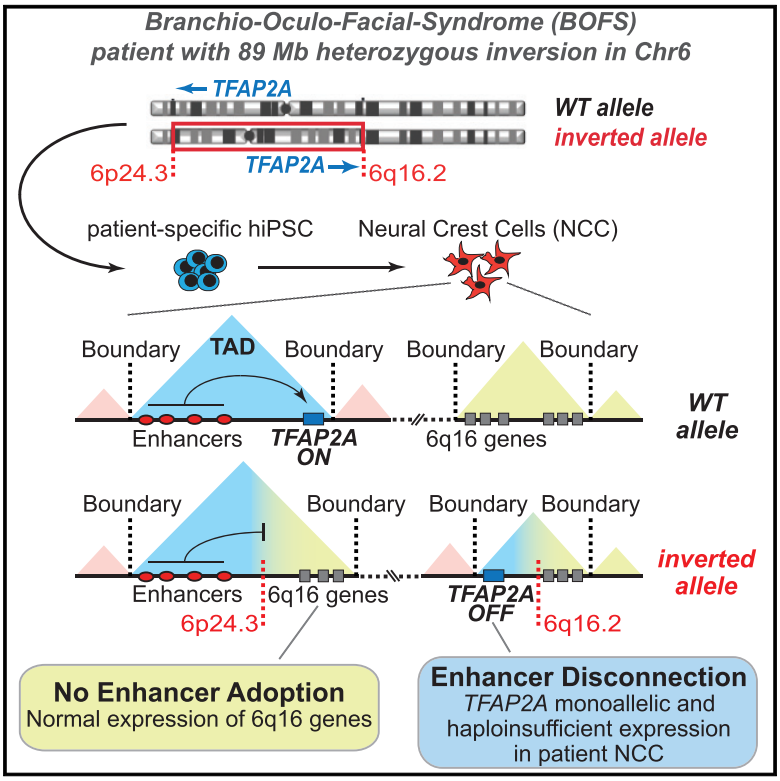
Article

Cell Stem Cell

Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs

Graphical Abstract



Highlights

- Disrupting the *TFAP2A* regulatory domain causes branchiooculofacial syndrome (BOFS)
- Enhancer disconnection leads to monoallelic and reduced *TFAP2A* expression
- Placing enhancers and genes within a TAD is not sufficient for enhancer adoption
- Patient-specific hiPSC-derived neural crest to uncover long-range pathomechanisms

Authors

Magdalena Laugsch, Michaela Bartusel, Rizwan Rehimi, ..., Katherine Lachlan, Julia Baptista, Alvaro Rada-Iglesias

Correspondence

julia.baptista@nhs.net (J.B.), aradaigl@uni-koeln.de (A.R.-I.)

In Brief

Elucidating the long-range pathomechanisms of structural variants *in vivo* is often difficult due to differences in gene dosage sensitivity between mice and humans. Laugsch et al. combined neural crest cells derived from patient-specific hiPSCs and functional genomics to demonstrate that an inversion causes BOFS by disconnecting *TFAP2A* from its enhancers.

Laugsch et al., 2019, Cell Stem Cell *24*, 736–752
May 2, 2019 © 2019 Elsevier Inc.
https://doi.org/10.1016/j.stem.2019.03.004

CellPress

CellPress

Cell Stem Cell
Article

Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs

Magdalena Laugsch,[1,2,13] Michaela Bartusel,[1,13] Rizwan Rehimi,[1] Hafiza Alirzayeva,[1] Agathi Karaolidou,[1] Giuliano Crispatzu,[1] Peter Zentis,[3] Milos Nikolic,[1] Tore Bleckwehl,[1] Petros Kolovos,[4] Wilfred F.J. van Ijcken,[5] Tomo Šarić,[6] Katrin Koehler,[7] Peter Frommolt,[3] Katherine Lachlan,[8,9] Julia Baptista,[10,11,*] and Alvaro Rada-Iglesias[1,3,12,14,*]

[1]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany
[2]Institute of Human Genetics, CMMC, University Hospital Cologne, Cologne, Germany
[3]Cluster of Excellence Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany
[4]Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark
[5]Erasmus Medical Center, University of Rotterdam, Rotterdam, the Netherlands
[6]Center for Physiology and Pathophysiology, Institute for Neurophysiology, Medical Faculty, University of Cologne, Cologne, Germany
[7]Department of Pediatrics, University Clinic Carl Gustav Carus, TU Dresden, Dresden, Germany
[8]Human Genetics & Genomic Medicine, University of Southampton, Southampton General Hospital, Southampton, UK
[9]Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Princess Anne Hospital, Southampton, UK
[10]Molecular Genetics Department, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
[11]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK
[12]Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), University of Cantabria, Cantabria, Spain
[13]These authors contributed equally
[14]Lead Contact
*Correspondence: julia.baptista@nhs.net (J.B.), aradaigl@uni-koeln.de (A.R.-I.)
https://doi.org/10.1016/j.stem.2019.03.004

SUMMARY

The pathological consequences of structural variants disrupting 3D genome organization can be difficult to elucidate *in vivo* due to differences in gene dosage sensitivity between mice and humans. This is illustrated by branchiooculofacial syndrome (BOFS), a rare congenital disorder caused by heterozygous mutations within *TFAP2A,* a neural crest regulator for which humans, but not mice, are haploinsufficient. Here, we present a BOFS patient carrying a heterozygous inversion with one breakpoint located within a topologically associating domain (TAD) containing enhancers essential for *TFAP2A* expression in human neural crest cells (hNCCs). Using patient-specific hiPSCs, we show that, although the inversion shuffles the *TFAP2A* hNCC enhancers with novel genes within the same TAD, this does not result in enhancer adoption. Instead, the inversion disconnects one *TFAP2A* allele from its cognate enhancers, leading to monoallelic and haploinsufficient *TFAP2A* expression in patient hNCCs. Our work illustrates the power of hiPSC differentiation to unveil long-range pathomechanisms.

INTRODUCTION

In order to control gene expression, enhancers establish long-range physical interactions with their target genes (Deng et al., 2012; Weintraub et al., 2017). Such contacts preferentially occur within large self-interacting genomic regions called topologically associating domains (TADs). TADs facilitate the communication between enhancers and their target genes and provide insulation by preventing the establishment of unwanted enhancer-gene interactions (Dixon et al., 2012; Lupiáñez et al., 2015; Nora et al., 2012; Schwarzer et al., 2017). The functional and medical relevance of TADs and the enhancers therein is supported by reports demonstrating that the disruption of enhancer activity or enhancer-gene communication can lead to human disease (Smith and Shilatifard, 2014; Spielmann et al., 2018). Structural variants (SVs) (i.e., deletions, duplications, inversions, and translocations) can disrupt the 3D architecture of TADs and cause pathological gains ("gain of function") and/or losses ("loss of function") in gene expression (Spielmann et al., 2018). Deletion and duplications spanning TAD boundaries can result in the "fusion" of neighboring TADs or in the appearance of "neo-TADs," respectively, which can lead to the emergence of ectopic enhancer-gene interactions (i.e., enhancer adoption or hijacking) and gains in gene expression (Franke et al., 2016; Gröschel et al., 2014; Lupiáñez et al., 2015). Translocations and inversions that cross TAD boundaries can result in the "shuffling" of TADs that are normally spatially separated, causing gains in gene expression by enhancer adoption or gene silencing through either position effect variegation (PEV) or loss of endogenous enhancer-gene interactions (i.e., enhancer disconnection; Goubau et al., 2013; Kleinjan and van Heyningen, 2005; Kleinjan et al., 2001; Lettice et al., 2002; Mehrjouy et al., 2018; Redin et al., 2017; Zepeda-Mendoza et al., 2018).

Based on the previous pathomechanisms, computational strategies were recently implemented to predict the pathological consequences of human SVs (Bianco et al., 2018; Weischenfeldt

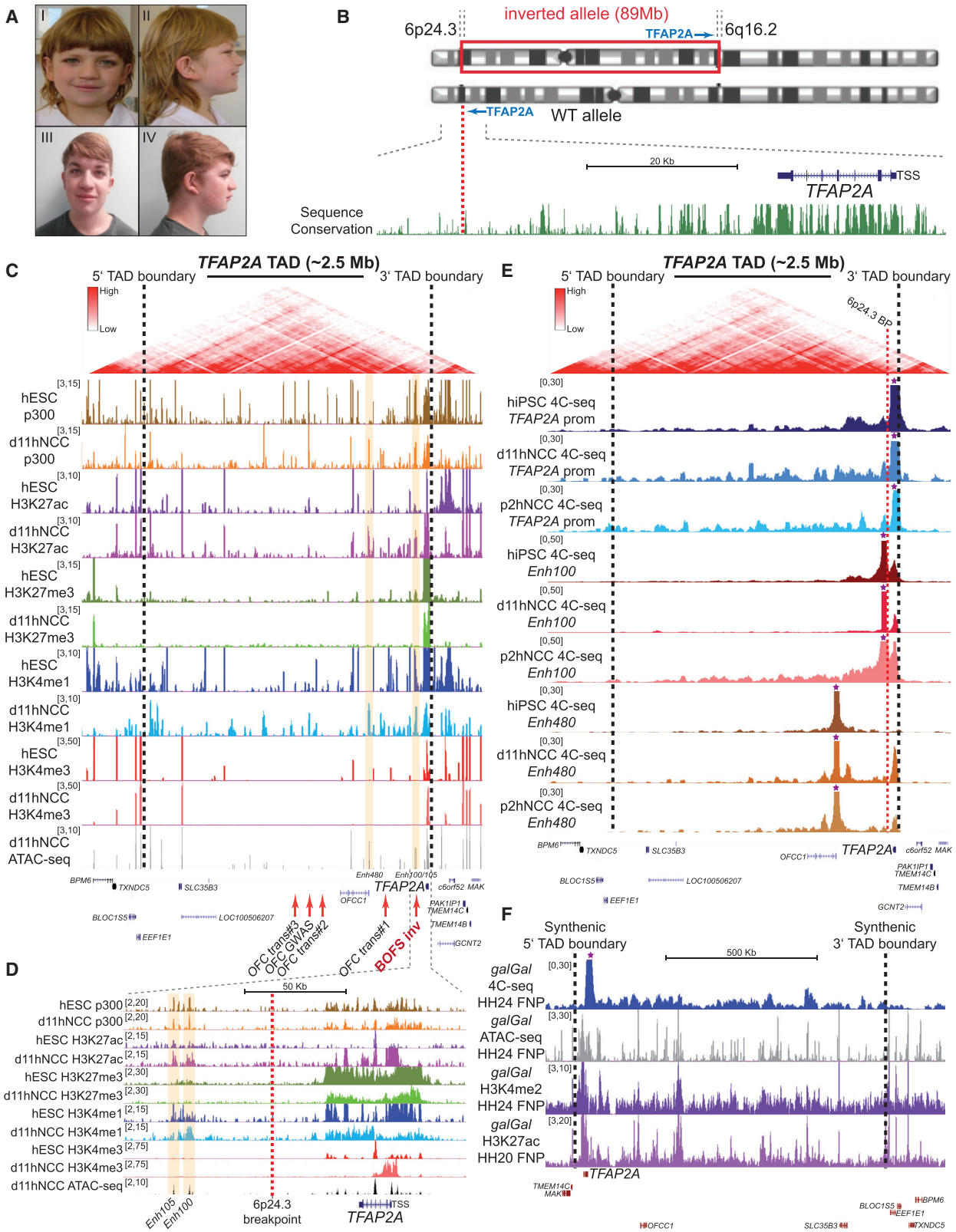


Figure 1. BOFS Patient with a *De Novo* Heterozygous Inversion that Can Potentially Alter the *TFAP2A* Regulatory Domain

(A) Facial appearance of the patient at 10 (I and II) and 17 (III and IV) years of age.

(B) Schematic view of the chr6 heterozygous inversion identified in the patient.

(legend continued on next page)

et al., 2017). Nevertheless, given the cell-type-specific activity of enhancers, these predictions need to be functionally validated in the relevant patient cell types or tissues, which, due to both ethical and technical difficulties, are often inaccessible. These limitations were recently overcome by using CRIPSR technology to recapitulate gain-of-function human SVs in mouse embryos and subsequently evaluate the downstream molecular and pathological consequences (Franke et al., 2016; Lupiáñez et al., 2015). However, these approaches are not always feasible in the case of loss-of-function SVs. Namely, differences in gene dosage sensitivity are frequently observed between mice and humans for developmental genes causally involved in human congenital disorders (i.e., humans, but not mice, are haploinsufficient for many developmental genes; Bedell et al., 1997; Wilkie, 2003). Therefore, certain loss-of-function SVs should be ideally investigated in the relevant human embryonic context. In addition, although CRISPR technology can be used to engineer SVs, the efficiencies at which this can be achieved are still rather low for certain variants, due to their large sizes or type of genomic re-arrangement (e.g., translocations).

The previous limitations (i.e., limited access to relevant patient material, differences in gene dosage sensitivity between mice and humans, and difficulties in recapitulating certain SVs) can be, in principle, overcome by combining the derivation of patient-specific human induced pluripotent stem cells (hiPSCs) with *in vitro* differentiation protocols toward specific human embryonic cell types (Shi et al., 2017). To illustrate this, here, we focus on branchiooculofacial syndrome (BOFS), a rare congenital syndrome characterized by a constellation of branchial, ocular, ear, and facial abnormalities (Milunsky et al., 2008, 2011). These abnormalities are likely to be caused by neural crest (NC) developmental defects, indicating that BOFS is a human neurocristopathy (Trainor, 2015). In agreement with this, all BOFS patients reported to date carry heterozygous mutations or deletions within *TFAP2A* that are predicted to result in either antimorphic, hypomorphic, or null alleles for this NC master regulator (Brewer et al., 2004; LeBlanc et al., 2013; Li et al., 2013; Milunsky et al., 2008, 2011). Furthermore, although *Tfap2a*-null mice display a BOFS-like phenotype, *Tfap2a*$^{+/-}$ animals appear morphologically normal and, thus, are not haploinsufficient (Brewer et al., 2004; Schorle et al., 1996; Zhang et al., 1996), highlighting the need to ideally investigate BOFS in a human embryonic context. However, due to its embryonic and migratory nature, the NC is a largely intractable cell population in humans. Here, we describe a unique BOFS patient without mutations within *TFAP2A* but with a *de novo* 89-Mb heterozygous inversion, in which one of the breakpoints is located downstream of this gene. Most importantly, combining the derivation of patient-specific hiPSCs with an *in vitro* human NC differentiation system (Bajpai et al., 2010; Prescott et al., 2015; Rada-Iglesias et al., 2012), we systematically evaluated the molecular consequences of this inversion. We conclusively showed that the inversion causes the physical disconnection between the inverted *TFAP2A* allele and its cognate enhancers, ultimately leading to monoallelic and haploinsufficient *TFAP2A* expression in human neural crest cells (hNCCs). Overall, our work provides a broadly applicable strategy to investigate the pathomechanisms of human SVs predicted to disrupt 3D genome organization.

RESULTS

Unique BOFS Patient with a *De Novo* Heterozygous Inversion in chr6 that Does Not Disrupt the *TFAP2A* Coding Sequence

Here, we describe a patient with a partial BOFS phenotype, which includes dysmorphic facial features and hearing and ocular abnormalities (Figure 1A; Table S1; STAR Methods). Sequencing of *TFAP2A* and comparative genomic hybridization (CGH) array analysis did not reveal any coding mutations or unbalanced chromosomal abnormalities, respectively (data not shown). However, cytogenetic analysis uncovered an 89-Mb *de novo* heterozygous inversion in chr6 (*46,XY inv(6)(p24.3;q16.2)*). Using DNA fluorescence *in situ* hybridization (FISH) and whole-genome sequencing (Figures S1A and S1B), the 6p24.3 and 6q16.2 inversion breakpoints were mapped to positions chr6:10355280 and chr6:99103873 (hg19), respectively. Interestingly, the 6p24.3 inversion breakpoint was located ~40 kb downstream of *TFAP2A* (Figure 1B), suggesting that the inversion could have a long-range regulatory effect on the expression of this gene (Kleinjan and van Heyningen, 2005; Spielmann et al., 2018).

Epigenomic and Topological Characterization of the *TFAP2A* Locus Regulatory Landscape in NCCs

To investigate the potential pathological consequences of the 89-Mb heterozygous inversion, we used an *in vitro* system whereby human embryonic stem cells (hESCs)/hiPSCs can be differentiated into hNCCs (Figure S1C; Bajpai et al., 2010; Prescott et al., 2015; Rada-Iglesias et al., 2012). Briefly, hESCs/hiPSCs are first differentiated into floating embryoid bodies with anterior neuroectodermal identity, which then attach to the tissue culture plate and give rise to migratory and proliferative cranial hNCCs. After ~11 days, a mixed population (referred to as *d11hNCC*) consisting of hNCCs and neuroectodermal

(C) Publically available Hi-C data from hESCs (Dixon et al., 2015), ChIP-seq data from hESCs and d11hNCCs (Rada-Iglesias et al., 2011, 2012) and newly generated d11hNCC ATAC-seq data are shown at the *TFAP2A* locus. The *TFAP2A*-TAD boundaries are shown as black dashed lines. Selected hNCC active enhancers located within the TFAP2A-TAD are highlighted in yellow: Enh480 is located ~480 kb downstream of *TFAP2A* and Enh100/105 denotes a couple of enhancers located ~100 and ~105 kb downstream of *TFAP2A*. The red arrows indicate the translocations breakpoints identified in three different OFC patients (Davies et al., 2004) and a risk-haplotype for OFC identified through GWAS (Yu et al., 2017).

(D) Close-up view of the previous epigenomic data centered on the 6p24.3 inversion breakpoint identified in our BOFS patient.

(E) 4C-seq profiles generated in WT hiPSCs (WT hiPSCs#1), WT d11hNCCs, and WT p2hNCCs using the *TFAP2A* promoter, Enh100, or Enh480 as viewpoints (purple stars).

(F) 4C-seq, ATAC-seq, H3K4me2 ChIP-seq, and H3K27ac ChIP-seq profiles generated in chick embryo FNP are shown at the *TFAP2A* chicken locus. The 4C-seq experiments were performed using a region located at the 3′ end of *TFAP2A* as a viewpoint (purple star). 4C-seq, ATAC-seq, and H3K4me2 ChIP-seq data were generated in FNP from stage HH24 chicken embryos, and the H3K27ac ChIP-seq data were previously generated in FNP from stage HH20 chicken embryos (Rada-Iglesias et al., 2012). The black dashed lines denote the approximate location of the human *TFAP2A*-TAD syntenic boundaries.

