Human-specific transposable elements shaped the evolution of craniofacial development through regulation of neural crest migration

Laura Deelen[1], Zoe H. Mitchell[1], Martina Demurtas[1], Beatriz Garcia Del Valle[1] and Marco Trizzino[1,*]

1 Department of Life Sciences, Imperial College London, London, UK

* **Corresponding author**: E-mail: m.trizzino@imperial.ac.uk

Abstract

Craniofacial development and neural crest specification are evolutionarily conserved processes, yet subtle modifications to their gene regulatory networks drive species-specific craniofacial diversity. Transposable elements (TEs) are increasingly recognized as contributors to genome evolution, but their role in shaping neural crest regulatory programs remains underexplored. Here, we investigate the domestication of human-specific TEs as transcriptional enhancers during cranial neural crest cell (CNCC) specification, a process critical for vertebrate head development. Using human iPSC-derived CNCCs, we identified ~250 human-specific TEs acting as active enhancers. These TEs were predominantly LTR5Hs and, to a lesser extent, SVA-E/Fs. We demonstrate that these elements have been co-opted through the acquisition of the conserved CNCC *coordinator* motif, and are bound by the CNCC signature factor TWIST1, and that their co-option appears to be largely exclusive to CNCCs.
To assess their functional relevance, we used CRISPR-interference to repress ~75% of all the LTR5Hs and SVAs active in CNCCs, which led to widespread transcriptional changes in genes associated with neural crest migration, a process essential for CNCCs to populate the embryo and form craniofacial structures. Using a cell migration assay, we showed that CNCC migration was significantly impaired by CRISPR-mediated TE repression. Finally, we demonstrate that genes near human-specific TEs are more highly expressed in human CNCCs relative to chimpanzee, and TE repression returns their expression to chimpanzee levels.
These findings reveal how human-specific TEs have been co-opted to fine-tune CNCC regulatory networks, potentially contributing to the evolution of lineage-specific craniofacial traits.

Keywords: LTR5Hs, SVA, enhancer, CNCC, co-option, TWIST1, coordinator motif, evolution of cell migration

Introduction

Craniofacial development is a highly complex process that requires precise spatiotemporal regulation and involves contributions from all three germ layers and in particular from the neural crest cells (NCCs) (Gans & Northcutt, 1983; Ahlstrom & Erickson, 2009; Martik & Bronner, 2021; Theveneau & Mayor, 2012). NCCs emerge during early embryogenesis, between weeks 3-4 in humans, at the neural plate border between the neuroectoderm and non-neural ectoderm. As neurulation progresses, the neural plate invaginates and separates from the dorsal ectoderm, forming the neural tube. At this stage, NCCs undergo epithelial-to-mesenchymal transition (EMT), allowing them to delaminate and migrate to specific regions throughout the developing embryo (Ahlstrom & Erickson, 2009; Martik & Bronner, 2021; Theveneau & Mayor, 2012). Among the NCC subtypes, cranial neural crest cells (CNCCs) play a pivotal role in the formation of key craniofacial structures, including bones and cartilage (Bronner & LeDouarin, 2012; Cordero et al., 2011; Jheon & Schneider, 2009).

Although craniofacial development is an evolutionary conserved process, recent adaptations to the modern human craniofacial complex include changes in shape and function to accommodate the enlargement of the brain, the transition to bipedal posture, laryngeal extension for speech as well as adjustments for the evolvement of sensory organs (Lieberman, 1998; Sambataro et al., 2022; Spoor et al., 1994). The evolution of human-specific craniofacial traits has required precise modifications in gene expression and increasing evidence suggests that regulatory changes, rather than protein-coding mutations, have been key to shaping species-specific features (Carroll, 2005; King & Wilson, 1975; Wray, 2007). One major source of these regulatory innovations are the transposable elements (TEs). Comprising nearly half of the human genome, TEs are now recognized as key contributors to genomic evolution through their ability to integrate into the genome and act as cis-regulatory elements (Bourque et al., 2008; Chuong et al., 2013, 2016; Cosby et al., 2021; Goubert et al., 2020; Kunarso et al., 2010; Lynch et al., 2011; Pontis et al., 2019; Schmidt et al., 2012; Sundaram & Wysocka, 2020). Transposable element-mediated rewiring of gene regulatory networks has previously been implicated as a major driver of species-specific gene expression patterns (Chuong et al., 2013; Feschotte, 2008; Fueyo et al., 2022; Jacques et al., 2013; Patoori et al., 2022; Playfoot et al., 2021; Prescott et al., 2015; Sundaram et al., 2014; Trizzino et al., 2017).

However, much remains to be uncovered about the precise mechanisms through which transposable elements (TEs) have been co-opted as cis-regulatory elements in humans, particularly in the context of craniofacial development. In the human genome, SINE-Vntr-Alus (SVAs) and LTR5Hs have been previously linked to gene regulatory activity (Barnada et al., 2022; Chuong et al., 2016; Fuentes et al., 2018; Patoori et al., 2022; Pontis et al., 2019; Trizzino et al., 2017). SVAs are the youngest TE family and include roughly 3000 copies in the human genome. They are composed of a hexamer repeat, an Alu-like element, a variable

number of tandem repeats (VNTRs), a SINE element, and a poly-A tail (Fig. 1A). The SVAs are hominoid-specific, comprise of six subfamilies (A–F), with SVA-E and SVA-F being found exclusively in humans, with ~1700 copies in total (Quinn & Bubb, 2014; Wang et al., 2005). LTR5Hs are also human-specific, and are remnants of an ancestral endogenous retrovirus of the HERV-K subtype.

Co-option of SVAs and LTR5Hs as active cis-regulatory elements has been observed in different human tissues (Barnada et al., 2022; Chuong et al., 2013; Fuentes et al., 2018; Ostertag et al., 2003; Patoori et al., 2022; Pontis et al., 2019; Trizzino et al., 2017). Since the expansion of LTR5Hs and SVA-E/F subfamilies occurred around the time of the human–chimpanzee split, we set out to investigate if these elements could have contributed to human-specific craniofacial development.

Human cranial neural crest cell (CNCC) specification and migration can be effectively modelled *in vitro* using human-induced pluripotent stem cells (hiPSCs). Therefore, in this study, we employed an inducible CRISPR-interference (CRISPRi) hiPSC line to investigate the impact of silencing human-specific SVA-E, SVA-F, and LTR5Hs elements to CNCC formation and migration, which are essential processes for the development of the craniofacial structures, including bones and cartilage. To achieve this, we employed previously published single-guide RNAs (sgRNAs) targeting approximately 80% of these transposable elements (Pontis et al., 2019).

With this approach, we identified approximately 250 human-specific SVAs and LTR5Hs that are accessible and depleted of the repressive histone mark H3K9me3 in human CNCCs. We found that the specific DNA sequence was the primary driver for the co-option of this set of TEs. Importantly, silencing these retroelements attenuated the expression of hundreds of genes involved in CNCC migration, which is a key process in craniofacial morphogenesis. Functional assays confirmed that migration was disrupted, suggesting a potential role for the transposons in species-specific craniofacial development. This was further supported by comparisons with previously published chimpanzee CNCC data, which revealed that the expression levels of genes located near accessible human-specific transposable elements resembled those observed in human CNCCs upon depletion of these elements.

Results

An inducible CRISPR-interference iPSC system for the repression of hominoid-specific TEs

To investigate the role of human-specific transposable elements (TEs) (SVAs and LTR5Hs; Fig. 1a) in human cranial neural crest cell (CNCC) development, we designed a stable human iPSC line with an inducible CRISPR-interference (CRISPRi) system with single-guide RNAs

(sgRNAs) targeting ~80% of all SVAs and LTR5Hs annotated in the human genome (Fig. 1b). The sgRNAs were originally designed and validated in a study by the Trono group (Pontis et al., 2019) and were further used in subsequent studies (Barnada et al., 2022; Patoori et al., 2022). Briefly, we cloned a stable iPSC line with a permanently integrated TET-inducible, catalytically dead, Cas9 fused to a repressive KRAB domain (dCas9-KRAB), along with the two gRNAs targeting the LTR5Hs and SVAs (hereafter +gRNA line). The KRAB domain recruits the transcriptional machinery necessary to deposit repressive histone methylation (H3K9me3) to the regions targeted by dCas9. To account for potential off-target effects caused by exposure to doxycycline (TET-ON) or by Cas9 expression, we cloned the same iPSC line with an identical dCas9-KRAB construct but without any gRNAs (hereafter -gRNAs line). Treating the cells with doxycycline for 24 hours was sufficient to activate dCas9 in both CRISPRi lines (i.e. with and without gRNAs; Fig. 1c,d).

Next, we generated CNCCs from our CRISPRi-iPSC lines using an established 5-day protocol (Fig. 1e-f; Leung et al., 2016). Since both SVAs and LTR5Hs have been shown to have important roles in human embryonic stem cells and iPSCs (Barnada et al., 2022; Fuentes et al., 2018; Pontis et al., 2019), doxycycline was only introduced 24h after differentiation (Fig. 1c). By day 5 of differentiation, both cell lines expressed markers typical of CNCC identity, both at gene (*SOX9*, *SOX10*, *TWIST1*, *TFAP2A*) and protein (SOX9, AP2α) level. We generally did not observe significant differences in CNCC marker expression between the two lines, apart from *TFAP2A* (AP2α) at the gene level (Fig. 1e,f), indicating that expression of the main genes essential for CNCC identity was largely unaffected by the CRISPRi.

