are as in panel A. C. Cartoon showing generation of the enhancer deletion clones. D. *CFTR* expression is reduced upon deletion of the DHS1 and DHS11 enhancers. RT-qPCR analysis showing *CFTR* expression relative to beta-2-microglobulin ( $\beta 2M$ ) in parental WT, clonal WT ( $n = 2$ ), del int1 clones ( $n = 4$ ), del int11 clones ( $n = 2$ ) and del int1/int11 clones ( $n = 3$ ). One-way ANOVA test, values are mean  $\pm$  SEM (standard error of the mean), technical replicates  $>3$ , \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \*\* $p < 0.01$ .

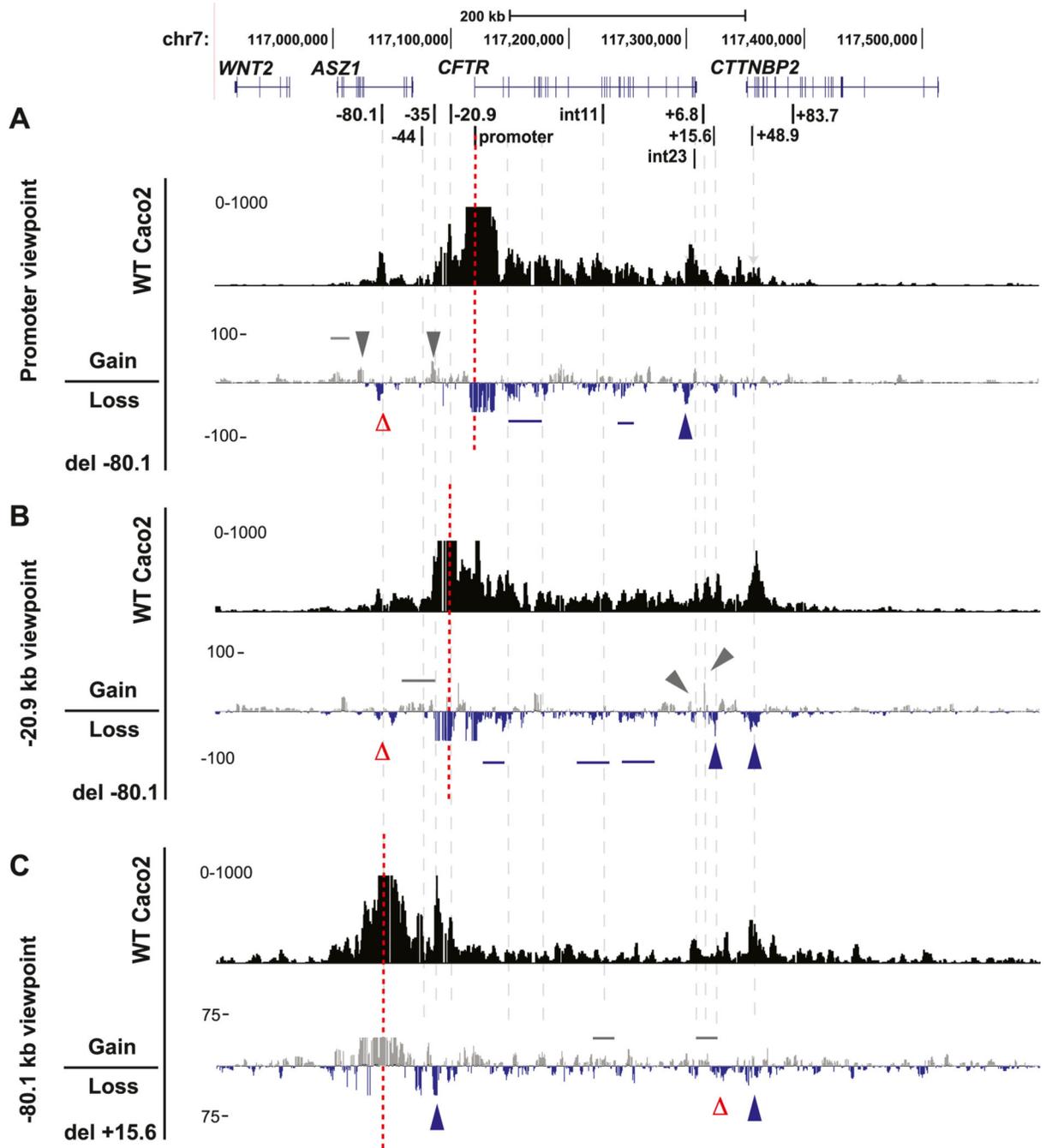


**Fig. 2.** Loss of the DHS1 and DHS11 enhancers alters higher-order chromatin structure at the *CFTR* locus. A, B. Chromatin structure across the *CFTR* locus assayed by 4C-seq using viewpoints at (A) the *CFTR* promoter or (B) the +15.6 kb insulator. A schematic at the top of each panel shows the *CFTR* locus on chromosome 