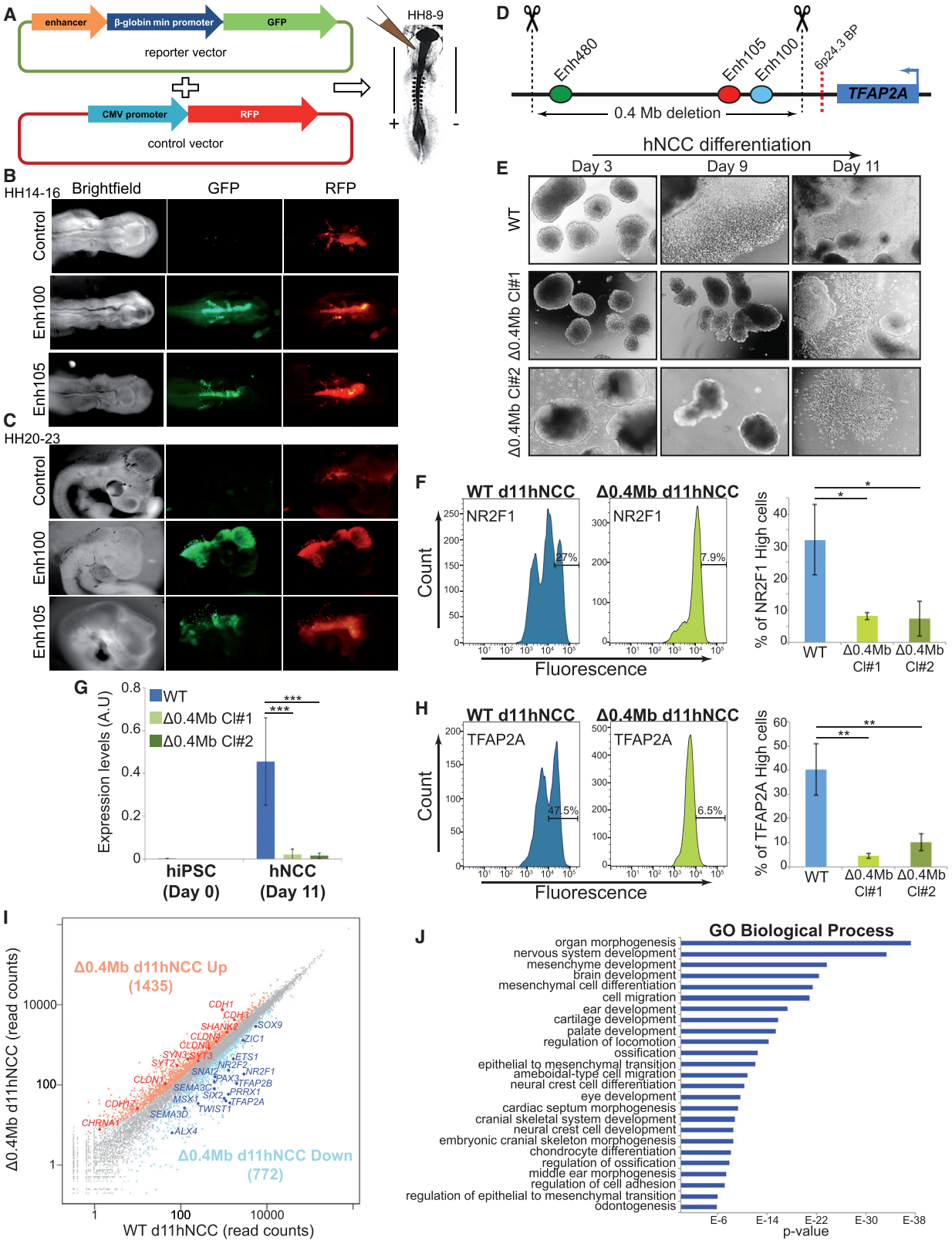


Figure 2. Functional Characterization of *cis*-Regulatory Elements Contained within the *TFAP2A*-TAD during NCC Development

(A) Schematic diagram illustrating the *in vivo* reporter assays performed in chicken embryos. Human enhancer sequences (Enh100, Enh105, and Enh480) were cloned into a reporter vector in which a minimal promoter (min β-globin promoter) drives GFP expression (reporter vector). Another vector in which a strong

(legend continued on next page)

progenitors can be obtained (Figures S1D and S1E). Subsequently, *d11hNCC* can be expanded for a couple of passages (Prescott et al., 2015), resulting in the elimination of neuroectodermal progenitors and yielding a more pure NCC population (referred to as *p2hNCC*; Figures S1E and S1F). The *p2hNCC* were further characterized by single-cell RNA sequencing (scRNA-seq), which confirmed that they represent a rather homogeneous cell population widely expressing multiple cranial NCCs and ectomesenchymal markers (Figure S1G).

To evaluate the *TFAP2A* regulatory landscape in hNCCs, we combined our previous epigenomic maps from hESCs and *d11hNCC* (Rada-Iglesias et al., 2012) with newly generated *d11hNCC* ATAC-seq data and publically available Hi-C data from hESCs (Dixon et al., 2012; Wang et al., 2018; Figures 1C and 1D). *TFAP2A* was located at the 3′ end of a large (>2.5 Mb) TAD (hereafter referred to as the *TFAP2A*-TAD) harboring many hNCC-active enhancers. Some of these hNCC-active enhancers were initially poised in hESCs (Rada-Iglesias et al., 2011; e.g., Enh100 and Enh105) and others became activated *de novo* in hNCCs (e.g., Enh480; Figures 1C and 1D). Interestingly, the 6p24.3 inversion breakpoint was located between *TFAP2A* and most of the hNCC-active enhancers identified within the *TFAP2A*-TAD (Figure 1D). Furthermore, the *TFAP2A*-TAD contains translocation breakpoints and risk haplotypes that were previously implicated in non-syndromic orofacial clefting (OFC) (Davies et al., 2004; Yu et al., 2017), a common congenital abnormality also considered as a neurocristopathy (Figure 1C). Therefore, the *TFAP2A-TAD* is likely to contain regulatory information that, when disrupted, can lead to NC-related abnormalities due to changes in *TFAP2A* expression. Next, we performed 4C-seq experiments in wild-type (WT) hiPSCs, *d11hNCC* and *p2hNCC* (Figure 1E), and using the *TFAP2A* promoter as a viewpoint, we confirmed the boundaries of the *TFAP2A*-TAD in hNCCs (Figure 1E). Furthermore, using Enh480 and Enh100 as viewpoints, strong and highly specific interactions with *TFAP2A* were observed in hNCCs (Figure 1E). In agreement with poised enhancers marked by H3K27me3 being able to interact with their target genes already in pluripotent cells (Cruz-Molina et al., 2017), Enh100-*TFAP2A* interactions were also detected in hiPSCs, although this was not the case for Enh480. In addition, we investigated the *TFAP2A* locus in NCC-derived facial mesenchyme isolated from the frontonasal prominences (FNPs) of chicken embryos (Figure 1F). 4C-seq experiments using the chicken *TFAP2A* gene as a viewpoint revealed an evolutionary conserved *TFAP2A*-TAD with syntenic boundaries to those observed in hNCCs (Figure 1F). Moreover, ATAC-seq, H3K27ac, and H3K4me2 profiles generated in the FNPs revealed that the chicken *TFAP2A*-TAD also contained many active enhancers (Figure 1F). Overall, the epigenomic and topological features of the *TFAP2A* regulatory domain appear to be conserved *in vivo*, thus supporting its physiological relevance (Dixon et al., 2012; Gómez-Marín et al., 2015).

The *TFAP2A*-TAD Contains Regulatory Elements that Are Essential for *TFAP2A* Expression in hNCCs

Next, we assessed whether the *TFAP2A*-TAD contained *cis*-regulatory information located 5′ of the 6p24.3 inversion breakpoint that was required for proper *TFAP2A* expression in hNCCs. First, we tested the activity of the previously described hNCC-active enhancers (Enh100, Enh105, and Enh480) using *in vitro* and *in vivo* reporter assays. According to the *in vitro* reporter assays, Enh100 and Enh105, but not Enh480, induced strong GFP expression in hNCC outgrowths (Figures S2A–S2C). Similarly, *in vivo* reporter assays in chicken embryos showed strong GFP expression in the dorsal neural tube, migrating cranial NCC and facial mesenchyme for Enh100 and Enh105 (Figures 2A–2C and S2D), although weaker expression within the dorsal neural tube was observed for Enh480 (Figure S2E).

promoter (cytomegalovirus [CMV] promoter) drives red fluorescent protein (RFP) expression was used as a positive control for electroporation (control vector). Reporter and control vectors were co-electroporated into the anterior neural tube of stage HH8-9 chick embryos to evaluate GFP and RFP expression levels at later developmental stages.

(B and C) Representative images of stage HH14-16 (B) and stage HH20-23 (C) chick embryos co-electroporated with the control vector (RFP) and reporter vectors (GFP) that were either empty or contained the Enh100 or Enh105 human sequences. In (B) and (C), the chick embryos are shown in a dorsal and lateral view, respectively.

(D) Schematic diagram illustrating the 0.4-Mb deletion generated within the *TFAP2A*-TAD in WT hiPSCs (WT hiPSCs#1).

(E) Representative images illustrating the compromised hNCC differentiation capacity of two different clonal hiPSC lines homozygous for the 0.4-Mb deletion (Δ0.4Mb Cl#1; Δ0.4Mb Cl#2) in comparison with their parental WT hiPSCs (WT hiPSCs#1).

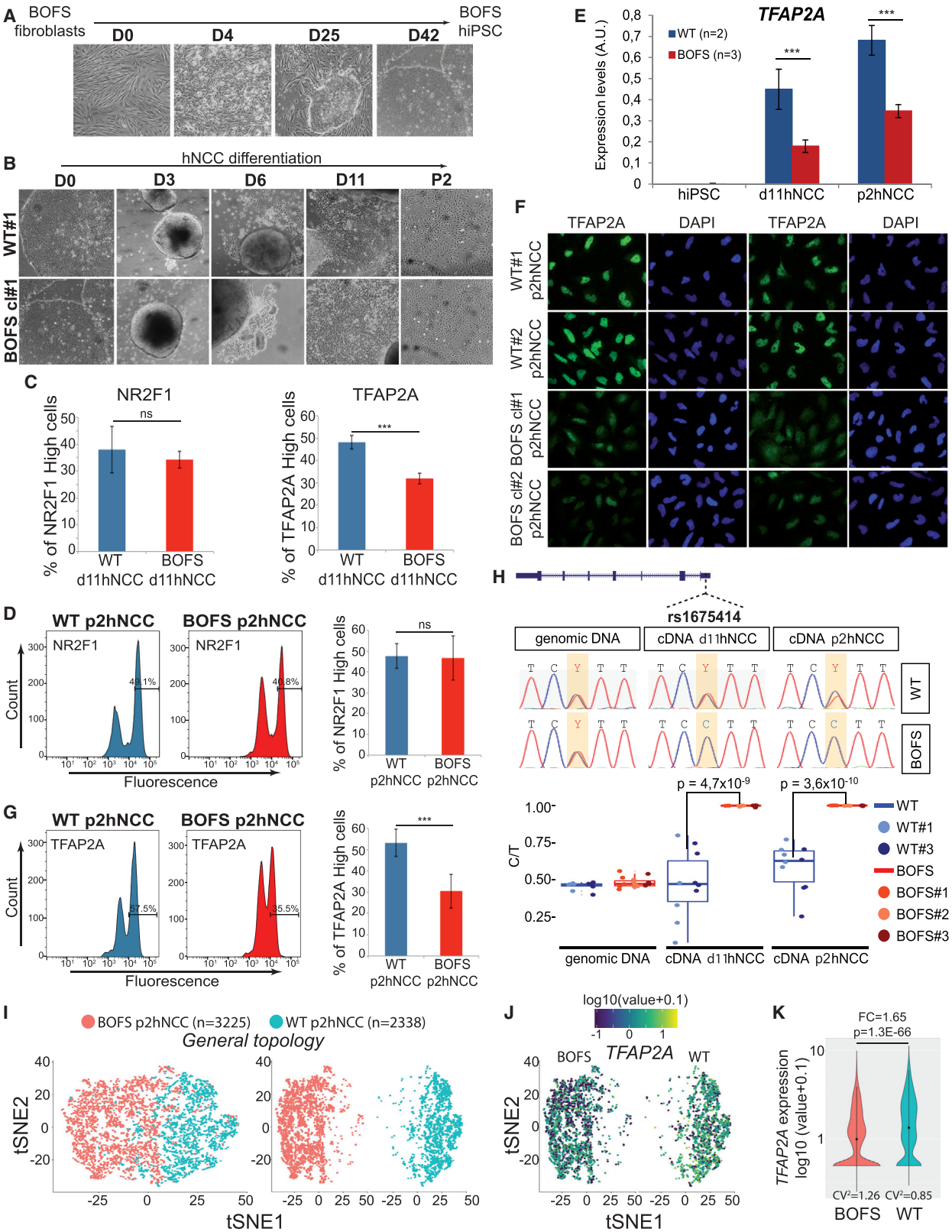
(F) Δ0.4Mb Cl#1 hiPSCs, Δ0.4Mb Cl#2 hiPSCs, and their isogenic WT hiPSCs line (WT hiPSCs#1) were differentiated into hNCCs, and at day 11 (d11hNCCs), the % of cells expressing high NR2F1 levels was quantified by flow cytometry (representative experiments are shown in the left and middle). The average results from six biological replicates for WT d11hNCCs, four biological replicates for Δ0.4Mb Cl#1 d11hNCCs, and four biological replicates for Δ0.4Mb Cl#2 d11hNCCs are presented in the right.

(G) *TFAP2A* expression levels were measured by qRT-PCR in Δ0.4Mb hiPSCs (Δ0.4Mb Cl#1; Δ0.4Mb Cl#2) and their isogenic WT hiPSCs (WT hiPSCs#1) at day 0 and after 11 days of hNCC differentiation (d11hNCC). Expression values were normalized to three housekeeping genes (*EEF2*, *GAPDH*, and *ACTB*) using the following number of measurements: twelve measurements in WT hiPSCs (technical triplicates in four different biological replicates), six measurements in Δ0.4Mb Cl#1 hiPSCs (technical triplicates in two different biological replicates), six measurements in Δ0.4Mb Cl#2 hiPSCs (technical triplicates in two different biological replicates), twelve measurements in WT d11hNCCs (technical triplicates in four different biological replicates), six measurements in Δ0.4Mb Cl#1 d11hNCCs (technical triplicates in two different biological replicates), and nine measurements in Δ0.4Mb Cl#2 d11hNCCs (technical triplicates in three different biological replicates).

(H) WT hiPSCs, Δ0.4Mb Cl#1 hiPSCs, and Δ0.4Mb Cl#2 hiPSCs were differentiated into hNCCs, and at day 11 (d11hNCCs), the % of cells expressing high TFAP2A protein levels was quantified by flow cytometry. The average results from six biological replicates for WT d11hNCCs, four biological replicates for Δ0.4Mb Cl#1 d11hNCCs, and four biological replicates for Δ0.4Mb Cl#2 d11hNCCs are presented in the right.

(I) Human genes plotted according to the average normalized RNA-seq read counts in WT d11hNCCs and Δ0.4Mb d11hNCCs. Genes considered as significantly up and downregulated in Δ0.4Mb d11hNCCs are shown in red and blue, respectively.

(J) Genes considered as significantly downregulated in Δ0.4Mb d11hNCCs were annotated according to Gene Ontology biological process terms. Selected terms among those with a $p < 10^{-5}$ are shown. In (F)–(H), p was calculated using t tests (*$p < 0.01$; **$p < 0.001$; ***$p < 0.0001$) and error bars represent SDs.



(legend on next page)

To more directly evaluate the functional relevance of the previous enhancers, we used CRISPR/Cas9 to engineer in WT hiPSCs a 0.4-Mb deletion 5′ of the 6p24.3 inversion breakpoint and spanning Enh100/105 and Enh480 (Figures 2D, S2F, and S2G). In comparison to their WT isogenic hiPSCs, hiPSCs homozygous for the 0.4-Mb deletion (Δ0.4Mb hiPSCs) showed a reduction in the attachment of floating EBs and diminished NCC outgrowths after 11 days of differentiation. Accordingly, the Δ0.4Mb hiPSCs yielded lower number of hNCCs as quantified by flow cytometry for several NCC markers (e.g., NR2F1 and p75; Prescott et al., 2015; Rada-Iglesias et al., 2012; Figures 2E, 2F, and S2H). To evaluate whether the 0.4-Mb deletion irreversibly impaired or, alternatively, delayed hNCC specification, Δ0.4Mb hiPSCs were kept for up to 13–18 days in hNCC differentiation media. Although, in some differentiations, hNCC specification was strongly compromised (Figure S2I, orange bars), in others, there was a progressive increase in the number of hNCCs over time (Figure S2I, green bars), indicating that the Δ0.4Mb deletion delays, but not fully abrogates, hNCC specification (Tchieu et al., 2017). Nevertheless, the 0.4-Mb deletion dramatically reduced *TFAP2A* induction during hNCC formation (Figures 2G and 2H), even upon prolonged maintenance of Δ0.4Mb cells in hNCC differentiation media (Figure S2I). To more globally evaluate how this severe reduction in TFAP2A levels could affect the hNCC differentiation process, we performed RNA-seq experiments in WT and Δ0.4Mb cells after 11 days of hNCC differentiation (Figures 2I and 2J; Table S2). Differential gene expression analysis revealed a large number of dysregulated genes in the Δ0.4Mb cells (772 downregulated and 1,435 upregulated), including, as expected, a strong downregulation of *TFAP2A* (Figure 2I; Table S2). Importantly, none of the genes immediately flanking *TFAP2A* and the 0.4-Mb deletion were differentially expressed (Figure S2J; Table S2), which together with the previous 4C-seq results (Figure 1E) supports that the 0.4-Mb deleted region contains *TFAP2A*-specific enhancers. Moreover, the genes downregulated in the Δ0.4Mb cells included major regulators of NCC development involved in NCC specification (e.g., *ZIC1*, *MSX1*, and *PAX3*), epithelial-to-mesenchymal transition (EMT) (e.g., *TWIST1*, *SNAI2*, and *PRRX1*), migration (e.g., *SEMA3D* and *SEMA3C*), or craniofacial morphogenesis (e.g., *ALX4* and *SIX2*; Figures 2J, S2K, and S2L; Table S2). In contrast, the genes upregulated in the Δ0.4Mb cells included genes involved in the maintenance of epithelial identity and EMT inhibition (e.g., CDH1, CLDN3, and CLDN4; Figure 2I; Table S2). Therefore, the 0.4-Mb deleted region within the *TFAP2A*-TAD contains enhancers that are necessary for proper *TFAP2A* expression during hNCC development.