Hundreds of SVAs and LTR5Hs are accessible in human CNCCs

We set out to determine whether any LTR5Hs and human-specific SVA exhibited chromatin accessibility in human CNCCs, and whether these elements could be repressed using our CRISPRi system. To this end, we performed ATAC-seq (paired-end, 150 bp reads) in hiPSC-derived CNCCs generated with our CRISPRi-iPSC lines (+ and -gRNAs). K-mer clustering of the ATAC-seq data identified a total of 256 accessible LTR5Hs and human-specific SVAs (Fig. 2a; Supplementary File S1). Notably, 77% of these 256 TEs displayed decreased accessibility in the +gRNA CRISPRi line relative to the -gRNA control (cluster 2, Fig. 2a). This is consistent with the assumption that these gRNAs can target ~80% of all the human LR5Hs and SVAs (Pontis et al. 2019). LTR5Hs were significantly overrepresented among the 256 accessible human-specific TEs (observed 87%, expected 29%; Fisher's Exact Test $p < 2.2 \times 10^{-16}$; Fig. 2b), while the SVAs were significantly underrepresented (13% observed vs 71% expected; Fisher's Exact Test $p < 2.2 \times 10^{-16}$; Fig. 2b), highlighting a potential primary role for LTR5Hs in CNCCs.

Furthermore, we wanted to confirm that the loss in chromatin accessibility observed in the +gRNA lines was a direct consequence of the CRISPRi-mediated repression. To this end, we performed ChIP-seq for H3K9me3 (paired-end 150 bp reads), which showed accumulation of this repressive histone mark at approximately 75% of the 256 accessible LTR5Hs and human-specific SVAs in the +gRNA samples (Fig. 2c). Notably, there was a remarkable overlap (92%) between the TEs that lost chromatin accessibility (Fig. 2a) and those gaining H3K9me3 (Fig. 2c).

We further investigated the genomic locations of these accessible human-specific TEs. Overall, 91% of the 256 accessible LTR5Hs and human-specific SVAs were located > 1 kb from the nearest transcription start site (TSS), and the median distance was 8.9 kb. This suggests that these accessible mobile elements could be putative CNCC enhancers. To confirm this, we leveraged publicly available H3K27ac ChIP-seq data previously generated by our group in hiPSC-derived CNCCs (Barnada et al., 2024). This analysis revealed that the vast majority of the 256 accessible human-specific TEs are also decorated by the H3K27ac, an established active enhancer mark, supporting their role as bona-fide human-specific CNCC enhancers (Supplementary Figure S1a).

To ensure that our findings were not biased by a specific differentiation protocol, we used an alternative iPSC-to-CNCC differentiation method (Bajpai et al., 2010) and performed ATAC-seq on the differentiated cells. With this protocol, migratory CNCCs are obtained in 2-3 weeks (Bajpai et al., 2010; Barnada et al., 2024; Mitchell et al., 2025; Pagliaroli et al., 2021). Using this approach, we similarly identified 374 LTR5Hs and human-specific SVA elements that are accessible in CNCCs, the majority of which exhibited reduced accessibility in the +gRNA line compared to the -gRNA control (Supplementary Fig. S1b). Moreover, comparable to our findings obtained with the 5-day protocol, LTR5Hs were significantly overrepresented in the set of accessible TEs (83% observed vs 29% expected; Fisher's Exact Test $p < 2.2 \times 10^{-16}$), while SVAs were significantly underrepresented (17% observed vs 71% expected; Fisher's Exact Test $p < 2.2 \times 10^{-16}$). Importantly, 91% of the 256 human-specific TEs identified as accessible in the 5-day protocol were also found accessible in CNCCs derived using the alternative protocol. This consistency suggests that our findings are robust and independent of the CNCC differentiation method used.

Finally, since our gRNAs also target non-human specific SVAs (i.e. SVA-A, -B, -C, -D), we examined the ATAC-seq signal across all existing SVA subfamilies. In total, 184 non-human-specific SVAs (A–D subfamilies) were accessible in CNCCs and 3922 were inaccessible (Supplementary Fig. S1c). Notably, even when considering all SVA groups, SVAs remained significantly underrepresented among accessible TEs (46% observed vs. 89% expected; Fisher's Exact Test, $p < 2.2 \times 10^{-16}$).

In summary, these experiments identified approximately 550 hominoid-specific TEs, half of which human-specific, that exhibit signatures of active enhancers in human iPSC-derived CNCCs, with LTR5Hs elements playing a prominent role. Notably, our CRISPR-based approach enabled the inducible repression of ~75% of these putative TE-derived human-specific CNCC enhancers.

Co-opted human-specific TEs are enriched with the CNCC *coordinator* motif

Next, we explored the genomic features driving the co-option of the LTR5Hs and SVAs as CNCC enhancers. We performed computational DNA motif analysis on the set of 256 accessible human-specific TEs using the non-accessible LTR5Hs and SVAs as background control for differential enrichment. Recent studies have identified a specific DNA motif, known as *coordinator*, which is enriched at enhancers critical for the regulation of CNCC identity (Kim et al., 2024; Prescott et al., 2015). *Coordinator* is a composite motif which consists of a fusion between a generic AT-rich homeobox motif (TTAATTA) and the binding motif of the CNCC master regulator TWIST1, typically joined by a stretch of A nucleotides (Kim et al., 2024; Prescott et al., 2015). Importantly, we found both components of the *coordinator motif* as highly enriched in the set of accessible LTR5Hs and SVAs (Fig. 2d; Supplementary File S2). Specifically, the TWIST1 motif and the AT-rich homeobox motif were found in 83% and 62% of accessible human-specific TEs, respectively (Fig. 2d; Supplementary File S2). Moreover, we found that 29% of the accessible human-specific SVAs and LTR5Hs harbour the full *coordinator motif* sequence, as opposed to only 14% of the non-accessible TEs. This suggests that the human-specific TEs accessible in CNCCs are significantly more enriched for the *coordinator motif* relative to the non-accessible ones (Fisher's Exact Test $p < 0.00001$). In addition, the motifs for other CNCC signature transcription factors, such as AP2α, SOX9/10, SLUG and FOXD3 were also found as significantly enriched (Fig. 2d; Supplementary File S2). Given the prevalence of the *coordinator* motif in accessible LTR5Hs and SVAs, we investigated whether these TEs are directly bound by TWIST1 in human CNCCs. To address this, we leveraged publicly available TWIST1 ChIP-seq data generated in iPSC-derived CNCCs (Kim et al., 2024), which revealed that over half of the accessible LTR5Hs and SVA elements are bound by TWIST1 in human CNCCs (Fig. 2e).

Overall, these findings suggest that DNA sequence and transcription factor binding are a major driver for TE co-option as active CNCC enhancers.

Repression of LTR5Hs and SVAs impairs expression of CNCC-migration genes

Our experiments so far identified ~550 LTR5Hs and SVAs (half of which human-specific) displaying active enhancer signature in human CNCCs. Since our CRISPRi system enables

the simultaneous repression of ~75% of these elements, we next investigated whether their repression had significant effects on the CNCC transcriptome. To this end, we differentiated our doxycycline-treated CRISPRi lines (+ and -gRNAs) into CNCCs and performed RNA-seq. First, we analyzed gene expression in relation to the nearest transcription start sites (TSS) of the 256 accessible human-specific SVAs and LTR5Hs. In total, we identified 107 genes located near these elements that were actively expressed in CNCCs (median TPM >1 across all -gRNA replicates). Overall, the expression of these genes was significantly decreased upon TE repression (Wilcoxon's Rank Sum test $p < 0.037$; Supplementary Fig. S2a), with 83/107 (=77.5%) displaying lower expression in the +gRNA sample relative to the -gRNA counterpart (Supplementary File S3).

We leveraged publicly available RNA-seq data from CNCCs derived from chimpanzee iPSCs (Prescott et al., 2015), which we reanalysed using our pipeline, to compare the expression of these 107 genes between the two closely related ape species. This analysis revealed that these genes are typically expressed at significantly higher levels in humans than in chimpanzees (Fig. 3a; Supplementary File S3). However, repressing the human-specific SVAs and LTR5Hs eliminated this expression difference between the two species (Fig. 3a), demonstrating a direct role for these TEs in species-specific CNCC gene regulation. Consistent with this, repressing the human-specific TEs increased the correlation between human and chimpanzee expression levels for these genes (Supplementary Fig. S2b).

Next, we examined transcriptome-wide effects. Differential gene expression analysis identified 795 genes that were significantly differentially expressed between +gRNA and -gRNA CNCCs (FDR < 0.05; FC < -1.5 or > 1.5; Fig. 3b; Supplementary File S4). Of these, 501 genes were downregulated, while 294 genes were upregulated in the +gRNA samples (Fig. 3b). Gene ontology enrichment analysis of the downregulated genes revealed a significant enrichment for cell migration-related processes, suggesting that many of these genes play key roles in CNCC migration (Fig. 3c; Supplementary File S5). Conversely, upregulated genes were primarily associated with mitochondrial and cell division processes (Supplementary File S6). Notably, only 3 of the differentially expressed genes had an intronic SVA or LTR5Hs, suggesting that the high number of differentially expressed genes is not an artifact of CRISPRi mediated repression at gene bodies.

Silencing human-specific TEs functionally affects CNCC migratory potential *in vitro*

Since cell migration was the predominant signature enriched in the downregulated genes, we investigated whether repressing human-specific SVAs and LTR5Hs could functionally affect CNCC migration. To test this, we performed a transwell migration assay, in which hiPSC-CNCCs were seeded on geltrex-coated transwell membranes overnight and subsequently exposed to medium supplemented with the general chemoattractant fetal bovine serum (FBS)

in the lower chamber (Fig. 4a). After 24 hours of incubation, comparison of the –gRNA and +gRNA lines revealed a significant reduction in CNCC migration across the membrane upon repression of human-specific TEs, suggesting impaired migratory capacity (Fig. 4b,c). These findings were consistent with the RNA-seq data and suggest that human-specific SVAs and LTR5Hs have been co-opted to regulate human CNCC migration.