7 and below are key CREs for the *CFTR* locus. Red arrows denote the orientation of key CTCF sites, Genome browser tracks of Caco2 ATAC-seq data are shown with 4C-seq quantification data (pipe4C) aligned below. For WT Caco2, the quantified track is in black. For the enhancer-deleted cell lines a

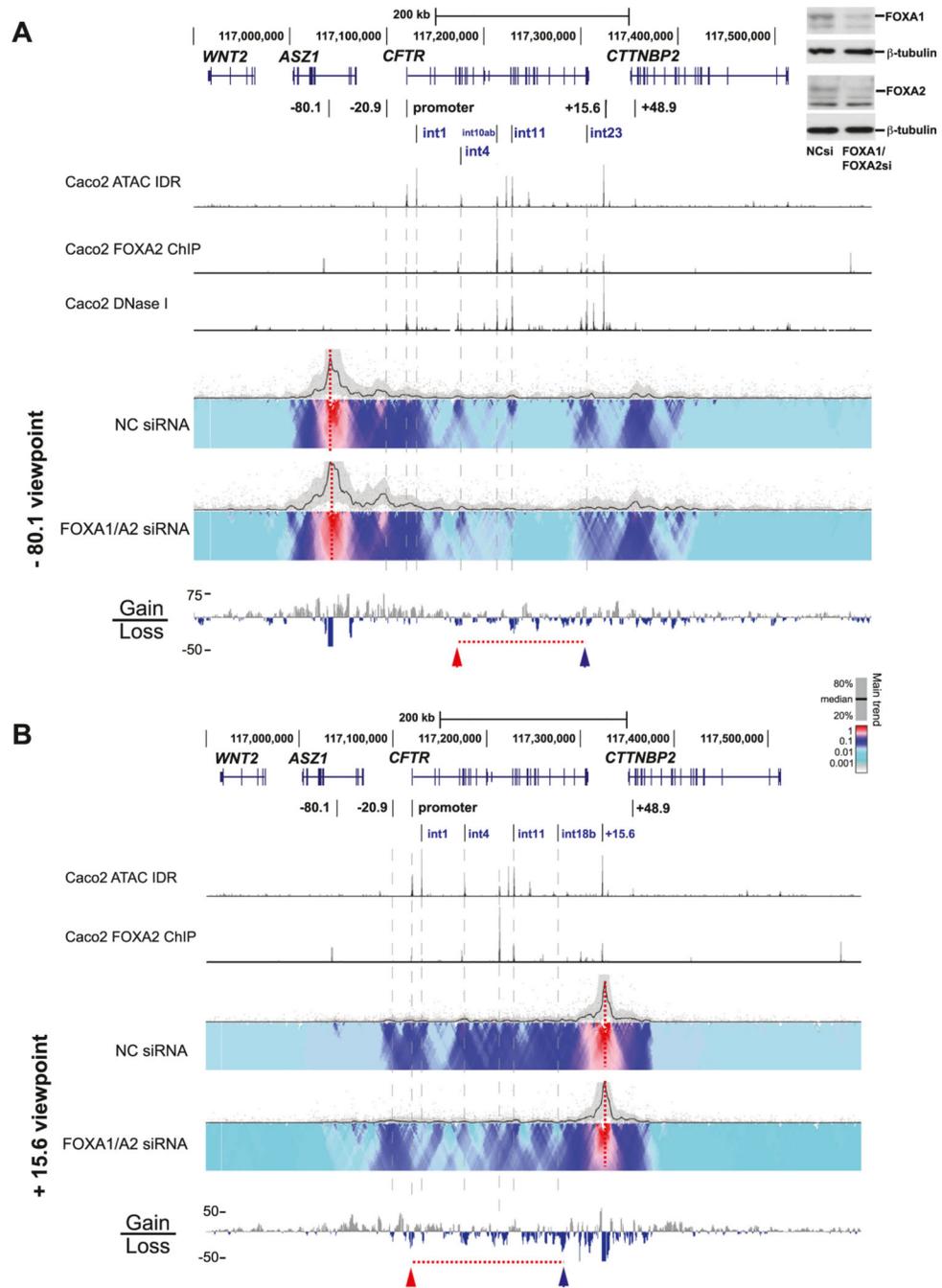
comparison of their interaction frequencies for each viewpoint (marked by red dashed line) with those seen in the unmodified (WT) cells is shown. These comparisons are generated by subtracting the read-density-normalized WT signal from the deletion clone, so signals above zero, colored grey, indicate a gain of interaction with viewpoint compared to WT, while signals below zero, colored blue, indicate a loss of interaction with viewpoint compared to WT. Subtraction signal is Log2 normalized. Red bars and arrows denote sites of specific interest referred to in the results section.