The 89-Mb Heterozygous Inversion Leads to Monoallelic and Reduced TFAP2A Levels in Patient hNCCs

The 89-Mb inversion in the BOFS patient could impair the interaction between the inverted *TFAP2A* allele and enhancers that are essential for its expression in hNCCs, thus resulting in reduced TFAP2A levels. A non-mutually exclusive possibility is that the inversion could place the inverted *TFAP2A* allele within a heterochromatic environment, thus leading to gene silencing due to PEV (Kleinjan and van Heyningen, 2005). To evaluate these potential pathomechanisms, we reprogrammed patient fibroblasts into hiPSCs (Seki et al., 2012; Figures 3A and S3A–S3D). The presence of the 89-Mb heterozygous inversion in the

Figure 3. BOFS Patient hNCCs Display *TFAP2A* Monoallelic Expression and Reduced and More Heterogeneous *TFAP2A* Expression Levels

(A) Reprogramming of fibroblasts from the BOFS patient into hiPSCs.

(B) WT hiPSCs (WT#1 hiPSCs) and BOFS hiPSCs (BOFS Cl#1 hiPSCs) were differentiated into hNCCs, and images were taken at the indicated time points during the differentiation.

(C) WT hiPSCs and BOFS patient hiPSCs were differentiated into d11hNCCs, and the % of cells expressing high NR2F1 (left) or high TFAP2A (right) levels was quantified by flow cytometry. The average results from seven biological replicates for the WT cells (three for WT#1, two for WT#2, and two for WT#3) and eight biological replicates for BOFS cells (four for BOFS Cl#1 and four for BOFS Cl#2) are presented for each protein.

(D) The percentage of cells expressing high NR2F1 levels was quantified by flow cytometry in p2hNCCs derived from WT hiPSCs and BOFS patient hiPSCs. The average results from ten biological replicates for the WT cells (four for WT#1, two for WT#2, and four for WT#3) and five biological replicates for BOFS cells (three for BOFS Cl#1 and two for BOFS Cl#3) are presented on the right.

(E) *TFAP2A* expression levels were measured by qRT-PCR in d11hNCCs and p2hNCCs derived from two different WT hiPSC lines (WT hiPSCs#1 and WT hiPSCs#2) and three different BOFS hiPSC lines (BOFS hiPSCs Cl#1, BOFS hiPSCs Cl#2, and BOFS hiPSCs Cl#3). Expression values were normalized to two housekeeping genes (*EEF2* and *GAPDH*) using the following measurements: 10 measurements for WT cells (five technical replicates in two different cell lines) and 15 measurements for BOFS cells (five technical replicates in three different cell lines).

(F) TFAP2A protein levels were investigated by immunofluorescence in p2hNCCs derived from the indicated WT and BOFS hiPSC lines. Nuclei were stained with DAPI.

(G) The percentage of cells expressing high TFAP2A levels was quantified by flow cytometry in p2hNCCs derived from WT hiPSCs and BOFS patient hiPSCs. The average results from ten biological replicates for the WT cells (four for WT#1, two for WT#2, and four for WT#3) and five biological replicates for BOFS cells (three for BOFS Cl#1 and two for BOFS Cl#3) are presented on the right.

(H) Allele-specific expression of *TFAP2A* was measured by quantitative genotyping of a SNP (rs1675414) located within the first exon of *TFAP2A* and that was heterozygous (C/T) in the patient cells and in WT hiPSCs#1 and WT hiPSCs#3. The rs1675414 SNP was genotyped in genomic DNA (gDNA) and cDNA from hNCCs (d11hNCCs and p2hNCCs) derived from two WT (WT hiPSCs#1 and WT hiPSCs#3) and three BOFS patient (BOFS hiPSCs Cl#1, BOFS hiPSCs Cl#2, and BOFS hiPSCs Cl#3) hiPSC lines. A representative chromatogram illustrating the genotyping of rs1675414 in gDNA and cDNA from both WT and BOFS hNCCs is shown in the upper part. The overall quantification of the genotyping results is shown as boxplots in the lower part. For each cell line, five genotyping measurements were performed in gDNA and cDNA obtained from at least two independent hNCC differentiations.

(I) scRNA-seq data were generated from 2,238 WT p2hNCCs and 3,225 BOFS p2hNCCs. WT (blue) and BOFS (red) p2hNCCs were visualized by tSNE (t-distributed stochastic neighbor embedding) on the same (left) or separate (right) plots using the same tSNE parameters.

(J) tSNE plot showing the expression of *TFAP2A* in the WT and BOFS p2hNCCs.

(K) scRNA-seq expression data for *TFAP2A* in WT and BOFS p2hNCCs are shown as violin plots. CV^2, squared coefficient of variation; FC, fold-change in the average expression of *TFAP2A* between WT and BOFS p2hNCC; p was calculated using the likelihood ratio test from Monocle (Qiu et al., 2017).

In (C)–(E) and (G), p values were calculated using t tests (***$p < 0.0001$; ns, not significant) and error bars represent SDs.

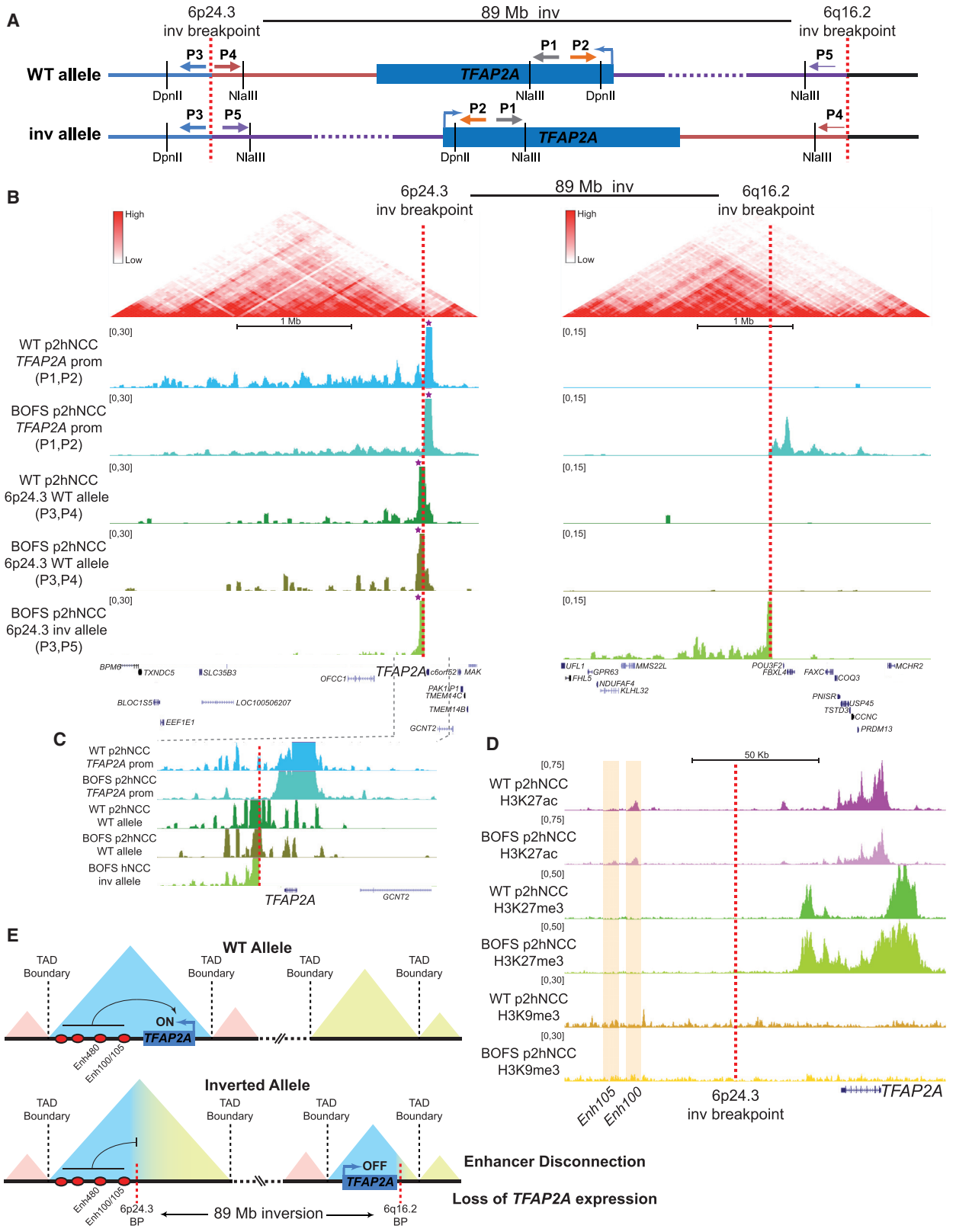


Figure 4. The Patient Inversion Physically Disconnects the Inverted *TFAP2A* Allele from Its Cognate Enhancers in hNCCs

(A) Primers used in the 4C-seq experiments performed in patient hNCCs. The P1/P2 primer pair ("*TFAP2A* promoter" viewpoint) does not distinguish between the WT and inverted alleles. The P3/P4 and P3/P5 primers pairs ("6p24.3 BP" viewpoint) are specific for the WT and inverted alleles, respectively.

(legend continued on next page)

patient hiPSCs was confirmed by PCR, Sanger sequencing, and targeted locus amplification (TLA) technology (de Vree et al., 2014; Figures S3B, S3E, and S3F; Table S3). Upon differentiation of the patient hiPSCs into hNCCs, we did not observe any notable differences in comparison to several WT hiPSCs lines used as controls (Figures 3B–3D, S3G, and S3H). Furthermore, we were able to similarly differentiate WT and patient hNCCs into cartilage, smooth muscle, and neurons, indicating that the patient hNCCs retain their multipotency and differentiation potential (Figures S3I–S3K). Nevertheless, *TFAP2A* expression levels were consistently reduced in patient hNCCs by ~50% with respect to WT hNCCs (Figures 3C and 3E–3G). To test whether the reduced *TFAP2A* expression in the patient hNCCs was a consequence of the 89-Mb inversion, we identified one SNP (rs1675414, C/T) within the first exon of *TFAP2A* that was heterozygous in both the BOFS patient and two of the WT hiPSCs lines, thus enabling us to measure *TFAP2A* expression in an allele-specific manner (Figure 3H). Furthermore, the data generated by TLA allowed us to phase rs1675414 in the BOFS patient cells with respect to the inversion (C in WT allele; T in inversion allele; Table S2; de Vree et al., 2014). Then, rs1675414 was quantitatively genotyped using cDNA from patient and WT hNCCs. Notably, although bi-allelic *TFAP2A* expression was observed in WT hNCCs, the patient hNCCs exclusively expressed the WT *TFAP2A* allele (i.e., C for rs1675414; Figures 3H and S5).

Although the bulk TFAP2A expression levels were reduced by ~50% in the patient p2hNCCs (Figures 3E–3G), we noticed that TFAP2A protein levels were variable between individual cells (Figures 3F and 3G). To evaluate whether this heterogeneity in *TFAP2A* expression was detectable at a transcriptional level, we extended the previous scRNA-seq analysis of WT p2hNCCs (Figure S1G) to BOFS p2hNCCs (Figures 3I–3K). We found that the reduced *TFAP2A* levels in BOFS p2hNCCs were the result of a significant increase in the number of BOFS p2hNCCs in which *TFAP2A* was lowly expressed, although in a few BOFS p2hNCCs, *TFAP2A* expression was high and comparable to that observed in WT p2hNCCs (Figures 3J and 3K). Hence, *TFAP2A* monoallelic expression leads to not only reduced but also more heterogeneous (>CV^2) *TFAP2A* expression in the patient hNCCs.

Monoallelic *TFAP2A* Expression in the Patient hNCCs Is Caused by the Physical Disconnection between the Inverted *TFAP2A* Allele and Its Cognate hNCC Enhancers

To test whether the 89-Mb inversion causes *TFAP2A* monoallelic expression by physically disconnecting the inverted *TFAP2A* allele from its cognate hNCC enhancers, we performed 4C-seq experiments in WT and patient p2hNCCs (Figures 4A–4C). Using the *TFAP2A* promoter as a viewpoint, the BOFS p2hNCCs showed extensive contacts not only within the *TFAP2A*-TAD, as observed in WT p2hNCCs, but also 3′ of the 6q16.2 inversion breakpoint (Figures 4B and 4C). However, these 4C-seq experiments cannot discriminate between the WT and inverted alleles. Therefore, we performed allele-specific 4C-seq experiments in the patient hNCCs using as a viewpoint the restriction fragment containing the 6p24.3 breakpoint and primer combinations that were specific for either the WT or inverted alleles (Figures 4A–4C; Franke et al., 2016). The WT allele displayed physical contacts both 5′ and 3′ of the 6p24.3 breakpoint, including clear interactions with the *TFAP2A* gene (Figures 4B and 4C). In contrast, the inverted allele only showed interactions 5′ of the 6p24.3 and 6q16.2 breakpoints, but not with the *TFAP2A* gene (Figures 4B and 4C), conclusively demonstrating a loss of physical proximity between the inverted *TFAP2A* allele and its cognate hNCC enhancers.

The patient inversion could also lead to monoallelic *TFAP2A* expression through PEV (Kleinjan and van Heyningen, 2005). To test this possibility, chromatin immunoprecipitation sequencing (ChIP-seq) for H3K27ac, H3K27me3, and H3K9me3 (a heterochromatic histone modification mechanistically implicated in PEV; Tchasovnikarova et al., 2015) was performed in WT and BOFS p2hNCCs (Figures 4D and S4A–S4C). For H3K27ac and H3K27me3, the ChIP-seq profiles within *TFAP2A* were in full agreement with our gene expression analyses. Namely, H3K27ac levels were higher in WT than in patient hNCCs, although the opposite was true for H3K27me3 (Figures 4D and S4C). Furthermore, allele-specific analysis of the ChIP-seq data using the previously described heterozygous SNP (i.e., *rs1675414*) showed that, in the patient hNCCs, the WT and inverted *TFAP2A* alleles were exclusively enriched in H3K27ac and H3K27me3, respectively (Figure S4C; Table S3). On the other hand, ChIP-seq for H3K9me3 showed low and similar enrichments within *TFAP2A* in both WT and patient NCCs (Figure 4D). Therefore, rather than becoming constitutively heterochromatinized (i.e., marked with H3K9me3), the inverted *TFAP2A* allele retains the H3K27me3 marking already present in pluripotent cells (Figures 4D and 1D), establishes interactions with other H3K27me3-marked promoters (e.g., *POU3F2*) within its new TAD (Cruz-Molina et al., 2017; Figure S4B), and fails to get induced in hNCCs. Overall, this suggests that a loss of enhancer-gene communication rather than PEV causes *TFAP2A* monoallelic expression and reduced TFAP2A levels in the patient hNCCs (Figure 4E).

The Patient Inversion Does Not Lead to Enhancer Adoption or Ectopic Gains in Gene Expression in hNCCs

The patient inversion places genes from the 6q16.2 locus (6q16.2 genes: *GPR63*, *NDUFAF4*, and *MMS22L*) in proximity of the *TFAP2A* hNCC active enhancers within a shuffled TAD (Spielmann et al., 2018; Figure 5A). This could lead to an enhancer adoption mechanism whereby these genes interact with the *TFAP2A* hNCC enhancers and become mis-expressed

(B) 4C-seq profiles generated in WT p2hNCCs and BOFS p2hNCCs (derived from WT#1 hiPSCs and BOFS hiPSCs Cl#1, respectively) are shown around the 6p24.3 and 6q16.12 breakpoint loci.

(C) Close-up view of the 4C-seq profiles shown in (B) centered on the 6p24.3 inversion breakpoint.

(D) ChIP-seq profiles for H3K27ac, H3K27me3, and H3K9me3 in WT p2hNCCs and BOFS p2hNCCs (derived from WT#1 hiPSCs and BOFS hiPSCs Cl#1, respectively) around the 6p24.3 breakpoint.

(E) Schematic representation of how the BOFS patient inversion physically disconnects the inverted *TFAP2A* allele from its cognate hNCC enhancers.