The LTR5Hs and SVAs active in CNCCs are silent in most human cell-types

Finally, we investigated whether the 256 human-specific SVAs and LTR5Hs with active enhancer signature in CNCCs were also co-opted as cis-regulatory elements in other human cell types. To address this, we leveraged publicly available H3K27ac ChIP-seq data from 14 cell types generated by the Roadmap Epigenomics Consortium, including distinct brain regions, lung, aorta, adipose tissue, pancreas and spleen (Kundaje et al., 2015). This analysis revealed that, in addition to CNCCs, these 256 TE exhibit active enhancer signature exclusively in endomesodermal cells, with a subset also active in iPSCs (Fig. 5). In contrast, they remain completely silenced in all other cell types (Fig. 5). The activation of these elements in endomesodermal cells is unsurprising, as a previous study from our lab has shown that human-specific SVAs are highly enriched for the motif of the key mesodermal regulator EOMES (= TBR2; Patoori et al., 2022). Similarly, the co-option of SVAs and LTR5Hs as regulatory elements in human iPSCs and ESCs has been suggested by previous research (Barnada et al., 2022; Fuentes et al., 2018; Pontis et al., 2019).

In summary, these findings support the notion that TE co-option as functional cis-regulatory elements is largely a cell-type-specific process, with distinct subsets of human-specific TEs contributing to regulatory networks in different developmental contexts.

Discussion

Transposable elements (TEs) have historically been viewed as genomic parasites with little relevance beyond their self-propagation mechanisms. Yet, over the past two decades, numerous studies have shown that TEs can substantially influence gene regulatory architectures, driving lineage-specific developmental programs and morphological variation (Chuong et al., 2017; Feschotte, 2008; Fueyo et al., 2022; Patoori et al., 2022; Sundaram & Wysocka, 2020; Trizzino et al., 2017). In this regard, the vertebrate cranial neural crest cell (CNCC) population, which plays a pivotal role in craniofacial morphogenesis, offers an especially intriguing system for the investigation of TE co-option as a source of regulatory novelty (Bronner & LeDouarin, 2012; Gokhman et al., 2021; Minoux & Rijli, 2010; Prescott et al., 2015). Over the course of vertebrate evolution, craniofacial diversity has been shaped by small but potent modifications in the gene regulatory circuits of the neural crest (Gokhman et al., 2021; Prescott et al., 2015). In this context, our findings suggest that hominoid-specific

TEs contribute to these modifications by providing new enhancer platforms that alter CNCC gene expression, pointing to a direct role for TEs in the emergence of human-specific craniofacial features (Gokhman et al., 2021; Prescott et al., 2015).

CNCCs are a multipotent, migratory cell population that emerges from the border region of the neural tube and subsequently disperses into developing craniofacial structures, the peripheral nervous system, and other tissues (Simões-Costa et al., 2015; Theveneau & Mayor, 2012). Understanding how species-specific genomic elements, such as hominoid-specific TEs, modulate CNCC development can shed light on the evolutionary mechanisms underlying morphological divergence, particularly in the craniofacial region (Capra et al., 2013; Prescott et al., 2015). Specifically, in this study we demonstrate that LTR5Hs and SVAs, two families of TEs unique to hominoids, function as enhancers within human CNCCs. Using integrative approaches that combined chromatin accessibility profiling, transcriptomic analysis, and functional perturbation via CRISPR-interference, we reveal that hundreds of these TEs display active enhancer signatures and exert direct regulatory control over genes crucial for CNCC migration. These findings are in line with previous studies showing that TEs can be repurposed as developmental enhancers, thereby contributing to species-specific gene regulatory landscapes in primates (Jacques et al., 2013; Patoori et al., 2022; Trizzino et al., 2017). Our data also implicate these TE-derived enhancers in driving human-specific craniofacial regulatory programs, as evidenced by the fact that repressing LTR5Hs and SVAs reduces the expression of dozens of genes in human CNCCs to levels normally observed in chimpanzees. Notably, LTR5Hs and SVAs have historically been linked to regulatory innovation in primates, though their roles in neural crest biology have remained unexplored (Fuentes et al., 2018; Imbeault et al., 2017; Patoori et al., 2022; Pontis et al., 2019).

One of the central questions in evolutionary developmental biology is how minor alterations in deeply conserved developmental pathways can yield significant morphological divergence (Carroll, 2005; King & Wilson, 1975). The CNCC population, which is indispensable for craniofacial bone and cartilage formation, serves as an excellent proof of principle in this regard. It has been suggested that relatively small changes in CNCC gene expression can profoundly impact facial shape and size (Khouri-Farah et al., 2025; Le Douarin & Dupin, 2018; Minoux & Rijli, 2010; Mitchell et al., 2025; Prescott et al., 2015). Our study extends this framework by proposing that TEs represent a flexible reservoir of regulatory motifs capable of integrating into pre-existing gene regulatory networks, fine-tuning CNCC specification and migration. These findings are consistent with previous reports showing that TEs have been exapted into tissue-specific enhancers in various lineages, including immune cells, brain cell-types, embryonic stem cells, and endometrial tissue (Chuong et al., 2017; Frost et al., 2023; Lynch et al., 2011; Trizzino et al., 2018; J. Wang et al., 2014).

While correlative evidence can implicate TEs as enhancers, functional validation is key to establishing their regulatory impact. By employing a CRISPRi approach that targeted approximately 75% of all the human LTR5Hs and SVAs active in CNCCs, we observed pronounced effects on gene expression with a strong enrichment for processes related to cell migration, which is indispensable for CNCCs to navigate into the developing facial primordia (Theveneau & Mayor, 2012). These results not only reinforce the notion that TEs influence CNCC gene networks but also demonstrate that their regulatory input is significant enough to modulate key developmental pathways. This model agrees with evidence from other contexts in which TEs have driven evolutionary innovations. For instance, MER41 elements in the human genome have been exapted as enhancers regulating immune responses (Chuong et al., 2017). Similarly, in the endometrium and placenta, endogenous retroviral LTRs have been implicated in the evolution of pregnancy-specific gene regulatory networks (Frost et al., 2023; Lynch et al., 2011). Our study adds craniofacial development to the growing list of developmental systems shaped by TE co-option, further highlighting the broad evolutionary relevance of these mobile elements.

TE co-option relies on the inherent or mutated presence of transcription factor binding sites within TE sequences (Barnada et al., 2022; Bejerano et al., 2006; Feschotte, 2008; Patoori et al., 2022; Trizzino et al., 2017, 2018). In this study we show that LTR5Hs and SVAs domesticated as enhancers in CNCCs harbour binding sites for the CNCC *coordinator* motif (Kim et al., 2024; Prescott et al., 2015), as well as motifs for TFAP2A, SOX9/10 and other key regulators of neural crest identity. We hypothesize that these factors bind cooperatively to multi-part motifs within the TE loci, enabling the formation of enhancer complexes that drive spatial and temporal gene expression programs in CNCCs.

Disruptions to CNCC gene regulatory networks can underlie congenital anomalies such as cleft lip and palate, craniosynostosis, and other craniofacial malformations (Khouri-Farah et al., 2025; Trainor, 2010). Given that TE-derived enhancers are integral to the CNCC transcriptional program, mutations or epigenetic dysregulations at these loci could contribute to such disorders. Future studies investigating the links between TE variations and craniofacial pathologies may reveal novel diagnostic markers or therapeutic targets.

In addition, comparative investigations in other hominoids, including gorillas and orangutans, could help unravel whether other TEs have been similarly co-opted in the cranial neural crest. Such research would clarify whether TE-driven craniofacial enhancer innovations are a hallmark across all hominoids or largely unique to the human lineage, shedding light on how small-scale genetic elements can produce macro-evolutionary changes in form and function.

In conclusion, our findings provide further evidence for TEs as highly adaptive genomic elements that actively contribute to shaping developmental and evolutionary trajectories. Through combined roles in CNCC specification, migration, and potentially other lineage-

specific developmental programs, TEs demonstrate how evolutionary tinkering at the regulatory level can yield morphological diversity.

Materials and Methods

Generation of doxycycline-inducible dCas9-KRAB hiPSC lines

A plasmid containing a tetracycline-inducible dCas9-KRAB expression cassette flanked by piggyBac recombination sites was obtained from the Wysocka Lab at Stanford University. This plasmid, referred to as 'p-dCas9-KRAB,' provides constitutive puromycin resistance, enabling selection of stable clones when co-expressed with the piggyBac transposase plasmid ('p-PB-Transposase,' Systems Bioscience). For the -gRNAs line, the C2277 hiPSC line was co-transfected with p-dCas9-KRAB and p-PB-Transposase by Applied StemCell Inc. (C2277-A, -gRNAs line). Next, a piggyBac transposon plasmid ('p-sgRNA,' Systems Bioscience) containing two sgRNAs targetting ~80% of annotated SVAs and LTR5Hs in humans was obtained based on a previously published study(Pontis et al., 2019). This plasmid confers constitutive dual sgRNA expression and geneticin resistance. For the +gRNAs line, the C2277 hiPSC line was co-transfected with p-dCas9-KRAB, p-PB-Transposase, and p-sgRNA (C2277-B, +gRNA line). Puromycin (0.125 µg/mL) only (-gRNAs line) or puromycin (0.125 µg/mL) and geneticin (100 µg/mL) (+gRNAs line) were introduced for multiple days to obtain purified colonies.

iPSC culture and NCC differentiation

Human iPSC lines C2277-A (-gRNAs) and C2277-B (+gRNAs) were cultured in feeder-free, serum-free 2D culture at 37°C and 5% CO_2. For the generation of NCCs, the differentiation method was derived from the protocol published by Leung et al. (2016). Briefly, hiPSCs were grown in mTeSR Plus (StemCell Technologies, 100-0276) on Geltrex-coated (Thermofisher, A1413302) 6-well plates until reaching ~80% confluency. Differentiation was initiated by plating $2x10^6$ hiPSCs in Neural Crest Differentiation media (DMEM/F12 (Gibco, A4192002), 2% B-27 supplement (Fisher Scientific, 15717988), 3 µM CHIR99021 (Stratech, S1263-SEL-5mg), 0.5% bovine serum albumin (Fisher Scientific, 12881630), 1X glutaMAX supplement (Gibco, 35050061), 1% penicillin-streptomycin (Gibco, 15070063)) onto Geltrex-coated 6-well plates supplemented with 10 µM Y-27632 (Cambridge Bioscience, HY-119937-5MG). The following days daily media changes were performed and 2 µg/mL doxycycline was added to the culture media from day 2 onwards (i.e. 24h after the start of differentiation). NCC differentiation was assessed after 5 days.