**Fig. 3.** Loss of the DHS1 and DHS11 enhancers alters higher-order chromatin structure at the *CFTR* locus. Chromatin structure at the *CFTR* locus analyzed from viewpoints at (A) the -20.9 insulator and (B) the +48.9 kb 3' TAD boundary. All features of the figures and analysis protocols are as described for Fig. 2.

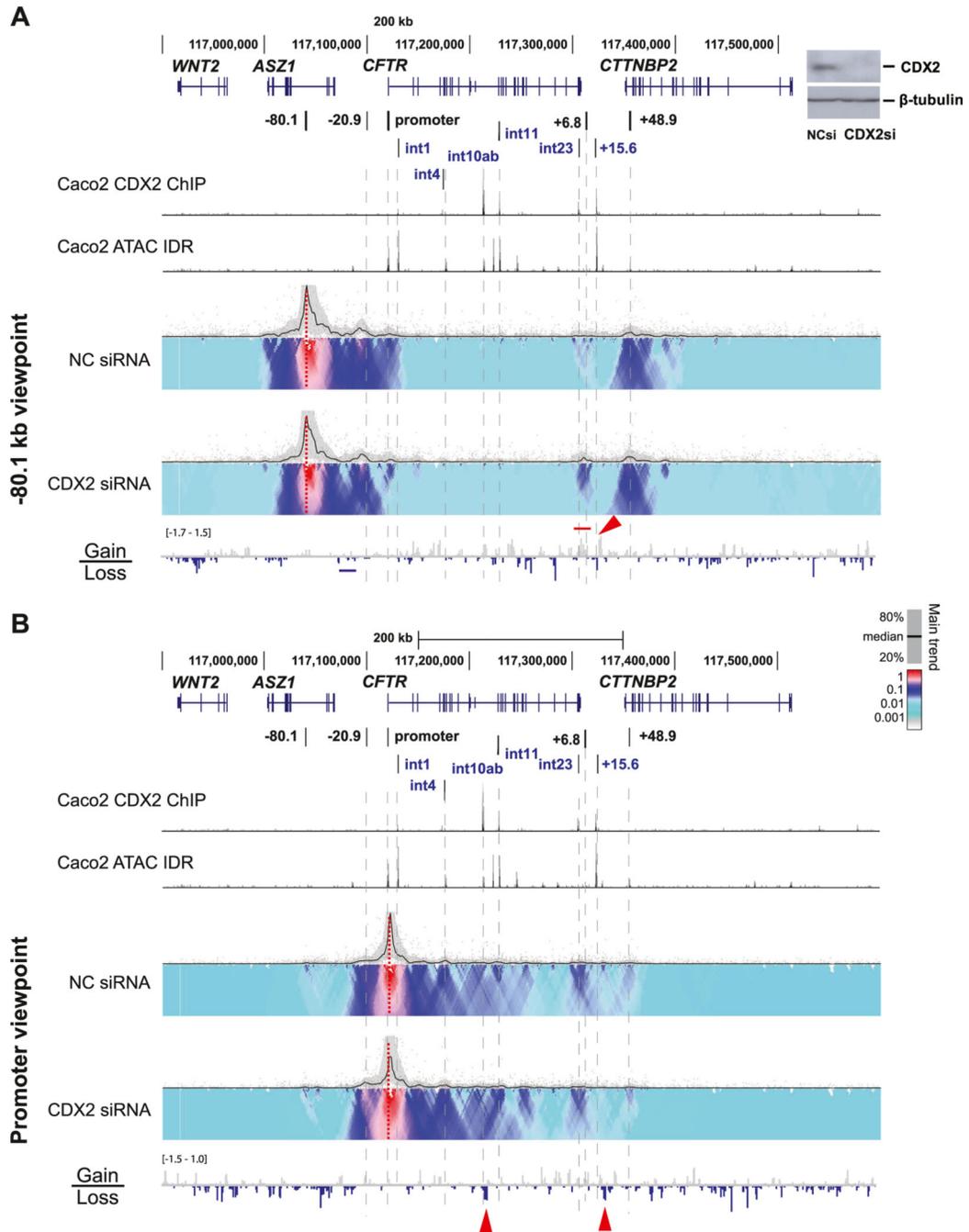


**Fig. 4.** Loss of the  $-80.1$  kb 5' TAD boundary or  $+15.6$  kb 3' insulator alters higher-order chromatin structure at the *CFTR* locus. Chromatin structure assayed by 4C-seq in (A, B) Del  $-80.1$  kb and (C) Del  $+15.6$  kb clones with viewpoints at the promoter (A),  $-20.9$  kb insulator (B) and  $-80.1$  kb TAD boundary (C). The sites of the deletions are marked by the red  $\Delta$ . All features of the figures