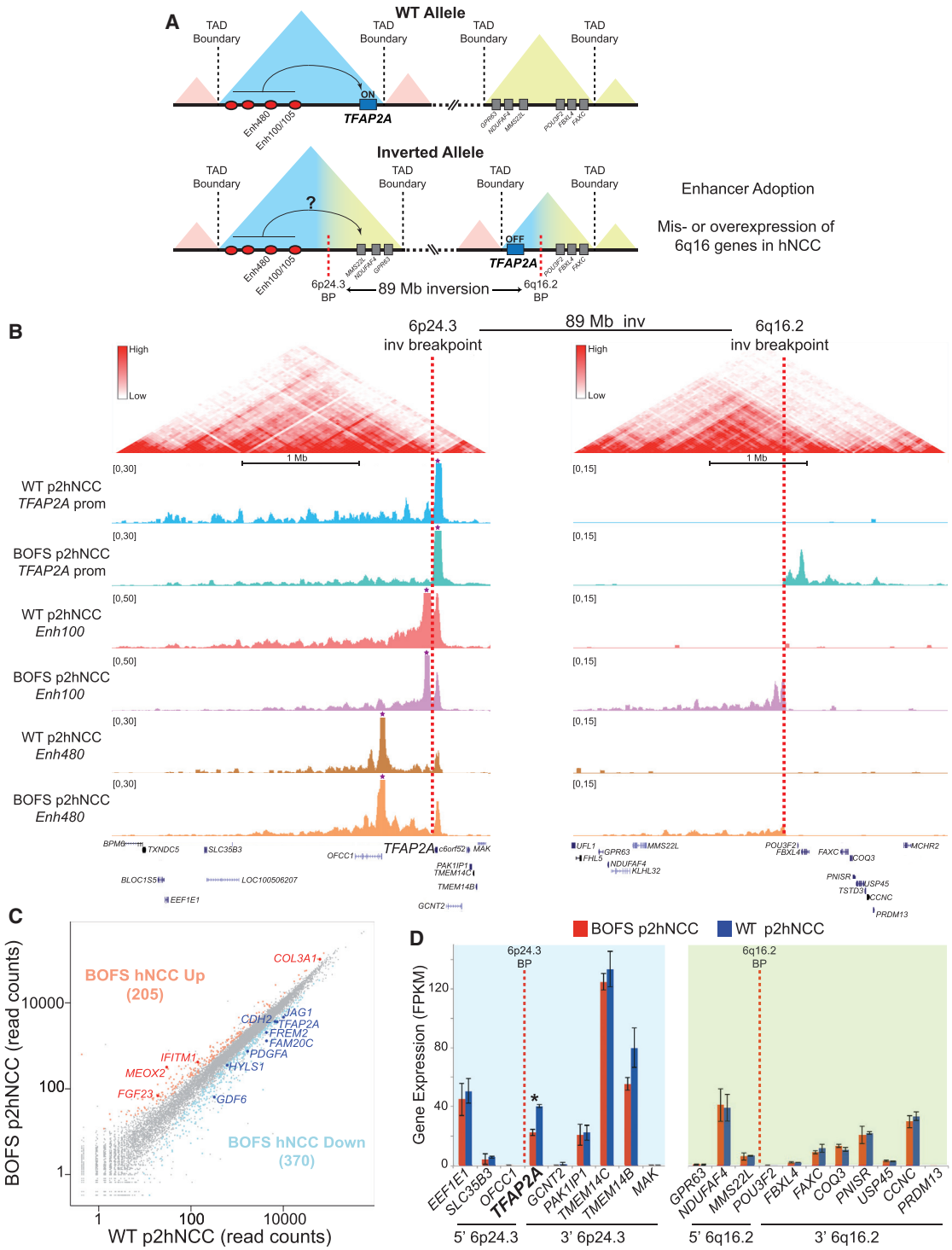


Figure 5. The Patient Inversion Does Not Cause an Enhancer Adoption Mechanism in hNCCs

(A) The BOFA patient inversion could lead to an enhancer adoption mechanism as genes within the 6q16.12 locus (i.e., *MMS22L, NDUFAF4*, and *GPR63*) are placed within a shuffled TAD together with the *TFAP2A* hNCC active enhancers.

(B) 4C-seq profiles generated in WT p2hNCCs and BOFS p2hNCCs (derived from WT#1 hiPSCs and BOFS hiPSCs Cl#1, respectively) are shown around the 6p24.3 and 6q16.12 breakpoint loci. 4C-seq experiments were performed using the *TFAP2A* promoter, the Enh100, and the Enh480 as viewpoints.

(C) Human genes plotted according to the average normalized RNA-seq read counts in WT p2hNCCs and BOFS p2hNCCs. Genes considered as significantly up and downregulated in BOFS p2hNCC are shown in red and blue, respectively.

(D) Average expression levels measured by RNA-seq in WT and BOFS p2hNCCs for genes immediately flanking the 6p24.3 and 6q16.2 inversion breakpoints. Error bars represent SDs. *Genes considered as differentially expressed.

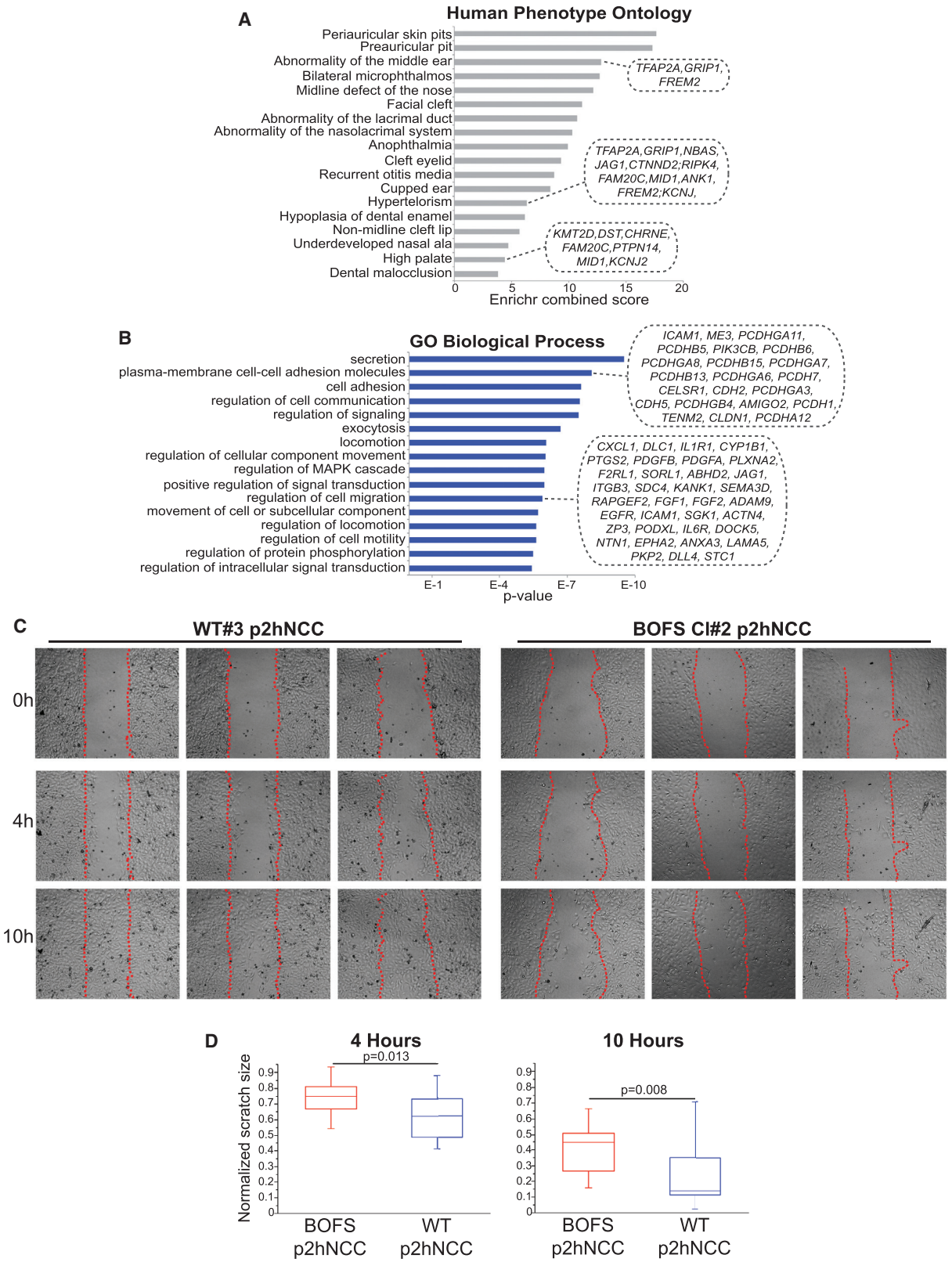


Figure 6. Transcriptional and Migratory Differences between WT and BOFS hNCCs

(A and B) Genes considered as significantly downregulated in BOFS p2hNCCs were annotated according to (A) human phenotype and (B) Gene Ontology biological process terms. Selected terms among the 50 most significantly overrepresented ones are shown.

(legend continued on next page)

in the patient hNCCs (Figure 5A). To test this possibility, we performed 4C-seq experiments in the patient p2hNCCs using the Enh100 and Enh480 as viewpoints (Figure 5B). These enhancers displayed extensive interactions not only within the *TFAP2A* TAD but also with regions located 5′ of the 6q16.2 inversion breakpoint (Figure 5B). However, these interactions declined with linear distance, and strong contacts were not observed between the enhancers and the 6q16.2 genes (Figure 5B). This argued against an enhancer adoption mechanism and, thus, against the mis-expression of the 6q16.2 genes in the patient hNCCs. To test this prediction, we generated RNA-seq data in WT and BOFS p2hNCCs and performed differential gene expression analysis (Figure 5C; Table S4). In addition to *TFAP2A*, which showed reduced and monoallelic expression in the patient p2hNCCs (Figures 5C and S5), a few hundred additional genes were differentially expressed between the WT and the BOFS p2hNCCs (Figure 5C). However, none of the genes flanking the 6p24.3 and 6q16.2 inversion breakpoints, except *TFAP2A*, was differentially expressed (Figure 5D). Therefore, BOFS in our patient is caused by *TFAP2A* haploinsufficiency and not by the ectopic activation of other genes flanking the inversion breakpoints.

BOFS Patient hNCCs Display a Defective Transcriptome and Reduced Migratory Capacity

To gain some insights into the molecular basis of *TFAP2A* haploinsufficiency, we analyzed in more depth the RNA-seq data generated in WT and BOFS p2hNCCs (Figures 6A, 6B, and S6A; Table S5). *In silico* functional annotation of the differentially expressed genes revealed that those downregulated in BOFS p2hNCCs were enriched in genes with major roles during craniofacial morphogenesis that, when mutated, result in typical BOFS phenotypes (e.g., periauricular skin pits, hypertelorism, and high palate; Figure 6A; Table S5). Furthermore, the genes downregulated in BOFS p2hNCCs were significantly enriched in cell adhesion and cell migration functions (Figure 6B; Table S5). Notably, *in vitro* migration assays in WT and BOFS p2hNCCs revealed a modest but reproducible reduction in the migratory ability of BOFS hNCCs (Figures 6C, 6D, and S6B). Moreover, these migratory defects were not accompanied by differences in proliferation rates between WT and BOFS p2hNCCs (Figures S6C and S6D). These results indicate that the heterozygous inversion in the patient leads to moderate, yet informative, transcriptional and phenotypic defects in hNCCs.

Gene Downregulation in BOFS Patient hNCCs Is Directly Mediated by Reduced TFAP2A Binding to hNCC Active Enhancers

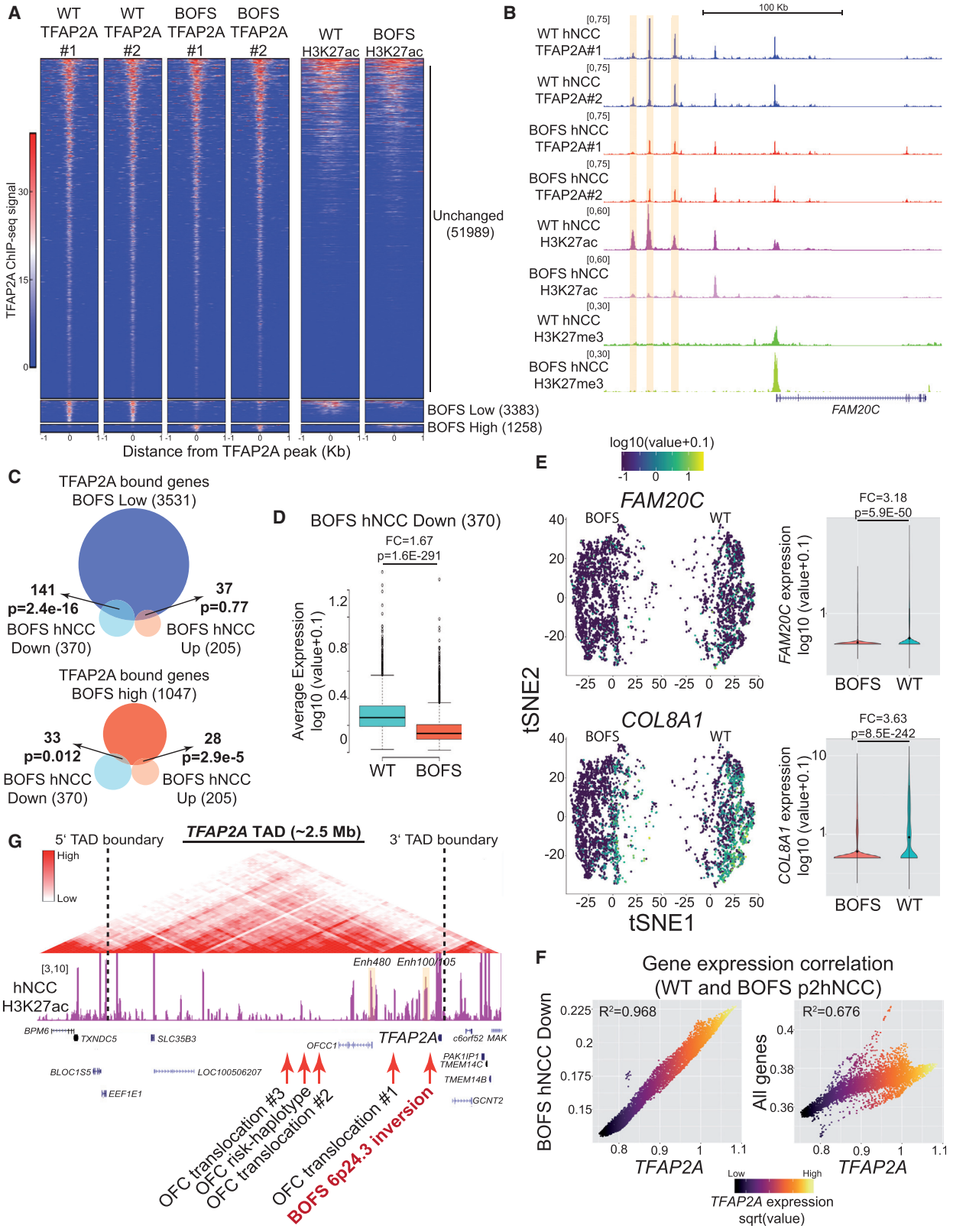
To determine whether the previous gene expression differences were directly mediated by TFAP2A, we performed ChIP-seq experiments for TFAP2A in both WT and BOFS p2hNCCs (Figures 7A and S7A). The overall TFAP2A binding patterns in WT and BOFS p2hNCCs were similar (Figures 7A, S7A, and S7B; Table S6). Nevertheless, a few thousand sites displayed differential TFAP2A binding between WT and BOFS p2hNCCs, with almost three times more sites displaying reduced binding in BOFS p2hNCCs (Figure 7A). Sites with reduced TFAP2A binding in BOFS p2hNCCs frequently overlapped distal enhancers that also showed lower H3K27ac levels in BOFS p2hNCCs (Figures 7A, 7B, and S7C–S7E; Table S6). The molecular basis of the differential TFAP2A binding in WT and BOFS p2hNCCs remains to be elucidated, but it is not due to differences in the underlying TFAP2A binding motifs (Figure S7F). After linking the differentially bound TFAP2A sites to nearby genes, we found a significant overlap between genes displaying reduced TFAP2A binding in BOFS p2hNCCs and genes downregulated in BOFS p2hNCCs ($p = 2.4e^{-16}$; Figures 7C and S7G). This suggests that TFAP2A directly controls the expression of the BOFS p2hNCC downregulated genes, which is also supported by our scRNA-seq data (Figures 7D–7F). Namely, a strong positive correlation was observed between the *TFAP2A* levels and the average expression of the BOFS downregulated genes present within individual p2hNCCs (Figure 7F). Overall, these results indicate that the downregulation of genes in BOFS hNCCs is directly mediated by reduced binding of TFAP2A to relevant enhancers. We conclude that this might be part of a general mechanism whereby reduced TFAP2A levels and/or activity in BOFS patients leads to transcriptional changes that affect the morphogenetic properties of hNCCs.

DISCUSSION

In addition to uncovering a long-range regulatory pathomechanism for BOFS, our extensive characterization of the *TFAP2A* regulatory domain provides major insights into the pathomechanisms of other neurocristopathies, including OFC (Ludwig et al., 2012). For example, a previous study reported three OFC patients with translocations breakpoints in the 6p24 region that could affect the expression and/or function of *OFCC1*, a poorly characterized gene located ~350 Kb downstream of *TFAP2A* (Davies et al., 2004; Figure 7G). However, this gene is barely expressed in hNCCs according to our RNA-seq data (Figures S2J and 5D), and *Ofcc1* null mice do not display any craniofacial abnormalities (Ohnishi et al., 2011). These observations argue against this gene playing any major role in the etiology of OFC. The translocation breakpoints of these three OFC patients are all located within the *TFAP2A-TAD*, 5′of the 6p24 inversion breakpoint identified in our BOFS patient (Figure 7G). Therefore, these translocations might disrupt the *TFAP2A* regulatory domain and, thus, *TFAP2A* expression in hNCCs. Moreover, as the translocated *TFAP2A* alleles would spare more hNCC-active enhancers than the inverted allele in our BOFS patient (e.g., Enh100/105 is spared in all three OFC patients; Figure 7G), the consequences for *TFAP2A* expression might be more subtle,

(C) Representative images of the scratch assays performed with p2hNCCs derived from WT hiPSCs and BOFS hiPSCs using cell culture inserts. Cell migration was monitored at the indicated times.

(D) The scratch assays described in (C) were quantified by measuring the outline of the scratched areas in images acquired at 0, 4, and 10 h. For each assay, the areas determined at 4 h and 10 h were normalized to the scratched area measured at 0 h. These normalized scratch sizes are shown as boxplots for measurements taken at 4 h (left) and 10 h (right). p was calculated using t tests, in which the normalized scratch sizes measured in WT p2hNCCs (n = 21 biological replicates) and BOFS p2hNCCs (n = 22 biological replicates) were compared.



(legend on next page)

perhaps explaining the milder phenotypes of these patients (i.e., OFC versus BOFS). Similarly, a recent genome-wide association study (GWAS) identified a susceptibility locus for OFC within the TFAP2A-TAD (Figure 7G; Yu et al., 2017). In this case, genetic variants within this OFC risk-haplotype might disrupt one or more hNCC-active enhancers and mildly affect *TFAP2A* expression.