Immunocytochemistry

HiPSCs-NCCs were harvested by incubation with StemPro Accutase (Gibco, A1110501) for 5-10 min followed by aspiration and detachment with DMEM/F12, and subsequently seeded onto 16 mm coverslips in 12-well plates coated with Geltrex and left overnight to attach before fixation with 4% paraformaldehyde/PBS for 15 min at 37°C. After washing with PBS, the fixed cells were stored in PBS until staining. HiPSC-NCCs were subsequently permeabilised with 0.1% Triton-X-100/PBS (Merck, 648463) for 10 min at RT and blocked using 10% donkey serum/PBS (Abcam, AB7475) for 1h at RT. Cells were stained with primary antibodies overnight at 4°C in the dark, washed and stained with secondary antibodies for 30 min at 37°C. The antibodies used can be found in Supplementary Table S1. After a subsequent washing step, cells were counterstained with 1 µg/mL DAPI (BioLegend, 422801). Washing steps were performed with 0.1% Tween:PBS (Promega, H5152). Images were acquired using a Zeiss Axio Observer and further processing was done using ImageJ. Normalised intensity was calculated by dividing the signal intensity of each frame by the number of cells.

Western blot

Cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (Fisher Scientific, 10230544) supplemented with a Pierce protease inhibitor tablet (Fisher Scientific, 15614189). The lysates were vortexed for 30 s and incubated on ice for 30 min. Samples were then centrifuged at 17,000 g for 10 min at 4°C, and the supernatant was collected. Protein concentration was determined using a Pierce BCA Protein Assay Kit (Fisher Scientific, 23225) according to the manufacturer's protocol. The plate was incubated at 37°C for 30 min and analyzed using the SpectraMax M2 microplate reader (Molecular Devices) at 562 nm. 10 ug of protein were denatured in Laemmli Buffer (Fisher Scientific, J61337.AC) at 95°C for 5 min. Proteins were separated using a 4–12% Novex Tris-Glycine Mini Protein Gel (Invitrogen, XP04122BOX) in Novex Tris-Glycine SDS Running Buffer (Fisher Scientific, 11559066) for 90 min. Proteins were transferred onto a nitrocellulose membrane using the Mini Trans-Blot Cell system with NuPAGE Transfer Buffer (Life Technologies, NP00061) for 1h. Membranes were blocked in Li-Cor Intercept Blocking Buffer (Li-Cor, 927-70001) for 1h at RT, then incubated overnight at 4°C with primary antibodies (see Table S1) diluted in Intercept Blocking Buffer with 0.2% Tween. The following day, washes were performed in 0.1% Tween/PBS and membranes were incubated with secondary antibodies (see Supplementary Table S1) for 30 min before imaging on a LiCor Odyssey XF imaging system.

Real-time qPCR

Total RNA was isolated using the Monarch Total RNA miniprep kit (New England BioLabs, T2010) following the manufacturer's instructions. Briefly, harvested cells were lysed in RLT

buffer. The lysate was then passed through a gDNA column and centrifuged at 16,000 g for 1 min. Next, 70% ethanol was added to the lysate, which was transferred to an RNeasy silica column and centrifuged at 16,000 g for 1 min. The column was washed followed by DNase I treatment for 15 min at RT. Additional wash steps were performed followed by a final dry centrifugation at 12,000 rpm for 1 minute. RNA was eluted in RNase-free ddH_2O, and centrifuged at 16,000 g for 1 min. The eluted RNA was either used immediately or stored at -80°C. Reverse transcription into cDNA was performed using the Maxima First Strand cDNA Synthesis kit with the following thermal cycler settings: 10 min at 25°C, 15 min at 50°C, and 5 min at 85°C. Gene expression was quantified via real-time PCR PowerUp SYBR Green Supermix (Life Technologies, A46112). Primers for each gene are listed in Supplementary Table S2. GAPDH was used as housekeeping gene. RT-qPCR was performed using a CFX Connect Real-Time PCR Detection System (Bio-rad) with the following cycling parameters: 5 min at 95°C, 40 cycles of 15 s at 95°C, 30 s at 60°C, and 30 s at 72°C.

Transwell migration assay

24-well inserts (8 μM, Starstedt) were coated on both sides with ~22 μg/cm² Geltrex for 1 h at 37°C. After removing excess liquid, 1×10^5 hiPSC-NCCs were seeded in Neural Crest Differentiation media + 10 μM Y-27632 in the upper compartment and incubated overnight for attachment. The following day, the media was replaced and Neural Crest Differentiation media supplemented with 10% FBS was added to the lower compartment. Migration was assessed by fixing the membranes in 4% paraformaldehyde/PBS for 15 min at RT, followed by PBS washing and DAPI staining for 15 min at RT. To ensure visualization of migrated cells only, the upper surface of the insert was carefully swabbed with a cotton swab before fixation. Membranes were then excised and mounted on microscope slides for imaging. 10X images were captured from four different fields of view, and the average number of migrated cells was counted. Images were acquired using a Zeiss Axio Observer and migration was quantified using ImageJ.

Bulk RNA-seq library preparation

Total RNA was isolated using the Monarch Total RNA miniprep kit (New England BioLabs, T2010) following the manufacturer's instructions. RNA was quantified using a Nanodrop spectrophotometer and RNA integrity was checked on a Tapestation 2200 (Agilent). Only samples with a RIN > 7.0 were used for transcriptome analysis. RNA libraries were prepared using 300 ng of total RNA input using the NEBNext Poly(A) mRNA Magnetic Isolation Module(E7490) and NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (E7760) according to the manufacturer's instructions. PE150 sequencing was performed on a NovaSeq X Plus instrument (Illumina).

RNA-seq analysis

Adapters were removed using TrimGalore!, and gene-level read counts were obtained with Kallisto (Bray et al., 2016). Differential gene expression analysis was performed using DESeq2 (Love et al., 2014). GO enrichment analysis was performed using Webgestalt 2019 (Liao et al., 2019). Additional statistical analysis was carried out using R (version 4.2.2) and GraphPad Prism 10.

ChIP-seq library preparation

10 million hiPSC-NCCs were crosslinked in 1% paraformaldehyde for 5 min at RT with mild rotation. Crosslinking was quenched with 0.125M glycine by rotation for 5 min at RT. Cells were washed twice with cold PBS and were snap-frozen for 15 min on dry ice. Samples were stored at -80 °C until further processing. Cells were resuspended in 1 mL ChIP buffer (150 mM NaCl, 1% Triton X-100, 500 mM DTT, 10 mM Tris-HCl, 5 mM EDTA) supplemented with Pierce protease inhibitor tablet and incubated on ice for 10 min. SDS was added to each sample to a final concentration of 0.3%, and chromatin was sheared using the following settings: 9 min, duty factor 5%, 6°C in a Covaris M220 Focused-Ultrasonicator. Chromatin fragment size was assessed on a Tapestation 2200.

The chromatin lysates were then centrifuged at 13,000 g for 10 min at 4 °C and supernatant was subsequently incubated with ChIP buffer supplemented with Pierce protease inhibitor and Dynabeads Protein G magnetic beads (Invitrogen) along with 3 µg of H3K9me3 antibody (Abcam, ab8898) and incubated overnight at 4 °C under mild rotation. The following day, samples were placed in a magnetic rack and washed twice with Mixed Micelle Wash Buffer (150 mM NaCl, 1% Triton X-100, 0.2% SDS, 20 mM Tris-HCl, 5 mM EDTA, 65% sucrose), twice with Buffer 200 (200 mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), twice with LiCl/detergent wash (250 mM LiCl, 0.5% sodium deoxycholate, 0.5% NP-40, 10 mM Tris-HCl, 1 mM EDTA), once with cold TE. The beads were resuspended in TE + 1% SDS and incubated at 65 °C for 10 min at 1200 rpm to elute immunocomplexes. The elution was repeated twice and the samples were incubated overnight at 65˚C to reverse cross-linking. The following day, Proteinase K (0.5 mg/mL) was added to digest samples at 65 °C for 1h. DNA was purified using the Zymo ChIP DNA Clean and Concentrator kit (Zymo, D5205) and quantified with the Quantifluor ONE dsDNA system (Promega, E4871). DNA libraries were prepared using the NEBNext Ultra II DNA library Prep Kit (E7645L) for Illumina. PE150 sequencing was performed on a NovaSeq X Plus instrument (Illumina).

ChIP-seq analysis

Adapters were removed with TrimGalore! and reads were aligned to hg38 using Burrows-Wheeler Alignment tool, with the MEM algorithm (Li & Durbin, 2009). Uniquely mapping reads were filtered (MAPQ>10) and PCR duplicates were removed and mitochondrial reads were discarded. Peaks were called using HOMER with default parameters at 5% FDR (Heinz et al., 2010). All statistical analyses were performed using BEDTools (Quinlan & Hall, 2010), deepTools (Ramírez et al., 2014), R (version 4.2.2) and GraphPad Prism 10.

ATAC-seq library preparation

DNA libraries were prepared following the ATAC-Seq kit (Active Motif, 53150) according to the manufacturer's instructions. 100,000 cells were tagmented per sample, DNA was purified using SPRI beads and amplified with dual index primers. Libraries were assessed for size distribution using the TapeStation 2200 and PE150 sequencing was performed on a NovaSeq X Plus instrument (Illumina).

ATAC-seq analysis

Adapters were removed with TrimGalore! and reads were aligned to hg38 using Burrows-Wheeler Alignment tool, with the MEM algorithm (Li & Durbin, 2009). With SAMTools (Li et al., 2009), uniquely mapping reads were filtered (MAPQ>10), PCR duplicates were removed and mitochondrial reads were discarded. Consensus peaks were determined in each cell line using BEDTools (Quinlan & Hall, 2010). Motif analysis was performed using HOMER (Heinz et al., 2010). All further downstream analysis was performed using BEDTools and deepTools (Ramírez et al., 2014).