and analysis protocols are as described for Fig. 2 except that blue arrows show features of interest referred to in the results.



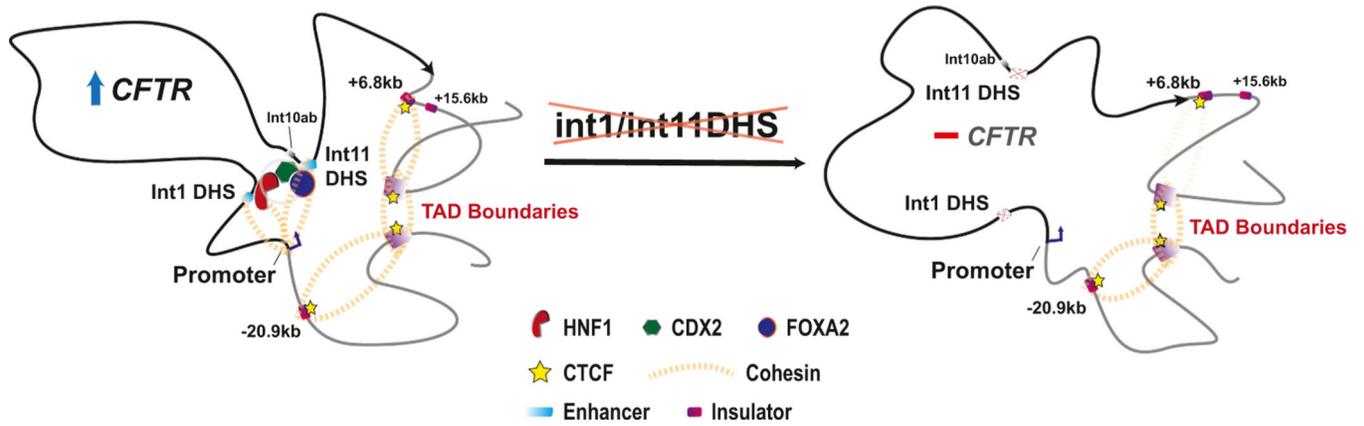
**Fig. 5.** Depletion of FOXA1/A2 by siRNA perturbs normal looping of CREs at the *CFTR* locus. A, B. Chromatin structure across the *CFTR* locus assayed by 4C-seq using viewpoints at (A) the -80.1 kb 5' TAD boundary or (B) the +15.6 kb insulator. A schematic at the top of each panel shows the *CFTR* locus on chromosome 7 and below key CREs. The inset panel shows FOXA1 and FOXA2 levels by western blot in NC siRNA or FOXA1/A2 siRNA treated cells. Caco2 ATAC-seq data show bigwig tracks, to which the 4C-seq data are aligned. Below are tracks showing ChIP-seq data for FOXA2 occupancy in Caco2 cells and DNase-