The structural disruption of TADs is emerging as a major etiological mechanism for human congenital disorders (Lupiáñez et al., 2016; Spielmann et al., 2018). Using genetic-engineering tools, human SVs can be recapitulated in mouse embryos to unveil their pathomechanisms *in vivo* (Franke et al., 2016; Lupiáñez et al., 2015). However, these approaches might not be generally applicable to investigate SVs that, due to their large size (e.g., 89-Mb inversion in our BOFS patient) or type of genomic re-arrangement (e.g., translocations), cannot be efficiently engineered with current tools. Similarly, animal models are not always suitable to investigate the pathological consequences of balanced SVs (i.e., inversions and translocations) that, through a "TAD shuffling" mechanism, can lead to not only gains but also losses in gene expression (Spielmann et al., 2018). Namely, there are a large number of human congenital disorders, including BOFS, caused by heterozygous mutations within developmental genes for which humans, but not mice, are haploinsufficient (Bedell et al., 1997; Wilkie, 2003). These frequent differences in gene dosage sensitivity suggest that balanced SVs predicted to cause congenital abnormalities through loss-of-function regulatory mechanisms should be ideally investigated in a human embryonic context. However, due to technical and ethical reasons, the etiologically relevant embryonic cell types involved in human congenital disorders (e.g., NCCs in BOFS) are extremely difficult to isolate from human patients. Consequently, the exact long-range regulatory mechanisms whereby balanced SVs cause human congenital disease remain in many cases theoretical and mostly supported by correlative observations (Kleinjan and van Heyningen, 2005; Mehrjouy et al., 2018; Redin et al., 2017; Zepeda-Mendoza et al., 2018).

Using a BOFS patient as a case study, we conclusively show that patient-specific hiPSCs, *in vitro* differentiation systems, and genomic and genetic engineering tools can be combined to systematically evaluate the molecular consequences of otherwise non-tractable SVs. We demonstrate that the 89-Mb inversion found in our patient causes BOFS by physically disconnecting the inverted TFAP2A allele from its cognate enhancers, leading to *TFAP2A* monoallelic and haploinsufficient expression in hNCCs. Interestingly, using scRNA-seq, we found that the monoallelic expression of *TFAP2A* did not simply reduce by 2-fold the amount of *TFAP2A* transcripts present in each patient hNCC. Instead, *TFAP2A* expression became more heterogeneous and there was an increase in the number of patient hNCCs that barely expressed *TFAP2A*. We hypothesize that, if only one *TFAP2A* allele is functionally available, transcriptional bursts might occur less frequently (Fukaya et al., 2016; Larsson et al., 2019), resulting in a larger fraction of cells expressing low *TFAP2A* levels at any given time point. Our results suggest the intriguing possibility that haploinsufficiency for *TFAP2A* and other developmental genes might be caused by not only reduced gene expression levels but also increased transcriptional heterogeneity (Deng et al., 2014; Raj and van Oudenaarden, 2008). Lastly, we showed that, although our patient's inversion shuffles within the same TAD the *TFAP2A* hNCC enhancers with potentially novel target genes, this leads to neither enhancer adoption nor ectopic gains in gene expression. Therefore, our data show that placing enhancers and genes within the same TAD is not sufficient to drive gene expression, suggesting that, in certain genomic and/or cellular contexts, additional genetic and/or epigenetic mechanisms might be necessary to ensure that enhancer-gene communication is transcriptionally productive (van Arensbergen et al., 2014; Arnold et al., 2017; Kraft et al., 2019). Overall, our work highlights our still poor understanding of the regulatory rules governing enhancer-gene communication and enhancer responsiveness and suggests that the pathological consequences of SVs predicted to disrupt 3D genome organization should be experimentally validated. In this respect, our work provides a valuable experimental pipeline to uncover the molecular basis of human congenital disorders caused by the structural disruption of gene regulatory landscapes. This strategy should be particularly useful when animal models do not fully recapitulate the human phenotypes or when the relevant embryonic cell types are not easily accessible.

Figure 7. Genes Downregulated in Patient hNCCs Represent Direct TFAP2A Targets

(A) ChIP-seq data for TFAP2A and H3K27ac were generated in p2hNCCs derived from WT hiPSCs (WT hiPSCs#1) and BOFS hiPSCs (BOFS hiPSCs Cl#1). All the TFAP2A peaks (n = 56,630) identified in either WT or BOFS p2hNCCs were clustered in three groups, according to the differential TFAP2A binding between WT and BOFS p2hNCCs: unchanged; lower in BOFS phNCCs; and higher in BOFS p2hNCCs.

(B) ChIP-seq profiles for TFAP2A, H3K27ac, and H3K27me3 in WT and BOFS p2hNCCs are shown at a representative locus (i.e., *FAM20C* locus) that contains three hNCC active enhancers displaying reduced TFAP2A binding and H3K27ac levels in BOFS p2hNCCs.

(C) Overlaps between genes linked to differentially bound TFAP2A sites and genes downregulated (light blue) or upregulated (light red) in BOFS p2hNCCs. p was calculated using hypergeometric tests.

(D) The expression of genes downregulated in BOFS p2hNCCs according to bulk RNA-seq is shown as boxplots using the scRNA-seq data generated in WT and BOFS p2hNCCs. FC, fold-change in the average expression between WT p2hNCCs and BOFS p2hNCCs; p was calculated using t tests.

(E) The expression of two genes considered as BOFS low according to ChIP-seq and downregulated in BOFS p2hNCCs according to bulk RNA-seq analysis (*FAM20C* and *COL8A1*) is shown as tSNE or violin plots using the scRNA-seq data generated in WT and BOFS p2hNCCs. FC, fold-change in the average expression between WT and BOFS p2hNCCs; p was calculated using likelihood ratio tests.

(F) p2hNCCs analyzed by scRNA-seq are shown as dots in which the expression of TFAP2A (x axis) is compared to the average expression of the genes downregulated in BOFS p2hNCCs (y axis). The scRNA-seq data were subject to library normalization, square root transformation, and dropout correction using MAGIC (van Dijk et al., 2018). R2, squared Pearson correlation coefficient.

(G) Schematic overview of the *TFAP2A*-TAD, with the red arrows indicating the translocations breakpoints identified in three OFC patients (Davies et al., 2004) as well as a OFC risk-haplotype identified by GWAS (Yu et al., 2017).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient description
  - WT hiPSCs
  - BOFS patient-specific hiPSCs
- METHOD DETAILS
  - Characterization of chromosomal abnormalities
  - Targeted Locus Amplification (TLA)
  - hiPSCs culture
  - Derivation of BOFS patient-specific hiPSCs
  - Engineering of the 0.4Mb deletion within TFAP2A-TAD in WT hiPSCs by CRISPR/Cas9
  - hNCC differentiation
  - Chicken Embryos
  - Isolation of chicken frontonasal prominences
  - Chromatin immunoprecipitation (ChIP)
  - Immunofluorescence (IF)
  - RNA extraction, cDNA preparation and RT-qPCR
  - Flow cytometry
  - *In vitro* reporter assays
  - *In vivo* reporter assays
  - Scratch assays
  - Proliferation assays
  - Allele-specific analysis of TFAP2A expression
  - RNA-seq
  - Single-cell RNA-seq (scRNA-seq)
  - ChIP-seq
  - 4C-seq
  - ATAC-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - TLA data analysis
  - RT-qPCR data analysis
  - Quantification of Scratch Assays
  - Quantification of Proliferation Assays
  - Quantification of TFAP2A allele-specific expression
  - RNA-seq analysis
  - scRNA-seq analysis
  - ChIP-seq analysis
  - 4C-seq analysis
  - Public genomic datasets and additional bioinformatics analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stem.2019.03.004.

ACKNOWLEDGMENTS

We are grateful to the patient and his family for participating in this study. We thank the Rada-Iglesias lab members for insightful comments and critical reading of the manuscript; Janine Altmüller for technical assistance with next-generation sequencing; the Exeter Sequencing Service facility for producing the whole-genome sequencing; Hana Lango Allen, Richard Caswell, and Sian Ellard for data analysis and helpful discussions; Leo Kurian's laboratory for sharing reagents; Institute of Molecular Haematology and Oncology at the University of Cologne, Germany for the cytogenetic analyses of WT and BOFS patient hiPSCs lines; and Maike Kreutzenbeck and Rebecca Dieterich for technical support in the derivation of WT hiPSC lines in the Tomo Šarić laboratory. The Tomo Šarić laboratory participates in the EBiSC project, which received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115582, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contributions. K.K.'s work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant agreement KO 3588/2-1. Work in the Rada-Iglesias laboratory was supported by CMMC intramural funding and grants from Fritz Thyssen Stiftung and Else Kröner Fresenius Stiftung.

AUTHOR CONTRIBUTIONS

M.L., M.B., J.B., and A.R.-I. conceived the study and designed the experiments. Experimental data were generated and analyzed by M.L., M.B., and J.B. with the help of H.A., A.K., R.R., and P.Z. Computational analyses were performed by G.C., M.N., T.B., P.K., W.F.J.v.I., and A.R.-I., with assistance from P.F. and P.Z. Patient diagnosis and collection of patient fibroblasts were performed by K.L. WT hiPSCs were provided by K.K. and T.Ś.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 26, 2018
Revised: January 3, 2019
Accepted: March 6, 2019
Published: April 11, 2019

REFERENCES

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

Arnold, C.D., Zabidi, M.A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., and Stark, A. (2017). Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. Nat. Biotechnol. *35*, 136–144.

Bajpai, R., Chen, D.A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.-P., Zhao, Y., Swigut, T., and Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. Nature *463*, 958–962.

Bedell, M.A., Largaespada, D.A., Jenkins, N.A., and Copeland, N.G. (1997). Mouse models of human disease. Part II: recent progress and future directions. Genes Dev. *11*, 11–43.

Bergsland, M., Ramsköld, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. Genes Dev. *25*, 2453–2464.

Bianco, S., Lupiáñez, D.G., Chiariello, A.M., Annunziatella, C., Kraft, K., Schöpflin, R., Wittler, L., Andrey, G., Vingron, M., Pombo, A., et al. (2018). Polymer physics predicts the effects of structural variants on chromatin architecture. Nat. Genet. *50*, 662–667.

Brewer, S., Feng, W., Huang, J., Sullivan, S., and Williams, T. (2004). Wnt1-Cre-mediated deletion of AP-2alpha causes multiple neural crest-related defects. Dev. Biol. *267*, 135–152.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W.F.J., Grosveld, F., Frommolt, P., Bazzi, H., and Rada-Iglesias, A. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised

Enhancer Function during Pluripotent Stem Cell Differentiation. Cell Stem Cell *20*, 689–705.e9.

Davies, S.J., Wise, C., Venkatesh, B., Mirza, G., Jefferson, A., Volpi, E.V., and Ragoussis, J. (2004). Mapping of three translocation breakpoints associated with orofacial clefting within 6p24 and identification of new transcripts within the region. Cytogenet. Genome Res. *105*, 47–53.

de Vree, P.J.P., de Wit, E., Yilmaz, M., van de Heijning, M., Klous, P., Verstegen, M.J.A.M., Wan, Y., Teunissen, H., Krijger, P.H.L., Geeven, G., et al. (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nat. Biotechnol. *32*, 1019–1025.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

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

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., Hudson, T.J., and Pastinen, T. (2005). Survey of allelic expression using EST mining. Genome Res. *15*, 1584–1591.

Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proc. Natl. Acad. Sci. USA *112*, 7542–7547.

Goubau, C., Devriendt, K., Van der Aa, N., Crepel, A., Wieczorek, D., Kleefstra, T., Willemsen, M.H., Rauch, A., Tzschach, A., de Ravel, T., et al. (2013). Platelet defects in congenital variant of Rett syndrome patients with FOXG1 mutations or reduced expression due to a position effect at 14q12. Eur. J. Hum. Genet. *21*, 1349–1355.

Griessinger, E., Anjos-Afonso, F., Vargaftig, J., Taussig, D.C., Lassailly, F., Prebet, T., Imbert, V., Nebout, M., Vey, N., Chabannon, C., et al. (2016). Frequency and Dynamics of Leukemia-Initiating Cells during Short-term Ex Vivo Culture Informs Outcomes in Acute Myeloid Leukemia Patients. Cancer Res. *76*, 2082–2086.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell *157*, 369–381.

Hamburger, V., and Hamilton, H.L. (1992). A series of normal stages in the development of the chick embryo. 1951. Dev. Dyn. *195*, 231–272.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kind, B., Koehler, K., Krumbholz, M., Landgraf, D., and Huebner, A. (2010). Intracellular ROS level is increased in fibroblasts of triple A syndrome patients. J. Mol. Med. (Berl.) *88*, 1233–1242.

Kleinjan, D.A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am. J. Hum. Genet. *76*, 8–32.

Kleinjan, D.A., Seawright, A., Schedl, A., Quinlan, R.A., Danes, S., and van Heyningen, V. (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. Hum. Mol. Genet. *10*, 2049–2059.

Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G., et al. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat. Cell Biol. *21*, 305–310.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44* (W1), W90–W97.

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

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature *565*, 251–254.

Laugsch, M., Rostovskaya, M., Velychko, S., Richter, C., Zimmer, A., Klink, B., Schröck, E., Haase, M., Neumann, K., Thieme, S., et al. (2016). Functional restoration of gp91phox-oxidase activity by BAC transgenesis and gene targeting in X-linked chronic granulomatous disease iPSCs. Mol. Ther. *24*, 812–822.

LeBlanc, S.K., Yu, S., and Barnett, C.P. (2013). 6p.24 microdeletion involving TFAP2A without classic features of branchio-oculo-facial syndrome. Am. J. Med. Genet. A. *161A*, 901–904.

Lettice, L.A., Horikoshi, T., Heaney, S.J.H., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. USA *99*, 7548–7553.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Sheridan, R., and Williams, T. (2013). Analysis of TFAP2A mutations in Branchio-Oculo-Facial Syndrome indicates functional complexity within the AP-2α DNA-binding domain. Hum. Mol. Genet. *22*, 3195–3206.

Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, H., Paul, A., Becker, J., Herberz, R., AlChawa, T., Nasser, E., et al. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat. Genet. *44*, 968–971.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Lupiáñez, D.G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: how alterations of chromatin domains result in disease. Trends Genet. *32*, 225–237.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mehrjouy, M.M., Fonseca, A.C.S., Ehmke, N., Paskulin, G., Novelli, A., Benedicenti, F., Mencarelli, M.A., Renieri, A., Busa, T., Missirian, C., et al. (2018). Regulatory variants of FOXG1 in the context of its topological domain organisation. Eur. J. Hum. Genet. *26*, 186–196.

Milunsky, J.M., Maher, T.A., Zhao, G., Roberts, A.E., Stalker, H.J., Zori, R.T., Burch, M.N., Clemens, M., Mulliken, J.B., Smith, R., and Lin, A.E. (2008). TFAP2A mutations result in branchio-oculo-facial syndrome. Am. J. Hum. Genet. *82*, 1171–1177.

Milunsky, J.M., Maher, T.M., Zhao, G., Wang, Z., Mulliken, J.B., Chitayat, D., Clemens, M., Stalker, H.J., Bauer, M., Burch, M., et al. (2011). Genotype-phenotype analysis of the branchio-oculo-facial syndrome. Am. J. Med. Genet. A. *155A*, 22–32.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial

partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Ohnishi, T., Yamada, K., Watanabe, A., Ohba, H., Sakaguchi, T., Honma, Y., Iwayama, Y., Toyota, T., Maekawa, M., Watanabe, K., et al. (2011). Ablation of Mrds1/Ofcc1 induces hyper-γ-glutamyl transpeptidasemia without abnormal head development and schizophrenia-relevant behaviors in mice. PLoS ONE *6*, e29499.

Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. Cell *163*, 68–83.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods *14*, 309–315.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell *11*, 633–648.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.-A., et al. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat. Genet. *49*, 36–45.

Rehimi, R., Nikolic, M., Cruz-Molina, S., Tebartz, C., Frommolt, P., Mahabir, E., Clément-Ziza, M., and Rada-Iglesias, A. (2016). Epigenomics-based identification of major cell identity regulators within heterogeneous cell populations. Cell Rep. *17*, 3062–3076.

Respuela, P., Nikolić, M., Tan, M., Frommolt, P., Zhao, Y., Wysocka, J., and Rada-Iglesias, A. (2016). Foxd3 promotes exit from naive pluripotency through enhancer decommissioning and inhibits germline specification. Cell Stem Cell *18*, 118–133.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schorle, H., Meier, P., Buchert, M., Jaenisch, R., and Mitchell, P.J. (1996). Transcription factor AP-2 essential for cranial closure and craniofacial development. Nature *381*, 235–238.

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

Seki, T., Yuasa, S., and Fukuda, K. (2012). Generation of induced pluripotent stem cells from a small amount of human peripheral blood using a combination of activated T cells and Sendai virus. Nat. Protoc. *7*, 718–728.

Shi, Y., Inoue, H., Wu, J.C., and Yamanaka, S. (2017). Induced pluripotent stem cell technology: a decade of progress. Nat. Rev. Drug Discov. *16*, 115–130.

Smith, E., and Shilatifard, A. (2014). Enhancer biology and enhanceropathies. Nat. Struct. Mol. Biol. *21*, 210–219.

Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat. Rev. Genet. *19*, 453–467.

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

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science *348*, 1481–1485.

Tchieu, J., Zimmer, B., Fattahi, F., Amin, S., Zeltner, N., Chen, S., and Studer, L. (2017). A modular platform for differentiation of human PSCs into all major ectodermal lineages. Cell Stem Cell *21*, 399–410.e7.

Thongjuea, S., Stadhouders, R., Grosveld, F.G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res. *41*, e132.

Trainor, P. (2015). Neural Crest and Placodes (Academic).

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

van Arensbergen, J., van Steensel, B., and Bussemaker, H.J. (2014). In search of the determinants of enhancer-promoter interaction specificity. Trends Cell Biol. *24*, 695–702.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell *174*, 716–729.e27.

Wagle, P., Nikolić, M., and Frommolt, P. (2015). QuickNGS elevates next-generation sequencing data analysis to a new level of automation. BMC Genomics *16*, 487.

Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. *19*, 151.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. Cell *171*, 1573–1588.e28.

Weischenfeldt, J., Dubash, T., Drainas, A.P., Mardin, B.R., Chen, Y., Stütz, A.M., Waszak, S.M., Bosco, G., Halvorsen, A.R., Raeder, B., et al. (2017). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat. Genet. *49*, 65–74.

Wilkie, A.O.M. (2003). Why study human limb malformations? J. Anat. *202*, 27–35.

Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., Meng, L., Wang, W., Song, Y., Cheng, Y., et al. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat. Commun. *8*, 14364.