Statistics

RT-qPCR, transwell migration and image quantification data were analysed using Graphpad Prism 10 software (Graphpad, San Diego, CA). Data is represented as mean ± standard error of mean (SEM). Statistical significance was accepted at $p < 0.05$. Sequencing data was analysed using BEDTools, DeepTools, and R as indicated. Motif analysis was performed using HOMER. GO enrichment analysis was performed using Webgestalt 2019.

Table 1. Antibodies used in this study.

Antibody	Manufacturer	Use	Dilution/quantity
Mouse AP2α	Fisher Scientific, 11594723	IF	1:100
Rabbit Sox9	Abcam, AB185230-1001	IF	1:250

Goat anti-Mouse IgG2b Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	Life Technologies, A21242	IF	1:500
Goat anti-Mouse IgG2b Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	Life Technologies, A21242	IF	1:500
Mouse Cas9	Cambridge bioscience, 61757	WB	1:1000
Rabbit GAPDH (D16H11)	Cell Signalling Technology, 5174	WB	1:1000
Goat anti-mouse IgG IRDye® 800CW	LI-COR, 926-32210	WB	1:15,000
Goat anti-Rabbit IgG IRDye® 800CW	LI-COR, 926-32211	WB	1:15,000
Rabbit H3 tri methyl K9 - ChIP Grade	Abcam, ab8898	ChIP	3 ug

Table 2. Primers used in this study for RT-qPCR.

Gene	**Forward (5'-3')**	**Reverse (5'-3')**
SOX9	GTACCCGCACTTGCACAAC	TCTCGCTCTCGTTCAGAAGTC
SOX10	GAGGGCTCCCCCATGTCAGAT	GTCTGCCTTGCCCGACTGC
TFAP2A	GCCTCTCGCTCCTCAGCTCC	CGTTGGCAGCTTTACGTCTCCC
TWIST1	GCCAGGTACATCGACTTCCTCT	TCCATCCTCCAGACCGAGAAGG
GAPDH	GAACGGGAAGCTTGTCATCAA	ATCGCCCCACTTGATTTTGG

Acknowledgements

The authors thank the Wysocka group at Stanford (and particularly Dr. Raquel Fueyo) for providing the dCAS9-KRAB vector, and Andrew Isopi and Dr. Samantha Barnada (Thomas Jefferson University), for designing the plasmid with the sg-RNAs. We thank the Facility for Imaging by Light Microscopy (FILM) at Imperial College London for providing access to instrumentation and technical support, which is part-supported by funding from the Wellcome Trust (grant 104931/Z/14/Z) and BBSRC (grant BB/L015129/1). For this work, MT was funded by Biotechnology and Biological Sciences Research Council (BBSRC).

Data availability

RNA-seq, ATAC-seq, and ChIP-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession code GEO: GSE292478 and are publicly available as of the date of manuscript submission

Author contributions

MT and LD designed the project. LD performed most of the experiments. ZHM, MD and BGDV contributed to some of the experiments. LD and MT analyzed the data and wrote the manuscript. All the authors read and approved the manuscript.

References

Ahlstrom, J. D., & Erickson, C. A. (2009). The neural crest epithelial-mesenchymal transition in 4D: A 'tail' of multiple non-obligatory cellular mechanisms. *Development*, *136*(11), 1801–1812. https://doi.org/10.1242/dev.034785

Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C. P., Zhao, Y., Swigut, T., & Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, *463*(7283), 958–962. https://doi.org/10.1038/nature08733

Barnada, S. M., Giner de Gracia, A., Morenilla-Palao, C., López-Cascales, M. T., Scopa, C., Waltrich, F. J., Mikkers, H. M. M., Cicardi, M. E., Karlin, J., Trotti, D., Peterson, K. A., Brugmann, S. A., Santen, G. W. E., McMahon, S. B., Herrera, E., & Trizzino, M. (2024). ARID1A-BAF coordinates ZIC2 genomic occupancy for epithelial-to-mesenchymal transition in cranial neural crest specification. *The American Journal of Human Genetics*. https://doi.org/10.1016/j.ajhg.2024.07.022

Barnada, S. M., Isopi, A., Tejada-Martinez, D., Goubert, C., Patoori, S., Pagliaroli, L., Tracewell, M., & Trizzino, M. (2022). Genomic features underlie the co-option of SVA transposons as cis-regulatory elements in human pluripotent stem cells. *PLoS Genetics*, *18*(6), 1–19. https://doi.org/10.1371/journal.pgen.1010225

Bejerano, G., Lowe, C. B., Ahituv, N., King, B., Siepel, A., Salama, S. R., Rubin, E. M., Kent, W. J., & Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, *441*(1), 87–90. https://doi.org/10.1038/nature04696

Bourque, G., Leong, B., Vega, V. B., Chen, X., Yen, L. L., Srinivasan, K. G., Chew, J. L., Ruan, Y., Wei, C. L., Huck, H. N., & Liu, E. T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research*, *18*(11), 1752–1762. https://doi.org/10.1101/gr.080663.108

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Bronner, M. E., & LeDouarin, N. M. (2012). Development and evolution of the neural crest: An overview. In *Developmental Biology* (Vol. 366, Issue 1, pp. 2–9). Academic Press Inc. https://doi.org/10.1016/j.ydbio.2011.12.042

Capra, J. A., Hubisz, M. J., Kostka, D., Pollard, K. S., & Siepel, A. (2013). A Model-Based Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes. *PLoS Genetics*, *9*(8). https://doi.org/10.1371/journal.pgen.1003684

Carroll, S. B. (2005). Evolution at two levels: On genes and form. In *PLoS Biology* (Vol. 3, Issue 7, pp. 1159–1166). Public Library of Science. https://doi.org/10.1371/journal.pbio.0030245

Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, *351*(6277), 1083–1087. https://doi.org/10.1126/science.aad5497

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. In *Nature Reviews Genetics* (Vol. 18, Issue 2, pp. 71–86). Nature Publishing Group. https://doi.org/10.1038/nrg.2016.139

Chuong, E. B., Rumi, M. A. K., Soares, M. J., & Baker, J. C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nature Genetics*, *45*(3), 325–329. https://doi.org/10.1038/ng.2553

Cordero, D. R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., & Helms, J. A. (2011). Cranial neural crest cells on the move: Their roles in craniofacial development. *American Journal of Medical Genetics, Part A*, *155*(2), 270–279. https://doi.org/10.1002/ajmg.a.33702

Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J., & Feschotte, C. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*, *371*(6531). https://doi.org/10.1126/science.abc6405

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Genetics*, *9*, 397–405.

Frost, J. M., Amante, S. M., Okae, H., Jones, E. M., Ashley, B., Lewis, R. M., Cleal, J. K., Caley, M. P., Arima, T., Maffucci, T., & Branco, M. R. (2023). Regulation of human trophoblast gene expression by endogenous retroviruses. *Nature Structural and Molecular Biology*, *30*(4), 527–538. https://doi.org/10.1038/s41594-023-00960-6

Fuentes, D., Swigut, T., & Wysocka, J. (2018). Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *ELife*. https://doi.org/10.7554/eLife.35989.001

Fueyo, R., Judd, J., Feschotte, C., & Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. *Nature Reviews Molecular Cell Biology*, *23*(7), 481–497. https://doi.org/10.1038/s41580-022-00457-y

Gans, C., & Northcutt, R. G. (1983). Neural Crest and the Origin of Vertebrates: A New Head. *Science*, *220*(4594), 268–273. https://doi.org/10.1126/science.220.4594.268

Gokhman, D., Agoglia, R. M., Kinnebrew, M., Gordon, W., Sun, D., Bajpai, V. K., Naqvi, S., Chen, C., Chan, A., Chen, C., Petrov, D. A., Ahituv, N., Zhang, H., Mishina, Y., Wysocka, J., Rohatgi, R., & Fraser, H. B. (2021). Human–chimpanzee fused cells reveal cis-regulatory divergence underlying skeletal evolution. *Nature Genetics*, *53*(4), 467–476. https://doi.org/10.1038/s41588-021-00804-3

Goubert, C., Zevallos, N. A., & Feschotte, C. (2020). Contribution of unfixed transposable element insertions to human regulatory variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1795). https://doi.org/10.1098/rstb.2019.0331

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Imbeault, M., Helleboid, P. Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, *543*(7646), 550–554. https://doi.org/10.1038/nature21683

Jacques, P. É., Jeyakani, J., & Bourque, G. (2013). The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements. *PLoS Genetics*, *9*(5). https://doi.org/10.1371/journal.pgen.1003504

Jheon, A. H., & Schneider, R. A. (2009). The cells that fill the bill: Neural crest and the evolution of craniofacial development. In *Journal of Dental Research* (Vol. 88, Issue 1, pp. 12–21). https://doi.org/10.1177/0022034508327757

Khouri-Farah, N., Winchester, E. W., Schilder, B. M., Robinson, K., Curtis, S. W., Skene, N. G., Leslie-Clarkson, E. J., & Cotney, J. (2025). *Gene expression patterns of the developing human face at single cell resolution reveal cell type contributions to normal facial variation and disease risk*. https://doi.org/10.1101/2025.01.18.633396

Kim, S., Morgunova, E., Naqvi, S., Goovaerts, S., Bader, M., Koska, M., Popov, A., Luong, C., Pogson, A., Swigut, T., Claes, P., Taipale, J., & Wysocka, J. (2024). DNA-guided transcription factor cooperativity shapes face and limb mesenchyme. *Cell*. https://doi.org/10.1016/j.cell.2023.12.032

King, M.-C., & Wilson, A. C. (1975). Evolution at Two Levels in Humans and Chimpanzees. *Science*, *188*, 107–116. https://www.science.org

Kunarso, G., Chia, N. Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y. S., Ng, H. H., & Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics*, *42*(7), 631–634. https://doi.org/10.1038/ng.600

Le Douarin, N. M., & Dupin, E. (2018). The “beginnings” of the neural crest. In *Developmental Biology* (Vol. 444, pp. S3–S13). Elsevier Inc. https://doi.org/10.1016/j.ydbio.2018.07.019