seq data for Caco2. For each panel two 4C-seq domainograms are shown with NC siRNA-treated cells above and FOXA1/A2 siRNA-treated cells below. Above each domainogram is a dotplot of all the post-filtered interactions pairs with the main trend of contact profiles using a sliding 5-kb window shown as a black line. The range between 20th and 80th percentiles is marked in grey flanking the median line. In the colored domainograms, relative interactions are normalized to the strongest interaction within each panel, using colour-coded intensity values to show relative interactions with window sizes varying from 2 to 50 kb. Here, red denotes the strongest interactions and dark blue, through turquoise, to grey represent gradually decreasing frequencies. Below each pair of domainograms a histogram shows alterations in the quantitative interaction frequencies (see Fig. 2 legend) in FOXA1/A2 siRNA-treated cells, subtracting the interaction frequencies in control-siRNA treated cells. Sites with value above zero (grey) are sites of gained interaction with the viewpoint after siRNA treatment and values below zero (blue) denote lost interactions. Red and blue arrowheads and red dotted lines denote features of interest described in the results.



**Fig. 6.** Depletion of CDX2 by siRNA had minor impact on normal looping of CREs at the *CFTR* locus. A, B. Chromatin structure across the *CFTR* locus assayed by 4C-seq using viewpoints at (A) the  $-80.1$  kb 5' TAD boundary or (B) the promoter. A schematic at the top of each panel shows the *CFTR* locus on chromosome 7 and below key CREs. The inset panel shows CDX2 levels by western blot in NC siRNA or CDX2 siRNA treated cells. Caco2 ATAC-seq data show bigwig tracks from the UCSC genome browser, to which the 4C-seq data are aligned. Below are tracks showing ChIP-seq data for CDX2 occupancy in Caco2 cells. For

each panel two 4C-seq domainograms are shown with NC siRNA-treated cells above and CDX2 siRNA-treated cells below. All other descriptions of the dotplots, domainograms and quantitative interaction tracks are as for the Fig. 5 legend.



**Fig. 7.** Model to show changes in higher order chromatin structure of the *CFTR* locus following deletion of the intron 1 and 11 DHS in Caco2 cells. In unperturbed Caco2 cells, there are strong looping interactions between the *CFTR* promoter, TAD boundaries, intron 1 and 11 DHS, and insulator elements at -20.9 kb and + 6.8 kb flanking the gene body. HNF1, CDX2, and FOXA2 TFs bind to the intronic CREs. Following deletion of both DHS, the close interactions between the intronic regions and 3' elements are lost and *CFTR* expression is abolished. Dissociation of CREs from the gene promoter may prevent recruitment of the enhancer-associated TFs and RNAPII.