Zepeda-Mendoza, C.J., Bardon, A., Kammin, T., Harris, D.J., Cox, H., Redin, C., Ordulu, Z., Talkowski, M.E., and Morton, C.C. (2018). Phenotypic interpretation of complex chromosomal rearrangements informed by nucleotide-level resolution and structural organization of chromatin. Eur. J. Hum. Genet. *26*, 374–381.

Zhang, J., Hagopian-Donaldson, S., Serbedzija, G., Elsemore, J., Plehn-Dujowich, D., McMahon, A.P., Flavell, R.A., and Williams, T. (1996). Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. Nature *381*, 238–241.

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

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-AP-2α (clone 3B5)	Santa Cruz Biotechnology	Cat#sc-12726; RRID: AB_667767
Rabbit polyclonal anti-H3K9me3	Active Motif	Cat#39161; RRID: AB_2532132
Rabbit polyclonal anti-H3K27ac	Active Motif	Cat#39133; RRID: AB_2561016
Rabbit polyclonal anti-H3K27me3	Active Motif	Cat#39155; RRID: AB_2561020
Rabbit polyclonal anti-H3K4me2	Active Motif	Cat#39141; RRID: AB_2614985
Goat polyclonal anti-SOX2 (clone Y-17)	Santa Cruz Biotechnology	Cat#sc-17320; RRID: AB_2286684
Mouse monoclonal anti-TRA-1-60	Santa Cruz Biotechnology	Cat#sc-21705; RRID: AB_628385
Mouse monoclonal anti-COUP-TF II/NR2F2 (clone H7147)	Perseus Proteomics	Cat#H7147; RRID: AB_2314222
Mouse monoclonal anti-COUP-TF I/NR2F1 (clone H8132)	R&D Systems	Cat#PP-H8132-00; RRID: AB_2155494
Rabbit polyclonal anti-TWIST1	Thermo Fisher Scientific	Cat#PA5-49688; AB_2635141; RRID: AB_2635141
Rabbit polyclonal anti-SOX9	Millipore	Cat#AB5535; RRID: AB_2239761
Rabbit polyclonal anti-SM22 alpha	Abcam	Cat#ab14106; RRID: AB_443021
Rabbit polyclonal anti-beta III Tubulin	Abcam	Cat#ab18207; RRID: AB_444319
Alexa Fluor® 488 goat anti-rabbit IgG (H+L)	Life Technologies	Cat#A11008; RRID: AB_143165
Alexa Fluor® 594 donkey anti-goat IgG (H+L)	Life Technologies	Cat#A11058; RRID: AB_2534105
Alexa Fluor® 594 goat anti-rabbit IgG (H+L)	Life Technologies	Cat#R37117; RRID: AB_2556545
Mouse monoclonal p75/CD271-PerCP (clone ME20.4)	Biolegend	Cat#345111; RRID: AB_11204078
Goat polyclonal anti-mouse F(ab)2 IgG (H+L) PerCP-conjugated	R&D Systems	Cat#F0114; RRID: AB_1207935
Goat polyclonal anti-mouse F(ab)2 IgG (H+L) PerCP-conjugated	Invitrogen	Cat#46-4010-82; RRID: AB_2573755
Goat polyclonal anti-rabbit IgG APC-conjugated	R&D Systems	Cat#F0111; RRID: AB_573127
DyLight 488 donkey anti-rabbit IgG (clone Poly4064)	Biolegend	Cat#406404; RRID: AB_1575130
Bacterial and Virus Strains		
E.Coli Top10	Kurian Lab	N/A
Biological Samples		
Patient-derived fibroblasts	Wessex Regional Genetics Laboratory, Salisbury, UK	N/A
Chemicals, Peptides, and Recombinant Proteins		
GelTrex	Thermo Fisher Scientific	Cat#A1413302
mTeSR1	Stem Cell Technologies	Cat#85850
StemMACS iPS-Brew XF	Miltenyl Biotec	Cat#130-104-368
Versene	Life Technologies	Cat#15040-066
Accutase	Sigma-Aldrich	Cat#A6964
Thiazovivin	Axon Med Chem	Cat#1535
FuGene HD transfection reagent	Promega	Cat#E2313
Human Plasma Fibronectin Purified Protein	Millipore	Cat#FC010
Neurobasal medium	Life Technologies	Cat#21103049
DMEM F12 medium	Life Technologies	Cat#10565018
B27 supplement	Life Technologies	Cat#17504044
N2 supplement	Life Technologies	Cat#17502048

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
epidermal fibroblast growth factor	Peprotech	Cat#AF-100-15
basic fibroblast growth factor	Peprotech	Cat#100-15-18B
Insulin	Sigma-Aldrich	Cat#I1882
BSA	Gemini Bio-Products	Cat#700-104P
MesenCult ACF Chondrogenic Differentiation Medium	StemCell Technologies	Cat#05455
Q5 High-Fidelity DNA Polymerase	New England Biolabs	Cat#M0491L
Platinum Taq DNA Polymerase	Life Technologies	Cat#10966034
T4 DNA ligase	Life Technologies	Cat#15224-041
TD buffer	Illumina	Cat#15027866
TDE1	Illumina	Cat#15027868
Critical Commercial Assays		
CytoTune-iPS 2.0 Sendai Reprogramming Kit	Invitrogen	Cat#A16517
TLA assay	Cergentis	https://www.cergentis.com/
CellTrace CFSE Cell Proliferation Kit	Invitrogen	Cat#C34554
innuPREP RNA Mini Kit	Analytic Jena	Cat#845-KS-2040250
ProtoScript II First Strand cDNA Synthesis Kit	New England Biolabs	Cat# E6560L
10x Single Cell 3′ Reagent Kit	Illumina	N/A
Expand Long Template PCR system	Roche	Cat#11681842001
QIAquick PCR Purification Kit	QIAGEN	Cat#28104
MinElute PCR Purification kit	QIAGEN	Cat#28004
TruSeq RNA Library Prep Kit v2	Illumina	RS-122-2001
Deposited Data		
ChIP-seq, RNA-seq and 4C-seq datasets	This paper	GEO: GSE108522
ChIP-seq Chicken HH20 frontonasal prominences: H3K27ac	Rada-Iglesias et al., 2012	GEO: GSM933349
ChIP-seq hESC: p300	Rada-Iglesias et al., 2011	GEO: GSM602291
ChIP-seq hESC: H3K27me3	Rada-Iglesias et al., 2011	GEO: GSM602293
ChIP-seq hESC: H3K27ac	Rada-Iglesias et al., 2011	GEO: GSM602294
ChIP-seq hESC: H3K4me1	Rada-Iglesias et al., 2011	GEO: GSM602295
ChIP-seq hESC: H3K4me3	Rada-Iglesias et al., 2011	GEO: GSM602296
ChIP-seq d11hNCC: p300	Rada-Iglesias et al., 2012	GEO: GSM714804
ChIP-seq d11hNCC: H3K27me3	Rada-Iglesias et al., 2012	GEO: GSM714806
ChIP-seq d11hNCC: H3K27ac	Rada-Iglesias et al., 2012	GEO: GSM714807
ChIP-seq d11hNCC: H3K4me1	Rada-Iglesias et al., 2012	GEO: GSM714808
ChIP-seq d11hNCC: H3K4me3	Rada-Iglesias et al., 2012	GEO: GSM714809
Human reference genome NCBI build 37, GRCh37/hg19	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
	International Chicken Genome Consortium	https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.1/
Chicken reference genome ICGSC 4.0/galGal4	International Chicken Genome Consortium	https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.3/
Hi-C data generated in hESC	Dixon et al., 2015	N/A
Experimental Models: Cell Lines		
Human: WT hiPSC line NP0040-8 (WT#1)	Saric lab, Center for Physiology and Pathophysiology, Institute for Neurophysiology, Medical Faculty, University of Cologne, Cologne, Germany	EBiSC accession name UKKi011-A; https://cells.ebisc.org/UKKi011-A/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human: WT hiPSC line S24 (WT#2)	Pediatric Endocrinology and Diabetology, Department of Pediatrics, University Clinic, Dresden, Germany	N/A
Human: WT hiPSC line T12 (WT#3)	Pediatric Endocrinology and Diabetology, Department of Pediatrics, University Clinic, Dresden, Germany	N/A
Human: Δ0.4Mb hiPSC	This paper	N/A
Human: BOFS hiPSC	This paper	N/A
Experimental Models: Organisms/Strains		
Chicken breed: white leghorn (*Gallus gallus domesticus*)	LSL Rhein-Main	N/A
Oligonucleotides		
Primers for RT-qPCR see Table S7	This paper	N/A
gRNA sequences for 0.4 Mb del CRISPR/Cas9 see Table S7	This paper	N/A
Primers for 0.4 Mb del PCR genotyping see Table S7	This paper	N/A
Primers for 4C sequencing see Table S7	This paper	N/A
Primers for Reporter Assays see Table S7	This paper	N/A
Primers for 89 Mb Inv PCR genotyping see Table S7	This paper	N/A
Primers for SNP genotyping see Table S7	This paper	N/A
Primers for TLA assay see Table S7	This paper	N/A
Recombinant DNA		
PiggyBac enhancer GFP Neo	Buecker et al., 2014	N/A
PiggyBac Enh100 GFP Neo	This paper	N/A
PiggyBac Enh105 GFP Neo	This paper	N/A
PiggyBac Enh480 GFP Neo	This paper	N/A
Super PiggyBac Transposase Expression Vector	System Biosciences	Cet#PB200PA-1
β-globin-GFP enhancer reporter system	Bergsland et al., 2011	N/A
β-globin-GFP Enh100 reporter	This paper	N/A
β-globin-GFP Enh105 reporter	This paper	N/A
β-globin-GFP Enh480 reporter	This paper	N/A
PiggyBac-CMV-MCS-EF1-RedPuro	System Biosciences	Cat#PB514B-2
pX330-hCas9-long-chimeric-grna-g2p	Kurian Lab; Cruz-Molina et al., 2017	N/A
pX330-hCas9-gRNA#34	This paper	N/A
pX330-hCas9-gRNA#L2	This paper	N/A
pX330-hCas9-gRNA#39	This paper	N/A
pX330-hCas9-gRNA#L5	This paper	N/A
Software and Algorithms		
Genome engineering toolbox	Zhang lab, MIT	http://www.genome-engineering.org/crispr/
BD FACSDIVA software	BD Bioscience	https://www.bdbiosciences.com/us/applications/s/flowcytometry
FloJo 7.6.5.	FlowJo, LLC	https://www.flowjo.com/
FIJI (ImageJ)	Schindelin et al., 2012	https://fiji.sc/
KNIME	KNIME AG. Zurich, Schwitzerland	https://www.knime.com/
OMERO	University of Dundee & Open Microscopy Environment	https://www.openmicroscopy.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PeakPicker2	Ge et al., 2005	N/A
FastQC	Babraham Bioinformatics	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
SAMtools	http://www.htslib.org	http://samtools.sourceforge.net
TopHat2	Kim et al., 2013	https://ccb.jhu.edu/software/tophat/index.shtml
DESeq2	Anders and Huber, 2010	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Cufflinks	Trapnell et al., 2010	https://github.com/cole-trapnell-lab/cufflinks
GREAT	McLean et al., 2010	http://bejerano.stanford.edu/great/public/html/
cellranger-2.1.0	Zheng et al., 2017	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
monocle_2.6.4	Qiu et al., 2017	http://cole-trapnell-lab.github.io/monocle-release/
Rmagic_1.3.0	van Dijk et al., 2018	https://github.com/KrishnaswamyLab/MAGIC
BWA	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
MACS2	Zhang et al., 2008	http://liulab.dfci.harvard.edu/MACS/
Bowtie	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/
R3C-seq	Thongjuea et al., 2013	http://bioconductor.org/packages/release/bioc/html/r3Cseq.html
3D Genome Browser	Wang et al., 2018	http://amp.pharm.mssm.edu/Enrichr/
Enrichr	Kuleshov et al., 2016	https://david.ncifcrf.gov/
DAVID	Huang et al., 2009	https://david.ncifcrf.gov/
deep Tools	Ramírez et al., 2014	http://deeptools.ie-freiburg.mpg.de/
MEME-ChIP software (part of the MEME suite)	Machanick and Bailey, 2011	http://meme-suite.org/tools/meme-chip
Gene Set Enrichment Analysis (GSEA)	Subramanian et al., 2005	http://software.broadinstitute.org/gsea/index.jsp
Other		
silicon culture-inserts	ibidi	Cat#80209
OVODYNE Electroporator	Intracel	Cat#TSS20

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to the Lead Contact, Alvaro Rada-Iglesias (aradaigl@uni-koeln.de; alvaro.rada@unican.es).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient description

The patient is a 17 year-old male born to non-consanguineous healthy parents (Figure 1; Table S1). He was born at 38 weeks gestation by normal delivery weighing 3.71 kg and with a head circumference of 35 cm. He sat at nine months and walked at 18 months of age, but with a continuing tendency to tiptoe walk. At two years of age, there were concerns about his development, in particular speech and language; and about his locomotor skills. The patient had dysmorphic facial features with long palpebral fissures, a flattened nasal tip, full lips, a short philtrum and upturned ear lobes as well as a high arched palate and a sector of heterochromia in his left iris. There were no hairy patches and no haemangiomas or unusual patches of skin on his neck. Aged four, he was described to have severe speech and language difficulties, in addition to an alternating convergent squint, right sided severe sensorineural hearing loss and fluctuating left-sided conductive hearing loss secondary to middle ear effusion. He also had palatopharnygoplasty surgery for velopharyngeal insufficiency. Sleep difficulties were reported and he was prescribed melatonin. There was delay in his social communication skills, but he did not meet the ICD-10 criteria for diagnosis of autism when assessed aged seven years. He reported significant difficulties with self-organization and social interactions. He had mild myopia and pooling of tears; however, a formal ophthalmic examination did not suggest a tear duct abnormality. Ultrasound scan of his kidneys and MRI of his inner auditory meatus were normal. The patient never experiences dizziness but his balance has always been poor in the dark. At age 17, his height is on 25th centile, weight on 50th centile and head circumference on 9th to 25th centile. His vestibular function tests were abnormal and underlie his poor balance in the dark. They showed a bilateral vestibular hypofunction of the horizontal semicircular canals and absent cervical

vestibular evoked myogenic potentials (a specific test of the saccule and inferior vestibular nerve) on the left. Echocardiogram was normal.

WT hiPSCs

The WT hiPSCs lines T12 (WT#3) and S24 (WT#2) were provided by the Pediatric Endocrinology and Diabetology laboratory (Department of Pediatrics, University Clinic Carl Gustav Carus, TU Dresden, Germany). These hiPSCs lines were obtained from fibroblasts from healthy donors (Kind et al., 2010), reprogrammed by lentiviral transduction vector and validated as described in Laugsch et al. (2016). The WT hiPSCs lines NP0040-8 (WT#1) provided by Center for Physiology and Pathophysiology, Institute for Neurophysiology, Medical Faculty, University of Cologne, Cologne, Germany was reprogrammed by episomal reprogramming vectors and validated according to EBiSC standards (NP0040-8 is available at EBiSC under accession name UKKi011-A; https://cells.ebisc.org/UKKi011-A/).

BOFS patient-specific hiPSCs

The patient was enrolled and sampled according to standard local practice at the University of Southampton (UK). The use of patient material in this study received consent from the patient's parents, was approved by the ethics committee at the University of Southampton and was in compliance with the Declaration of Helsinki concerning Ethical Principles for Medical Research Involving Human Subjects.

METHOD DETAILS

Characterization of chromosomal abnormalities

Genome analysis by array-CGH was carried out with Agilent 244K oligonucleotide arrays according to standard methods.

Mapping of the inversion breakpoints was undertaken in metaphase chromosomes by two-color FISH using standard protocols with probes derived from BAC, PAC and fosmid constructs. Briefly, the 6q16.2 breakpoint was mapped by FISH to a genomic region containing no known genes within BAC RP11-390H11. The 6p24.3 breakpoint was mapped within PAC RP1-290I10, in a region containing *TFAP2A*. The breakpoint location was further refined by using fosmid WI2-506N5, which excluded *TFAP2A* from the breakpoint region.

Paired-end genome sequencing using 100bp reads was performed on a single lane of the Illumina HiSeq2500 in order to resolve the inversion breakpoints at the nucleotide level.

Targeted Locus Amplification (TLA)

TLA was performed by Cergentis (https://www.cergentis.com/) according to their standard protocol (de Vree et al., 2014). Briefly, hiPSCs derived from our BOFS patient were used as input and analyzed with four sets of TLA primers (Table S7). Primer-sets 1A and 1B were used to amplify both *TFAP2A* alleles in order to (i) phase rs1675414, a heterozygous SNP in the BOFS patient cells, with respect to the inversion, (ii) identify additional heterozygous SNPs in the BOFS patient cells and (iii) to phase them with respect to the inversion. Primer-sets 2 and 3 were designed upstream of the 6p24 and downstream of the 6q16 inversion breakpoints, respectively, in order to confirm and refine the position of the two inversion breakpoints. The primer sets were used in individual TLA amplifications. PCR products were purified, and libraries prepared using the NEBNext Hyper Plus II protocol (NEB) and sequenced on an Illumina sequencer.

hiPSCs culture

All hiPSCs were maintained on GelTrex (Thermo Fisher Scientific) in mTeSR1 (Stem Cell Technologies) with daily medium change or in StemMACS iPS-Brew XF (Miltenyl Biotec), with medium changed every other day. Confluent hiPSCs were regularly split by Versene (GIBCO by Life Technologies) or Accutase (Life Technologies) treatment and supplemented with Thiozavivin (Axon Med Chem) for the first 24-48h.

Derivation of BOFS patient-specific hiPSCs

Fibroblasts from the patient carrying the 89Mb inversion in chromosome 6 were provided by the Wessex Regional Genetics Laboratory, Salisbury, UK. The fibroblast were reprogrammed into hiPSCs using standard factors (OSKM) by non-integrating Cyto Tune-iPS Sendai Kit (Thermo Fisher Scientific) (Seki et al., 2012) and validated by immunostaining and quantitative reverse transcription–PCR for the expression of pluripotency-associated genes. Conventional cytogenetic (G-banding) of used BOFS-hiPSCs clones confirmed their normal karyotype, (Institute of Molecular Haematology and Oncology, University Clinics of Cologne, Germany) except for the already diagnosed 89Mb inversion. The 89Mb heterozygous inversion in patient fibroblasts and hiPSCs was confirmed by PCR using the primers listed in Table S7 (Figure S3) and TLA analysis.