Leung, A. W., Leung, A. W., Murdoch, B., Salem, A. F., Prasad, M. S., Gomez, G. A., & García-Castro, M. I. (2016). WNT/β-catenin signaling mediates human neural crest induction via a

pre-neural border intermediate. *Development (Cambridge)*, *143*(3), 398–410. https://doi.org/10.1242/dev.130849
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, *47*(W1), W199–W205. https://doi.org/10.1093/nar/gkz401
Lieberman, D. E. (1998). Sphenoid shortening and the evolution of modern human cranial shape. *Nature*, *393*(6681), 158–162. https://doi.org/10.1038/30227
Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8
Lynch, V. J., Leclerc, R. D., May, G., & Wagner, G. P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics*, *43*(11), 1154–1159. https://doi.org/10.1038/ng.917
Martik, M. L., & Bronner, M. E. (2021). Riding the crest to get a head: neural crest evolution in vertebrates. In *Nature Reviews Neuroscience* (Vol. 22, Issue 10, pp. 616–626). Nature Research. https://doi.org/10.1038/s41583-021-00503-2
Minoux, M., & Rijli, F. M. (2010). Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. In *Development* (Vol. 137, Issue 16, pp. 2605–2621). https://doi.org/10.1242/dev.040048
Mitchell, Z. H., den Hoed, J., Claassen, W., Demurtas, M., Deelen, L., Campeau, P. M., Liu, K., Fisher, S. E., & Trizzino, M. (2025). *CHD3 regulates BMP signalling response during cranial neural crest cell specification*. https://doi.org/10.1101/2025.02.14.638260
Ostertag, E. M., Goodier, J. L., Zhang, Y., & Kazazian, H. H. (2003). SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans. In *Am. J. Hum. Genet* (Vol. 73).
Pagliaroli, L., Porazzi, P., Curtis, A. T., Scopa, C., Mikkers, H. M. M., Freund, C., Daxinger, L., Deliard, S., Welsh, S. A., Offley, S., Ott, C. A., Calabretta, B., Brugmann, S. A., Santen, G. W. E., & Trizzino, M. (2021). Inability to switch from ARID1A-BAF to ARID1B-BAF impairs exit from pluripotency and commitment towards neural crest formation in ARID1B-related neurodevelopmental disorders. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-26810-x
Patoori, S., Barnada, S. M., Large, C., Murray, J. I., & Trizzino, M. (2022). Young transposable elements rewired gene regulatory networks in human and chimpanzee hippocampal intermediate progenitors. *Development (Cambridge)*, *49*(23). https://doi.org/10.1242/dev.200413
Playfoot, C. J., Duc, J., Sheppard, S., Dind, S., Coudray, A., Planet, E., & Trono, D. (2021). Transposable elements and their KZFP controllers are drivers of transcriptional innovation in the developing human brain. *Genome Research*, *31*(9), 1531–1545. https://doi.org/10.1101/gr.275133.120
Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T. W., Jaenisch, R., & Trono, D. (2019). Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell*, *24*(5), 724-735.e5. https://doi.org/10.1016/j.stem.2019.03.012
Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F. H., Swigut, T., & Wysocka, J. (2015). Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell*, *163*(1), 68–83. https://doi.org/10.1016/j.cell.2015.08.036
Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Quinn, J. P., & Bubb, V. J. (2014). SVA retrotransposons as modulators of gene expression. *Mobile Genetic Elements*, *4*(4), e32102. https://doi.org/10.4161/mge.32102

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*(W1). https://doi.org/10.1093/nar/gku365

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–329. https://doi.org/10.1038/nature14248

Sambataro, S., Fiorillo, L., Bocchieri, S., Stumpo, C., Cervino, G., Herford, A. S., & Cicciù, M. (2022). Craniofacial Evolution: From Australopithecus to Modern Man. *Journal of Craniofacial Surgery*, *33*(1), 325–332. https://doi.org/10.1097/SCS.0000000000007841

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonalves, Â., Kutter, C., Brown, G. D., Marshall, A., Flicek, P., & Odom, D. T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, *148*(1–2), 335–348. https://doi.org/10.1016/j.cell.2011.11.058

Simões-Costa, M., Stone, M., & Bronner, M. E. (2015). Axud1 Integrates Wnt Signaling and Transcriptional Inputs to Drive Neural Crest Formation. *Developmental Cell*, *34*(5), 544–554. https://doi.org/10.1016/j.devcel.2015.06.024

Spoor, F., Wood, B., & Zonneveld, F. (1994). Implications of early hominid labyrinthine morphology for evolution of human bipedal locomotion. *Nature*, *369*(6482), 645–648. https://doi.org/10.1038/369645a0

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P., & Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research*, *24*(12), 1963–1976. https://doi.org/10.1101/gr.168872.113

Sundaram, V., & Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Issue 1795). Royal Society Publishing. https://doi.org/10.1098/rstb.2019.0347

Theveneau, E., & Mayor, R. (2012). Neural crest migration: interplay between chemorepellents, chemoattractants, contact inhibition, epithelial–mesenchymal transition, and collective cell migration. *WIREs Developmental Biology*, *1*(3), 435–445. https://doi.org/10.1002/wdev.28

Trainor, P. A. (2010). Craniofacial birth defects: The role of neural crest cells in the etiology and pathogenesis of Treacher Collins syndrome and the potential for prevention. In *American Journal of Medical Genetics, Part A: Vol. 152 A* (Issue 12, pp. 2984–2994). https://doi.org/10.1002/ajmg.a.33454

Trizzino, M., Kapusta, A., & Brown, C. D. (2018). Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4850-3

Trizzino, M., Park, Y. S., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G. H., Lynch, V. J., & Brown, C. D. (2017). Transposable elements are the primary source of novelty in primate gene regulation. *Genome Research*, *27*(10), 1623–1633. https://doi.org/10.1101/gr.218149.116

Wang, H., Xing, J., Grover, D., Hedges Kyudong Han, D. J., Walker, J. A., & Batzer, M. A. (2005). SVA elements: A hominid-specific retroposon family. *Journal of Molecular Biology*, *354*(4), 994–1007. https://doi.org/10.1016/j.jmb.2005.09.085

Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., Schumann, G. G., Chen, W., Lorincz, M. C., Ivics, Z., Hurst, L. D., & Izsvák, Z. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*, *516*(7531), 405–409. https://doi.org/10.1038/nature13804

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. In *Nature Reviews Genetics* (Vol. 8, Issue 3, pp. 206–216). https://doi.org/10.1038/nrg2063

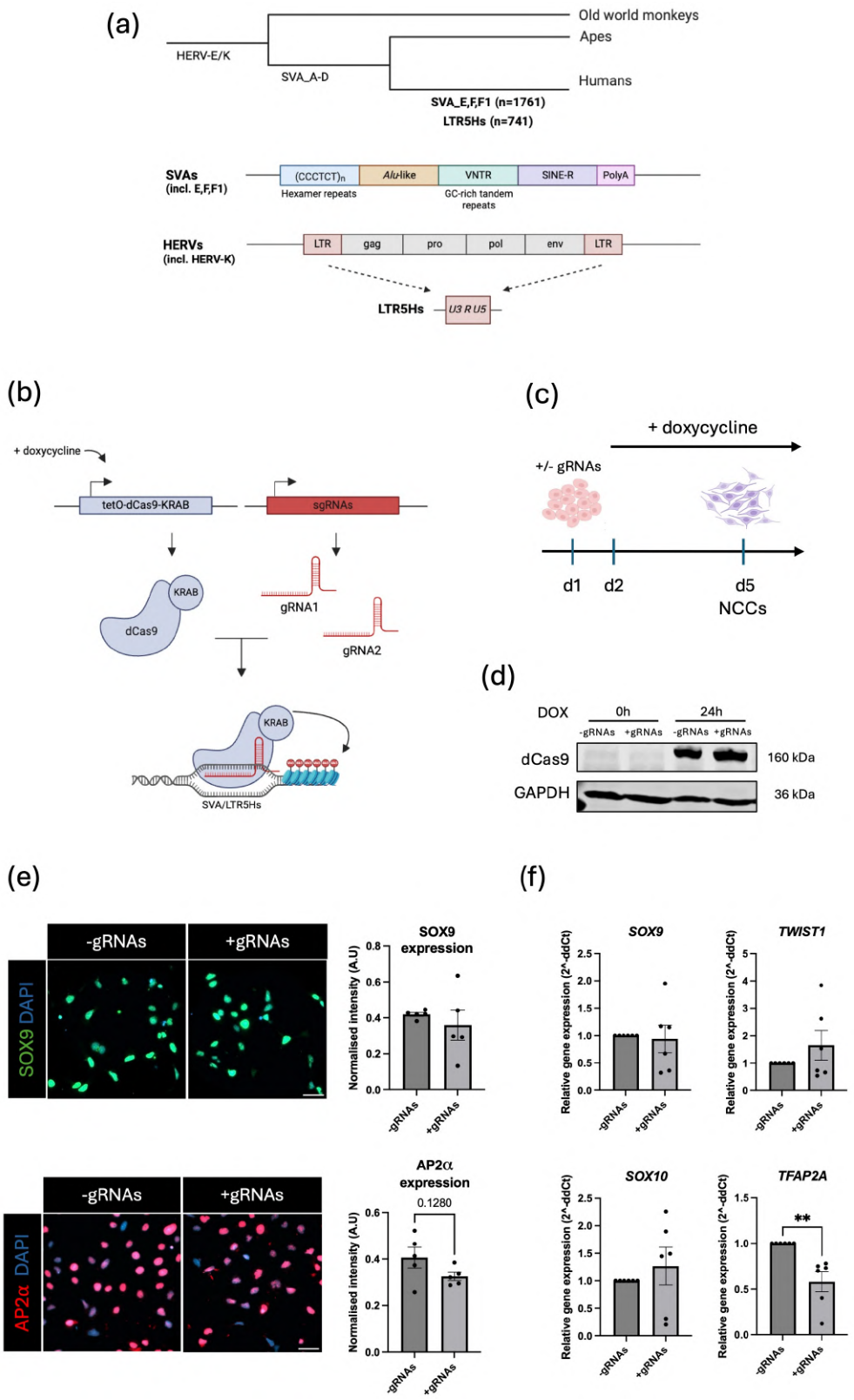

Figure 1. An inducible CRISPRi tool to repress human-specific SVAs and LR5Hs in iPSC-derived CNCCs – (a) Schematic overview of SVA and LTR5Hs transposable elements, illustrating their structure and phylogenetic distribution across Old World monkeys, apes, and humans. Human-specific SVA subfamilies (i.e., SVA_E/F/F1) and the LTR5Hs are highlighted, along with the approximate number of elements in each category. **(b)** CRISPR-dCas9-KRAB

targeting strategy for SVAs and LTR5Hs. Two guide RNAs (gRNA1 and gRNA2) target ~80% of all SVAs and LTR5Hs, directing dCas9-KRAB to deposit repressive epigenetic marks at the TEs, thereby reducing their transcriptional activity. **(c)** Experimental timeline for doxycycline (dox)-induced expression of dCas9-KRAB. Dox is added to the media 24 hours after the start of differentiation into cranial neural crest cells, which spans 5 days. **(d)** Western blot confirming dox-inducible dCas9-KRAB expression. Lysates from cells cultured with or without TE-targetting gRNAs were probed with antibodies against dCas9 (top band, ~160 kDa) and GAPDH (bottom band, loading control). **(e)** Immunofluorescence displaying expression of CNCC signature markers AP2α (red) and SOX9 (green) in CNCCs differentiated using CRISPRi-iPSCs without (left) and with gRNAs (right). Signal intensity was quantified and normalised to the cell number per frame. **(f)** RT-qPCR of CNCC markers *SOX9*, *TWIST1*, *SOX10* and *TFAP2A*. n=6. Statistical analyses were performed using unpaired t-tests. Error bars indicate mean ± SEM. **, $p<0.01$.

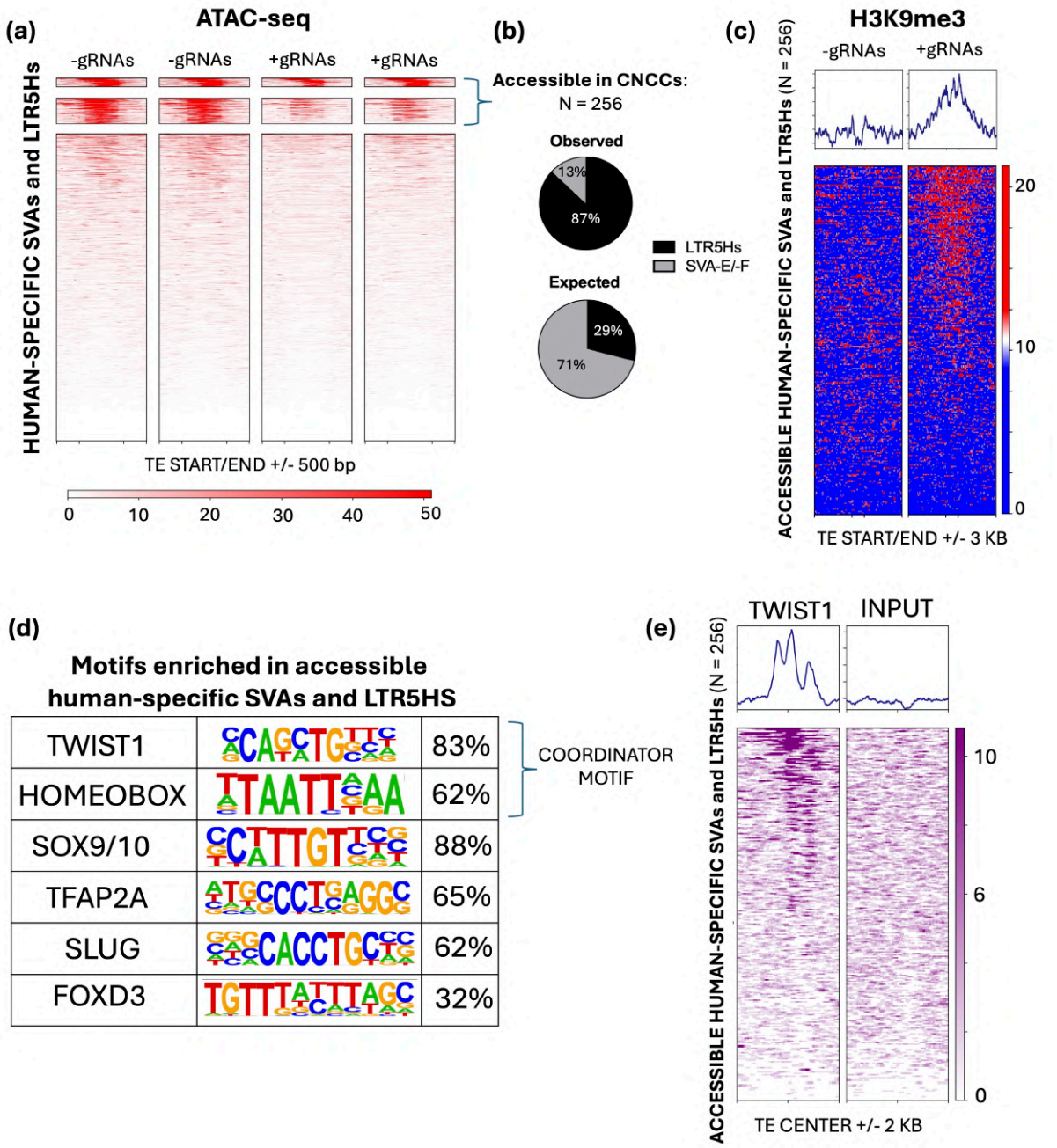

Figure 2 – Hundreds of human-specific SVAs and LTR5Hs function as enhancers in human CNCCs – (a) ATAC-seq signal across human-specific SVAs and LTR5Hs. Heatmap rows represent individual transposable elements (TEs) aligned at their start or end (±500 bp), and the intensity reflects chromatin accessibility. Clusters were generated with deepTools using k-mer algorithm. **(b)** Pie charts showing the observed versus expected fractions of accessible TEs belonging to the LTR5Hs/SVA subfamilies. **(c)** H3K9me3 enrichment at the accessible human-specific SVAs and LTR5Hs. The top panel displays the average H3K9me3 signal (average profile) aligned to the TE start or end (±3 kb), while the heatmap below shows individual TEs ranked by signal intensity. **(d)** Transcription factor motif analysis of accessible human-specific SVAs and LTR5Hs. Representative enriched motifs (TWIST1, HOMEOBOX, SOX9/10, AP2α, SLUG, and FOXD3) in accessible human-specific TEs are shown with the percentage of TEs containing each motif. **(e)** Heatmaps of TWIST1 ChIP-seq signal (left) and corresponding input control (right) centred on the accessible human-specific SVAs and LTR5Hs (±2 kb). Colour intensity indicates normalized enrichment, highlighting TWIST1 occupancy at these accessible elements.