Engineering of the 0.4Mb deletion within TFAP2A-TAD in WT hiPSCs by CRISPR/Cas9

The template sequence flanking the chosen 0.4Mb region was obtained from the UCSC Genome Browser and used for gRNA design. The gRNA sequences (#34: ATTCTAAACATTCCCCGCACTGG; #L2: TGTGCTGAAAGGGTCGTCGCTGG, #39, CTTTGAAAGAT TATCTGCCTAGG and #L5: GGATAACTTGCTTACCCCAGTGG) were designed according to the genome engineering toolbox

(http://www.genome-engineering.org/crispr/) and cloned into the pX330-hCas9-long-chimeric-grna-g2p (kindly provided by Leo Kurian's laboratory, CMMC, University of Cologne), as described in Cruz-Molina et al. (2017). To validate positive bacterial clones, the corresponding forward gRNA oligo and the reverse primer (5′ aacgccaatagggactttcc 3′) and were used for colony PCR and the resulting products were analyzed by Sanger sequencing (SeqLab). To optimize the CRIPSR/cas9 targeting efficiency, NP0040-8 (WT#1) hiPSCs were subject to two rounds of transfections with vectors containing #L2 and #L5 gRNAs (round 1) or #34 and #39 gRNAs (round 2). The vectors were delivered to the cells by non-liposomal formulation (FuGene HD transfection reagent, Promega) upon cell splitting and using mTeSR1 supplemented with Thiozavivin for the first 24h. Two days post transfection, the mTeSR medium was supplemented with 1mg/mL Puromycin for 24h and then gently washed. The hiPSCs were then cultured for additional 4-6 days and tested for the desired 0.4Mb deletion by PCR-based genotyping and Sanger sequencing using the primers shown in Table S7. To generate clonal lines, the targeted hiPSCs were split by accutase and resuspended in mTeSR1 supplemented with Thiazovivin (150 cells per 1cm^2). Medium was carefully changed every other day until colonies were ready for picking. After the first transfection round, the resulting cell lines were tested for the 0.4Mb deletion by PCR-based genotyping using the primers describe in Table S7 and a clone heterozygous for the deletion was subjected to a second round of transfection. Following derivation of clonal lines and subsequent genotyping using PCR and Sanger sequencing, two independent clonal lines homozygous for the 0.4Mb deletion were obtained and used for subsequent experiments.

hNCC differentiation

For the differentiation of hiPSCs into hNCC we used previously reported protocols (Bajpai et al., 2010; Prescott et al., 2015; Rada-Iglesias et al., 2012). Briefly, confluent hiPSCs colonies were detached by 2mg/mL collagenase treatment, washed with PBS and plated in Petri dishes in a hNCC differentiation medium (Neurobasal and DMEM F12 media in 1:1 ratio, 0.5x B27 with Vitamin A and 0.5x N2 supplements, 20ng/mL epidermal fibroblast growth factor, 20ng/mL basic fibroblast growth factor and 5ug/mL Insulin). Embryoid body (EB) formation was induced already 24h post splitting and medium was changed every 2-3 days. Typically, at Day 7 EBs attached to the Petri dishes and gave rise to NCC outgrowths. On Day 11, the hNCC (d11hNCC) were either harvested for downstream analyses or dissociated by accutase treatment and seeded (50.000 cells per cm^2) on cell culture dishes coated with 5mg/mL fibronectin in hNCC maintenance medium in which insulin was replaced by 1mg/mL BSA (Prescott et al., 2015). Following one additional passage, hNCC cells were harvested in passage number 2 (p2) for downstream analyses (p2hNCC).

For differentiation of p2hNCC into later lineages, cells were cultured for 7 days in media that promoted differentiation into smooth muscle (DMEM F12 + 10%FBS) or for 14 days in media that promoted differentiation into neurons/glia (DMEM F12 + B27 + 2 mM glutamine + 50 ng/ml LIF + 1% FBS). The cells were analyzed by immunofluorescence in order to determine the expression of representative lineage-specific markers. For chondrogenic differentiation 5x10^5 p2hNCC were pelleted and subsequently cultured in 0.5 mL of MesenCult ACF Chondrogenic Differentiation Medium (StemCell Technologies, Inc.) in 15ml polypropylene tubes. Three days later, 0.5 mL of fresh medium was added. Subsequently, the medium was changed every three days and the tubes were gently flicked to ensure that the cell pellets were not completely attached to the tube. On day 21, the cell pellets were fixed with 10% formalin for 30 minutes at RT and embedded in paraffin. Next, 8 μm sections were obtained using a microtome, followed by two xylene washes (5 minutes each) in order to deparaffinised them. Sections were rehydrated with washes in decreasing ammounts of ethanol (100%–70%) and rinsed under running tap water for 5 minutes. For alcian blue staining, sections were incubated in 1% alcian blue solution (pH2.5) in 3% acetic acid for 30 minutes. Slides were washed with tap water and counterstained with nuclear fast red solution for 5 minutes. After the staining process, sections were washed in running tap water and through increasing amounts of ethanol (50%–100%) for dehydration. Ethanol was cleared with one wash of acetone and slides were mounted using aquatex permanent aqueous mounting agent (Sigma). Sections were imaged using a brightfield stereomicroscope (Olympus SZX16).

Chicken Embryos

Fertilized chicken eggs (white leghorn; Gallus gallus domesticus) were obtained from a local breeder (LSL Rhein-Main) and incubated at 37°C and 80% humidity in a normal poultry egg incubator (Typenreihe Thermo-de-Lux). Following microsurgical procedures, the eggs were re-incubated until the embryos reached the desired developmental stages. The developmental progress was determined according to the staging system of Hamburger and Hamilton (HH) (Hamburger and Hamilton, 1992).

Isolation of chicken frontonasal prominences

Fertile chicken eggs were incubated at 37°C for 4 days until they reached stage HH24 (Hamburger and Hamilton, 1992). Embryos were isolated from the eggs, transferred into 20 mL of 1x PBS, and the extraembryonic membranes were removed. Then, FNP were isolated using surgical scissors. After isolation, FNP sections were pooled in a 1.5-mL tube and used immediately or flash-frozen in liquid nitrogen and stored at −80°C.

Chromatin immunoprecipitation (ChIP)

ChIPs were performed in p2hNCC and chicken FNP as previously described (Rada-Iglesias et al., 2012; Rehimi et al., 2016). Briefly, cells were crosslinked with 1% formaldehyde for 10 min at room temperature and formaldehyde was quenched by 0.125 M glycine. For TFAP2A ChIPs and histone ChIPs, approximately 20 and 7 million p2hNCC were used, respectively. H3K4me2 ChIP in chicken FNPs was performed with 15 FNP from HH24 chicken embryos. For all samples, chromatin was sonicated in lysis buffer (10 mM Tris, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine) to an average size of 0.5–2 kb using

Bioruptor (Diagenode). A total of 5 μg of antibody for histones (H3K9me3 (39161, Active Motif), H3K27ac (39133, Active Motif), H3K27me3 (39155, Active Motif), H3K4me2 (39141, Active Motif)) and 10 μg antibody for TFAP2A (sc-12726, Santa Cruz Biotechnology)) was added to the sonicated chromatin and incubated overnight at 4 °C. 10% of chromatin used for each ChIP reaction was kept as input DNA. Subsequently, 100 μl of protein G Dynabeads magnetic beads (Life Technologies) were added to the ChIP reactions and incubated for four additional hours at 4 °C. Magnetic beads were washed and chromatin eluted by addition of elution buffer (50 mM Tris, 10 mM EDTA, and 1% SDS) and incubation at 65°C. Crosslinking was then reversed at 65°C overnight followed by RNase (0.2 μg/μL) and Proteinase K (0.2 μg/μL) treatment, and DNA purification by phenol-chloroform and ethanol precipitation. Resultant ChIP DNA was dissolved in water.

Immunofluorescence (IF)

hiPSCs were cultured on coverslips coated with GelTrex and hNCC on coverslips coated with fibronectin. Confluent cells were fixed with 3.7% formaldehyde in PBS for 15 min, permeabilized with 0.5% Triton X-100 and 5% BSA in PBS for 10 min as well as blocked for an additional hour at RT with 5% BSA-PBS solution. Primary antibodies, diluted in 1% BSA-PBS were incubated over night at 4°C (TFAP2A (sc-12726, Santa Cruz Biotechnology), SOX2 (sc-17320, Santa Cruz Biotechnology), TRA-1-60 (sc-21705, Santa Cruz Biotechnology), NR2F2 (H7147, Perseus Proteomics), NR2F1 (PP-H8 132-00, R&D Systems), TWIST1 (PA5-49688, Thermo Fisher), SOX9 (AB5535, Millipore), SMA (anti-SM22a, ab14106, Abcam), TUJ1 (anti-beta III Tubulin, ab18207, Abcam)). Following washing steps with 0.05% Tween-PBS, appropriate secondary antibodies (A-11001, A-11058, R37117, Thermo Fisher) were diluted in 1% BSA-PBS, incubated with the cells for 30 min at 37°C, washed with 0.05% Tween-PBS and incubated with 0.1μM DAPI (Sigma) for 10 min at RT. After washing steps (first with 0.05% Tween-PBS and then with PBS only), the cells were permanently mounted in FluormountG (SouthernBiotech) for subsequent fluorescence microscopy analysis (Olympus IX 81).

RNA extraction, cDNA preparation and RT-qPCR

Total RNA was isolated using the innuPREP RNA Mini Kit (Analytic Jena) or the innuPREP RNA/DNA Mini Kit (Analytic Jena) according to the manufacturer's instructions. cDNA was generated using the ProtoScript II First Strand cDNA Synthesis Kit (E6560L, New England Biolabs). RT-qPCRs were performed on the Light Cycler 480II (Roche) using ORA qPCR Green ROX L Mix (highQu).

Flow cytometry

For analysis of p75, hNCC were detached by accutase, resuspended in 0.5% BSA-PBS solution and cells were stained with p75/CD271-PerCP (345111, BioLegend) antibody for 1h at 4°C and washed with 0.5% BSA-PBS. To investigate the expression of nuclear proteins (i.e., TFAP2A, NR2F1, SOX9), cells were fixed with 2% formaldehyde in PBS for 15 min at RT, permeabilized by 0.1% Triton and 5% BSA in PBS for 30 min at RT and blocked for additional 30 min at RT with 5% BSA-PBS. Primary antibody against TFAP2A (sc-12726, Santa Cruz Biotechnology), NR2F1 (PP-H8 132-00, R&D Systems) and SOX9 (AB5535, Millipore) were diluted in 5% BSA-PBS and incubated for 1h and RT. Following washing steps with 0.5% BSA-PBS, secondary antibodies (conjugated to PerCP (F0114, R&D Systems; 46-4010-82, Invitrogen), conjugated to APC (F0111, R&D Systems) or conjugated to Dylight694 (406404, Biolegend)) diluted in 5% BSA-PBS were added for an additional 30 min at RT. Finally the cells were washed with 0.5% BSA-PBS. All samples were then analyzed by BD FACSDIVA (BD Bioscience) and FloJo softwares.

In vitro reporter assays

In vitro enhancer reporter assays were performed using a previously described PiggyBAC transposon-based enhancer reporter system (Respuela et al., 2016). Enhancer sequences (Enh100, Enh105 and Enh480) were amplified using the Q5 High-Fidelity DNA Polymerase (NEB) with primers listed in Table S7 and cloned into the target vector *PiggyBacB enhancer GFP Neo* with BamHI and EcoRI or BstI in front of a minimal TK promoter driving GFP expression. hiPSCs were co-transfected with the resulting PiggyBac reporter vectors and a vector expressing the Super PiggyBac Transposase (System Biosciences) using FuGENE HD transfection reagent (Promega) according to the manufacturer's instructions. In addition, hiPSCs were also transfected with an empty PiggyBac reporter vector and the Super PiggyBac Transposase to generate a negative control reporter hiPSCs line. Starting 24 hours after transfection, cells were constantly treated with neomycin (0.2 mg/ml) in order to select cells that stably incorporated the reporter vector. When confluent, surviving cells were expanded and differentiated into hNCC. GFP levels were evaluated using a Nikon ECLIPSE TS100 Microscope.

In vivo reporter assays

In vivo reporter assays were performed using a previously described β-globin-GFP enhancer reporter system (Bergsland et al., 2011). Enhancer sequences (Enh100, Enh105 and Enh480) were amplified using the Q5 High-Fidelity DNA Polymerase (NEB) with primers listed in Table S7 and cloned into the reporter vector with KpnI and NheI or BamHI in front of a minimal β-globin promoter driving GFP expression. Electroporation of chicken embryos were performed as previously described (Rehimi et al., 2016). Briefly, eggs of stage HH8-9 were windowed, and the extra embryonic membrane was partially removed. The reporter vector and an RFP expressing control vector (PiggyBac-CMV-MCS-EF1-RedPuro (PB514B-2, System Biosciences)) were mixed 2:1 with Fast Green solution (Sigma) and microinjected into the neural tubes of the chicken embryos. The neural tube was then electroporated with five square pulses of 20 V within 20 ms width using the Intracel TSS20 OVODYNE Electroporator. Following electroporation, the eggs were sealed with

tape and re-incubated until the desired developmental stages (HH14-16 and HH20-23). GFP and RFP signals were then evaluated using Olympus SZX16 stereomicroscope with EXFO X-cite series 120PC Q for florescence illumination.

Scratch assays

Scratch assays were performed using cell culture inserts. Briefly, 24 well plates were coated with fibronectin and hNCC in passage 1 (p1hNCC) ($3x10^5$ cells per mL) were split for p2hNCC into one well of a silicon culture insert (ibidi) attached to the surface. The inserts allow cell growth in the designated areas while creating a cell–free gap. Following 24h and cell attachment the inserts were carefully removed and the medium was changed. Using EVOS FL Auto Cell imaging System and Onstage Incubator Life, we visualized cell migration within the gap and acquired images every two hours for a total period of 24h.

In addition, traditional "scratch assays" was performed. Briefly, WT and BOFS p2hNCC were grown in 6-well plates and scratches (one per well) were manually generated using pipette tips. Cell migration into the generated scratched was monitored at 0h, 10h and 22h by brightfield microscopy.

Proliferation assays

$5x10^6$ WT or BOFS p2hNCC were labeled with 10 μM Carboxyfluorescein succinimidyl ester (CFSE) (CellTrace CFSE Cell Proliferation Kit; MolecularProbes) in 1ml PBS or in PBS without dye and incubated for 10 min at 37°C. The reaction was quenched by addition of 5 mL ice-cold hNCC maintenance medium followed by a 5 min incubation on ice. After centrifugation, the cell pellets were washed twice with fresh medium, seeded on fibronectin coated dishes and cultured in hNCC maintenance medium for one to four days. Samples were collected (day1, day2, day3 and day4) and analyzed using flow cytometry with 488 nm excitation.

Allele-specific analysis of TFAP2A expression

Total RNA was isolated using the innuPREP RNA Mini Kit (845-KS-2040250, Analytic Jena) according to the manufacturer's instructions. RNA was then treated with TURBO DNase (Life Technologies) before cDNA was generated using the ProtoScript II First Strand cDNA Synthesis Kit (E6560L, New England Biolabs). Reference DNA was extracted using QuickExtract Solution (50 mM KCl, 10 mM TRIS pH8.3, 2.5 mM $MgCl_2$, 0.45% NP40, 0.45% Tween20, 0.4 mg/ml Proteinase K). The heterozygous SNP rs1675414 was then amplified with suitable primers (Table S7) using Platinum Taq Polymerase (ThermoFisher Scientific). Subsequently, PCR products were purified using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and Sanger sequenced with the reverse primer.

RNA-seq

RNA-seq libraries were prepared using the RNA sample preparation kit (TruSeq v2; Illumina) as previously described (Respuela et al., 2016) and sequenced with a 2 × 75-bp strand-specific protocol on a HiSeq 2500 sequencer (Illumina). The following RNA-seq libraries were prepared:

(i) WT d11hNCC (three samples derived from WT hiPSCs#1) and Δ0.4Mb d11hNCC (four samples: two derived from Δ0.4Mb hiPSCs Cl#1 and two from Δ0.4Mb hiPSCs Cl#2).
(ii) WT p2hNCC (two samples: one derived from WT hiPSCs#1 and one from WT hiPSCs#2) and BOFS p2hNCC (three samples: one derived from BOFS hiPSCs#1, one from BOFS hiPSCs#2 and one from BOFS hiPSCs#3).

Single-cell RNA-seq (scRNA-seq)

WT and BOFS p2hNCC were dissociated to generate a suspension of viable single cells. The suspension was centrifuged at 300 *rcf.* for 5 min and the cell pellet was washed in 1 mL of 1xPBS 0.04% BSA. Cells were centrifuged again and resuspended in 1xPBS 0.04% BSA to the desired cell concentration. Cells were passed through a strainer and cell concentration was determined again. cDNA synthesis and library preparation was performed following the 10x Single Cell 3′ Reagent Kit protocol. Briefly, cells were partitioned in nanoliter-scale Gel Bead-In-Emulsions (GEMs) containing, in addition to a single cell, a master mix and primers that consist of (i) an Ilumina R1 sequence, (ii) a 16bp 10x barcode unique for each cell, (iii) a 10 bp Unique Molecular Identifier (UMI) and (iv) a poly-dT sequence. Full-length barcoded cDNAs were generated from poly-adenylated mRNAs. GEMs were broken and the pooled fractions were recovered. The cDNAs were exposed to enzymatic fragmentation and size selection prior to library construction and P5, P7, a sample index and an Ilumina R2 primer were added during library construction. The final libraries containing the P5 and P7 primers were used for Illumina bridge amplification. Finally, R1 and R2 primers were used for paired-end Illumina sequencing.

ChIP-seq

ChIP-seq libraries from ChIP and input DNAs obtained from WT p2hNCC (TFAP2A ChIP (two replicates), H3K27ac ChIP, H3K27me3 ChIP, H3K9me3 ChIP, input) and BOFS p2hNCC (TFAP2A ChIP (two replicates), H3K27ac ChIP, H3K27me3 ChIP, H3K9me3 ChIP, input) were prepared according to Illumina protocol and sequenced with a 1 × 50–bp protocol on a HiSeq 2500 sequencer (Illumina). H3K4me2 ChIP DNA obtained from HH24 chicken FNP was prepared according to Illumina protocol and sequenced with a 2 × 100–bp strand-specific protocol on a HiSeq 2500 sequencer (Illumina).

4C-seq

Circular Chromatin Conformation Capture (4C) assays were performed as previously described with slight modifications (Cruz-Molina et al., 2017). 2-3x10^7 cells were crosslinked with 1% formaldehyde for 10 min and quenched with 0.125 M glycine. Cells were washed with PBS, resuspended in lysis buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100 and 1X protease inhibitor) and incubated on ice for 10 min. Following centrifugation for 5 min at 4000x*g* and 4°C, nuclei were re-suspended in 0.5 mL of 1.2x CutSMART restriction buffer (NEB) with 0.3% SDS and incubated at 37°C and 900 rpm for 1h. Then, Triton X-100 was added to a final concentration of 2%, reducing the final CutSMART dilution to 1x, followed by 1h incubation at 37°C and 900 rpm. Afterward, chromatin was digested overnight at 37°C and 900 rpm with 600 U of NlaIII (R0125L, NEB). NlaIII was inactivated by adding SDS to a final concentration of 1.6% and incubating the mixture for 20 min at 65°C and 900 rpm. The digested chromatin was transferred to 50 mL tubes and filled up with 1x ligation buffer (50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 1 mM ATP, 1mM DTT) to a final volume of 7 ml. Then, Triton X-100 was added to a final concentration of 1% and the solution was incubated for 1h at 37°C and 300 rpm. Afterward, digested chromatin was ligated using 100 U of T4 DNA ligase (15224-041, Life Technologies) for 4h at 16°C, followed by RNase A treatment (300 μg total, Peqlab) for 45 min at 37°C. Subsequently, chromatin was de-crosslinked by treating with Proteinase K (300 μg total, Peqlab) and incubating at 65°C overnight. DNA was then purified by phenol/chloroform extraction followed by ethanol precipitation and re-suspended in 100 μL of water. At this point, the digestion and ligation efficiencies were evaluated by analyzing a small fraction of the purified DNAs by agarose electrophoresis on a 1% agarose gel. The remaining DNA was digested with 50U of DpnII (R0543M, NEB) in 500 μL of 1x NEBuffer DpnII at 37°C overnight. DNA samples were subsequently purified by phenol/chlorophorm extraction and ethanol precipitation. Purified DNA samples were re-suspended in 500 μL of water and a second ligation was performed using 200U of T4 DNA Ligase in a final volume of 14 mL 1x ligation buffer followed by overnight incubation at 16°C. DNA samples were subjected to another round of phenol/chlorophorm extraction and ethanol precipitation, re-suspended in 150 μL of water and purified by QIAquick PCR Purification Kit (28104, QIAGEN). Efficiencies of the second digestion and ligation were also evaluated by agarose electrophoresis. Finally, the resulting 4C DNA products were amplified by inverse PCR using primers located within selected regions, which were designed as previously described (Cruz-Molina et al., 2017) (Table S7). The inverse PCRs were performed with the expand long template PCR system (11681842001, Roche) using 32 amplification cycles (94°C 2 min, 32x [94°C 10 s, 58°C 1 min, 68°C 3 min], 68°C 5 min). 4C-seq libraries were sequenced on the HiSeq2500, generating reads of either 74 or 100 bp in length.

ATAC-seq

Assay for Transposase Accessible Chromatin coupled to high-throughput sequencing (ATAC-seq) was performed as previously described (Buenrostro et al., 2013) with slight modifications. Briefly, two isolated FNP of chicken embryos HH23 were disaggregated using 0.125% (w/v) collagenase in PBS with 10% BSA for 30 min at 37°C and 600 rpm and then carefully resuspended. Approximately 50.000 single cells from chicken FNP or hNCC were then lysed in 500 μl cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0,1% IGEPAL CA-630) substituted with 1x protease inhibitor for 15 min on ice. Immediately after lysis, nuclei were centrifuged at 4000xg for 10 min at 4°C. Nuclei were re-suspended in 25 μl 2x TD buffer (15027866, Illumina), 10 μl TDE1 (*i.e.* transposase, 15027868, Illumina) and 15 μl nuclease-free H_2O, and incubated for 30 min at 37°C. Afterward, the sample was purified using the MinElute PCR Purification Kit (28004, QIAGEN) according to the manufacturer's protocol and eluted with 10 μl elution buffer. Subsequent whole genome sequencing was performed using the NexteraXT Kit (Illumina).

QUANTIFICATION AND STATISTICAL ANALYSIS

TLA data analysis

Reads were mapped to the human genome (hg19) using BWA-SW (Li and Durbin, 2009), allowing partial mapping which is optimally suited for identifying reads that span breakpoints.

First, the TLA data was used to confirm and refine the breakpoints of the inversion identified in our patient:

6p24 Breakpoint: *chr6:99,103,875 (rev)* fused to **chr6:10,355,280 (fwd)**, with 3bp of homologous sequence at the breakpoint.

CCTCTGAGCTGCAGTTGACAACGTTACAGTTGGTAGCTCTTACAGTGGCCTAGCTCCTGGGCTGTTTGAGGTAAAATGATATGAA ATGGA**TTGAAGCAAGAAGGTAACAAGCATGATTCCCAGCATTTGTGGGGGTAACTTGTACTTTTT**

6q16 Breakpoint: *chr6:99,103,872 (fwd)* fused to **chr6:10,355,274 (rev)** with a 7bp insertion

*CAAAATCTCAGTGTCTTCAAACAACAGATTTATTTCTTATTCACATTATCTGTCCATTATAGGGCAGATGAGCCCTATAA***TGTATTT GGTAAGCACTTTAGAAGCCTTTAAAATAGTACTCTAAGGCAGGTAACGTGGTGTGCCTCTAGT**

Next, the TLA data was used to identify heterozygous SNPs and link them to the inversion allele in the BOFS patient. The first step was to find heterozygous SNPs. Therefore, SNP calling was performed on the regions 100kb upstream and downstream of the inversion breakpoints. Heterozygous SNPs were called by selecting the positions with $\geq$ 30X coverage and with a minor allele frequency of 20%–80%. Assigning SNP alleles to either the wild-type (WT) or inversion (INV) allele was done based on the concept that different primer-sets amplify different alleles. The linkage analyses on the heterozygous SNPs was performed in four sub-regions:

(i) SNP-region 1 is the region upstream of the 6p24.3 breakpoint. In this region, SNPs were called in the data of primer-set 2 that amplified both alleles of this region. This resulted in the detection of 10 heterozygous SNPs in region 1. Linkage of these SNPs to the inversion was performed using the data of primer-sets 1a and 1b, knowing that these two primer-sets only have coverage

of the WT allele in this region. All 10 SNPs were indeed homozygous in the 1a and 1b dataset, and both 1a and 1b showed the same linkage results for all positions. So, all 10 SNPs were phased relative to the inversion.

(ii) SNP-region 2 is the region downstream of the 6p24.3 breakpoint. In this region, SNPs were called in the data of primer-set 1a and 1b that each amplified both alleles in this region. In total, 79 heterozygous SNPs were detected in region 2. Linkage of these SNPs to the inversion was determined using the data of primer-sets 2 and 3, knowing that primer-set 2 shows coverage on the WT allele, and primer-set 3 shows coverage on the INV allele. In total 71 heterozygous SNPs were phased with the inversion in region 2. Among these is the rs1675414 SNP at position chr6:10,412,188, for which the TLA data showed that the "T" allele is found on the INV chr6, whereas the "C" allele is found on the WT chr6.

(iii) SNP region 3 is the region upstream of the 6q16.2 breakpoint. In this region, SNPs were called in the data of primer-set 2 and 3. There were no SNPs that met the coverage requirement in both primer-sets, so no SNPs could be linked to either WT or INV in region 3.

(iv) SNP region 4 is the region downstream of the 6q16.2 breakpoint. In this region, SNPs were called in the data of primer-set 3 that amplified both alleles of this region. 55 heterozygous SNPs were detected in region 4. Linkage of these SNPs to the inversion was performed using the data of primer-sets 1a and 1b, knowing that these two primer-sets only have coverage of the INV allele in this region. In total, 52 heterozygous SNPs were phased relative to the INV in region 4.

Altogether, 133 heterozygous SNPs have been specifically linked to either the WT or the INV allele in our BOFS patient cells (Table S2).

RT-qPCR data analysis

Gene expression fold changes were calculated using the delta-delta Ct method and *EEF2, ACTB* and *GAPDH* as housekeeping genes. Primers used are listed in Table S7.

Quantification of Scratch Assays

To quantify cell migration, the gaps (i.e., scratched areas) were measured using FIJI (ImageJ) (Schindelin et al., 2012) ROI tools to manually outline the borders of the gap in images acquired at 0, 4 and 10 hours. The determined areas were normalized to the initial gap size by dividing the areas obtained after 4 and 10 hours by the area measured at 0 hours. Student's t test was used to assess the statistical significance of the difference between wild-type and patient samples after 4 and 10 hours. Representative images were obtained using KNIME, OMERO and Photoshop CS2 (Adobe Systems, San Jose, CA).

Quantification of Proliferation Assays

Cell proliferation is inversely proportional to the MFI (Mean Florescence Intensity) of CFSE. Therefore, we measured the average MFI of each sample and calculated proliferation rates as a Fluorescence Dilution Factor (FDF) (Griessinger et al., 2016) by dividing the MFI measured on Day 1 by the MFI measured on Days 2, 3 or 4.

Quantification of TFAP2A allele-specific expression

Quantitative allele ratio analysis was performed using PeakPicker software (Ge et al., 2005). The rs1675414 SNP was genotyped in genomic DNA (gDNA) and cDNA from hNCC (d11hNCC and p2hNCC) derived from two WT (WT hiPSCs#1, WT hiPSCs#3) and three BOFS patient (BOFS hiPSCs Cl#1, BOFS hiPSCs Cl#2, BOFS hiPSCs Cl#3) hiPSCs lines, respectively. For each cell line, five genotyping measurements were performed in gDNA and cDNA obtained from at least two independent hNCC differentiations.

RNA-seq analysis

RNA-seq data were analyzed using a high-throughput next-generation sequencing analysis pipeline (Wagle et al., 2015): Basic read quality check was performed using FastQC (Babraham Bioinformatics) and read statistics were obtained with SAMtools. Reads were mapped to the human reference assembly (hg38) using TopHat2 (Kim et al., 2013). Read count means, fold-change (FC) and p values were calculated with DESeq2 (Anders and Huber, 2010) and gene expression for the individual samples was calculated with Cufflinks (Trapnell et al., 2010) as FPKMs, using in both cases genomic annotation from the Ensembl database. Log FC values reported by DESeq2 were transformed to linear FC by taking 2^x. Linear FC values between 0 and 1 were then transformed to $-1/x$ (Tables S2 and S4).

Genes differentially expressed between: (i) WT d11hNCC (three samples derived WT hiPSCs#1) and Δ0.4Mb d11hNCC (four samples derived from two different Δ0.4Mb hiPSCs lines) or (ii) WT p2hNCC (two samples derived from two different WT hiPSCs lines) and BOFS p2hNCC (three samples derived from three different BOFS hiPSCs lines) were determined according to the following criteria, considering only Ensembl Genes used in GREAT; McLean et al., 2010) (Tables S2 and S4):

- Downregulated in Δ0.4Mb d11hNCC (Δ0.4Mb Down): p value < 0.01, FC < −1.5, Average FPKM in WT d11NCC samples > 0.1.
- Upregulated in Δ0.4Mb d11hNCC (Δ0.4Mb Up): p value < 0.01, FC > 1.5, Average FPKM in Δ0.4Mb d11hNCC samples > 0.1.
- Downregulated in BOFS p2hNCC (BOFS Down): p value < 0.01, FC < −1.5, Average FPKM in WT p2NCC samples > 0.1.
- Upregulated in BOFS p2hNCC (BOFS Up): p value < 0.01, FC > 1.5, Average FPKM in BOFS p2hNCC samples > 0.1.

scRNA-seq analysis

UMIs were counted using cellranger-2.1.0 (Zheng et al., 2017) with default parameters on hg19. Initial dimensionality reduction, clustering and visualization were performed with cellranger-2.1.0 and cellrangerRkit_2.0.0 within R-3.4.0. Counts were aggregated into a single matrix with default normalization ("–normalize=mapped") by cellranger-2.1.0. This matrix was further processed with monocle_2.6.4 (Qiu et al., 2017) within R-3.4.0. Dropouts were corrected with Rmagic_1.3.0 (van Dijk et al., 2018) within R-3.4.0 for correlation analysis.

ChIP-seq analysis

ChIP-seq sequencing reads were mapped to the human (hg19 assembly) or chicken (galGal4 assembly) genomes using BWA (Li and Durbin, 2009) and duplicate reads were discarded. In the case of TFAP2A ChIP-seq samples (WT p2hNCC TFAP2A#1, WT p2NCC TFAP2A#2, BOFS p2hNCC TFAP2A#1, BOFS p2hNCC TFAP2A#2), the resulting BAM files were then analyzed with MACS2 (Zhang et al., 2008) using the following settings in order to identify genomic regions significantly enriched in TFAP2A binding in comparison to their corresponding total genomic input DNA: $q \leq 10^{-5}$; Fold-enrichment ≥ 5; Broad Region Calling OFF. Subsequently, in order to determine which regions were differentially bound by TFAP2A in WT and BOFS p2hNCC, we first considered the union of all TFAP2A peaks identified in the four TFAP2A ChIP-seq experiments (n = 56630 peaks). Among all these TFAP2A peaks, those displaying statistically differential TFAP2A binding between WT and BOFS p2hNCC were identified with DiffBind (Ross-Innes et al., 2012), using FDR $\leq 10^{-2}$; Fold ≥ 2 (Fold in DiffBind refers to the difference in mean read concentrations between the WT and BOFS TFAP2A ChIP-seq signals). As a result, TFAP2A peaks were classified as *Unchanged*, *BOFS Low* and *BOFS High* (Table S6).

4C-seq analysis

Reads were assigned to samples based on the first 10 bases of the read. Subsequently, the primer sequence was removed from the read and the remaining sequence was trimmed to 36 bases. These 36 bases were aligned to the human (hg19) or chicken (gg3) reference genomes using Bowtie (Langmead et al., 2009). The resulting mapped reads were analyzed with R3C-seq (Thongjuea et al., 2013) in order to generate RPM (reads per million) normalized bedgraph files for downstream visualization and analysis. In all cases, the sequence that occurred most often at the restriction site matched the NlaIII recognition sequence.

Public genomic datasets and additional bioinformatics analyses

Publically available ChIP-seq datasets used in this study (Rada-Iglesias et al., 2011, 2012):

- Chicken HH20 frontonasal prominences: H3K27ac (GEO: GSM933349)
- hESC: p300 (GEO: GSM602291), H3K27me3 (GEO: GSM602293), H3K27ac (GEO: GSM602294), H3K4me1 (GEO: GSM602295), H3K4me3 (GEO: GSM602296).
- d11hNCC: p300 (GEO: GSM714804), H3K27me3 (GEO: GSM714806), H3K27ac (GEO: GSM714807), H3K4me1 (GEO: GSM714808), H3K4me3 (GEO: GSM714809).

Publically available Hi-C data generated in hESC (Dixon et al., 2015) was visualized using the *3D Genome Browser* (Wang et al., 2018).

Genes differentially expressed between WT and Δ0.4Mb d11hNCC or between WT and BOFS p2hNCC were functionally annotated using Enrichr (Kuleshov et al., 2016) and DAVID (Huang et al., 2009). For the Enrichr analysis, the *MGI* and *Human Phenotype* ontologies were considered and the categories showing *Enrichr Combined Scores* > 10 or at least the top 50 functional categories displaying the highest *Enrichr Combined Scores* were considered (Tables S2 and S5). For the DAVID analysis, the *GO_BP_FAT* ontology was considered and the categories showing p values $< 10^{-5}$ or at least the top 50 functional categories displaying the lowest p values were selected (Tables S2 and S5).

As stated above, for the TFAP2A ChIP-seq experiments, two biological replicates were performed in both WT and BOFS p2hNCC. Pearson correlation coefficients between TFAP2A ChIP-seq experiments were determined by the *bamCorrelate* tool from *deepTools* (*bins* mode and a bin size of 10 Kb across the whole human genome) (Ramírez et al., 2014) (Figure S7A).

To visualize ChIP-seq and ATAC-seq signal profiles, corresponding BAM files were transformed into bedgraph or bigwig files in which signals were normalized as RPGC (reads per genomic content; 1x depth of coverage) using five bp bins according to *deepTools*. Normalized bigwig files were used to generate heatmap signal profiles around the center of TFAP2A peaks using *deepTools*.

GREAT (McLean et al., 2010) was used to assign TFAP2A peaks to their putative target genes using the *Basal plus extension* association rule: each gene was assigned a basal regulatory domain of a minimum of 5 Kb upstream and 1 Kb downstream of the TSS (regardless of other nearby genes). The gene regulatory domain was extended in both directions to the nearest gene's basal domain but no more than the 250 Kb in one direction. Once every gene got assigned a regulatory domain, each genomic region was associated with all genes whose regulatory domain it overlaps.

Motif analysis in *Unchanged*, *WT High* and *BOFS High* TFAP2A peaks was performed with the MEME-ChIP software (part of the MEME suite) using standard settings (Machanick and Bailey, 2011).

Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) was performed using the *GSEAPreranked* option following the provided instructions (https://www.broadinstitute.org/gsea/index.jsp). The considered gene sets consisted of:

- Genes associated with *WT High* TFAP2A peaks according to GREAT (*Basal plus extension* association rule with a maximum distal extension of 250 kb) (3531 genes).
- Genes associated with *BOFS High* TFAP2A peaks according to GREAT (*Basal plus extension* association rule with a maximum distal extension of 250 kb) (1047 genes).

GSEAPreranked analysis was performed for the previous gene sets against all human genes (Ensembl Genes used in GREAT) ranked according to the expression FC obtained in RNA-seq experiments where BOFS p2hNCC versus WT p2hNCC were compared. GSEAPreranked analyses were performed using 1000 iterations.

DATA AND SOFTWARE AVAILABILITY

All ChIP-seq, RNA-seq, scRNA-seq, ATAC-seq and 4C-seq datasets generated in this work have been deposited into the GEO repository under accession number GEO: GSE108522.