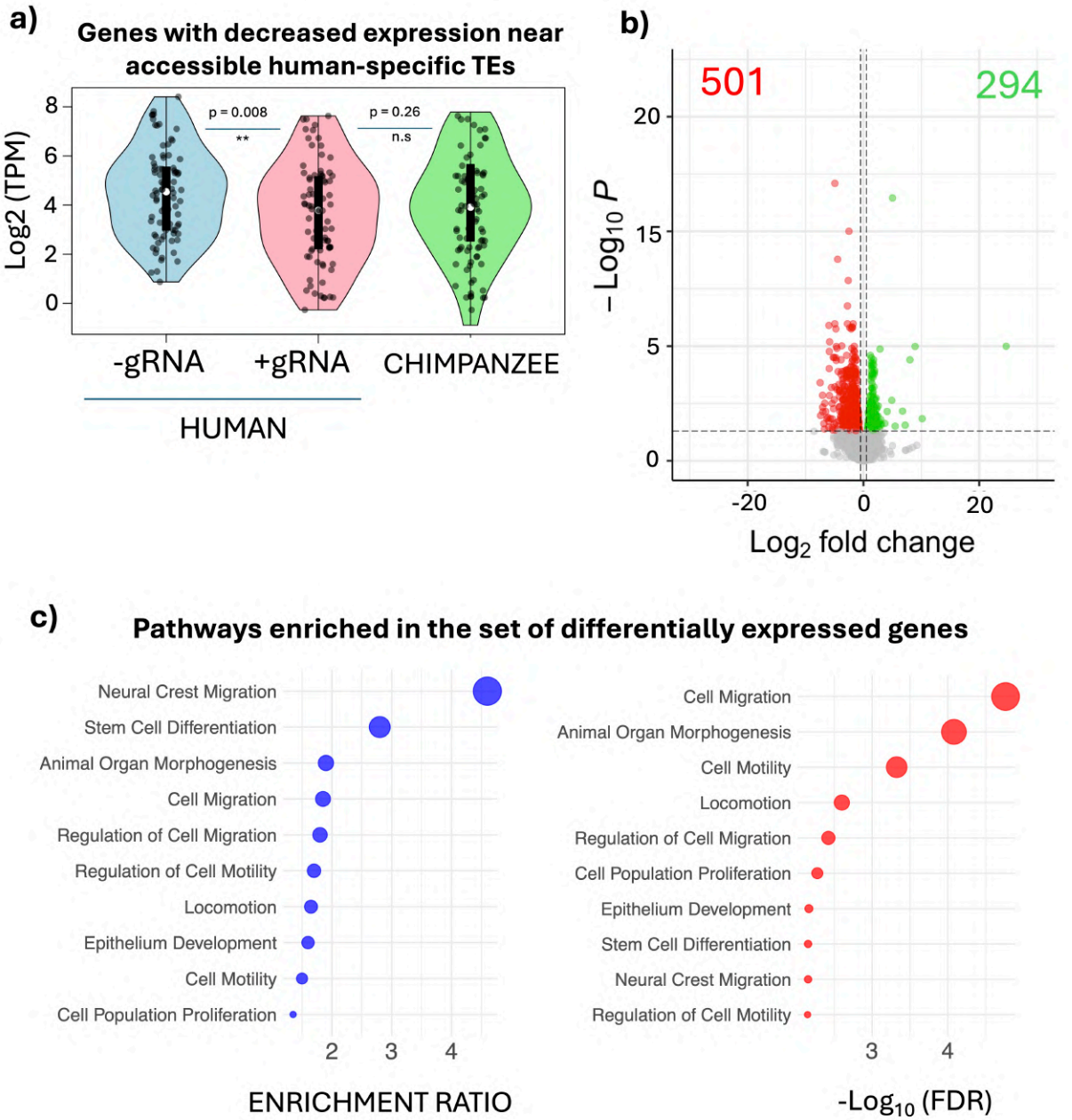

Figure 3. Human-specific SVAs and LTR5Hs regulate the expression of CNCC migration genes – (a) Violin plots displaying expression (log2 TPM) of genes located near accessible human-specific SVAs and LTR5Hs. Three conditions are shown: human CNCCs generated using CRISPRi iPSC line without guide RNAs (−gRNA), human CNCCs generated using CRISPRi iPSC line with guide RNAs (+gRNA), and chimpanzee CNCCs. The decreased expression in the +gRNA condition indicates potential regulatory contributions of these human-specific TEs. **(b)** Volcano plot illustrating differentially expressed genes between the −gRNA and +gRNA conditions in human CNCCs. The x-axis shows the log2 fold change, while the y-axis represents the −log10 P-value. Significantly upregulated (green) and downregulated (red) genes are highlighted, with total numbers indicated. Dashed lines mark common significance thresholds. **(c)** Functional enrichment analysis of downregulated genes. Bubble plots display top enriched pathways, with the x-axis indicating enrichment ratio (blue) or the -log10 FDR (red) and the y-axis listing pathway terms. The size of each bubble reflects statistical significance. Pathways involved in cell migration and motility are among the most prominently enriched.

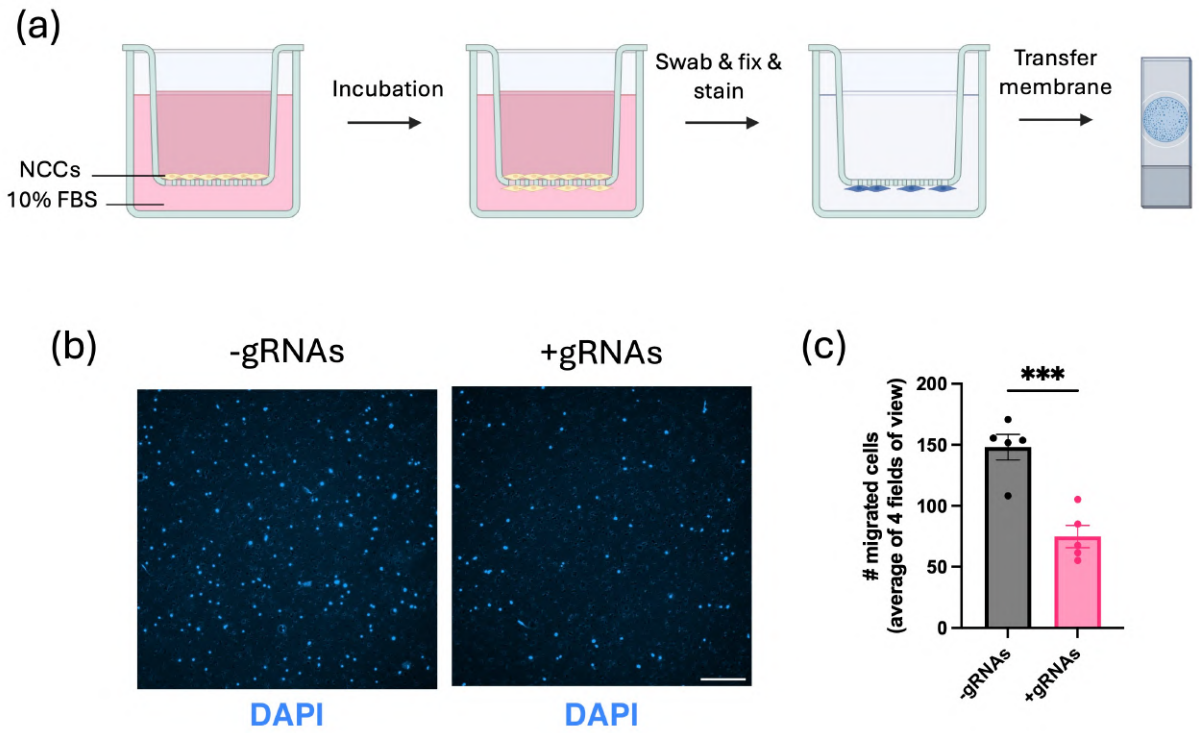

Figure 4. HiPSC-NCCs show inferior migration capability when human-specific TEs are silenced. (a) Outline of the transwell migration assay. **(b)** 1 x 10^5 hiPSC-NCCs (+/- gRNAs) were seeded overnight, followed by a media change in the upper chamber and incubation with differentiation medium +10% FBS in the lower chamber for 24h. At 24h, non-migrated hiPSC-NCCs on the top of the membrane were swabbed away and migrated hiPSC-CMs were fixed, stained and imaged. DAPI staining was used to visualise nuclei. Scale bar: 50 µm. **(c)** Quantification of transwell membranes. 10X images were taken from 4 different fields of view and the average number of migrated cells was calculated. n=5. Statistical analyses were performed using an unpaired t-test. Error bars indicate mean ± SEM. This experiment was repeated twice in two independent rounds of differentiation, which yielded comparative results.

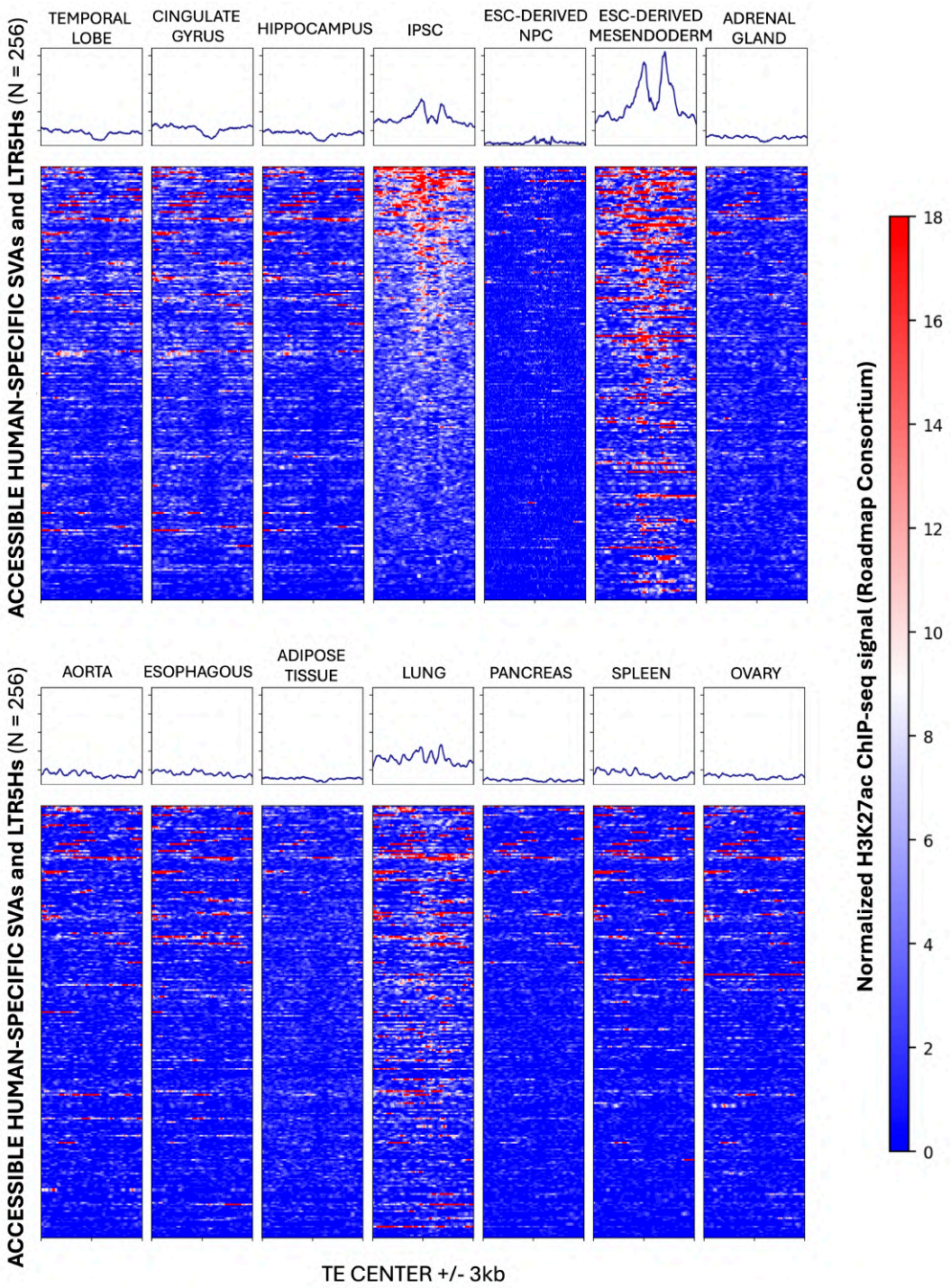

Figure 5. Co-option of SVAs and LTR5Hs as active enhancers is cell-type specific – Heatmaps of normalized H3K27ac ChIP-seq signal (Roadmap Epigenomics Consortium) across 256 accessible human-specific SVAs and LTR5Hs in multiple tissues and cell types. Each row corresponds to an individual transposable element (TE), aligned at its centre (±3 kb), while columns represent different tissues/cell lines. The colour bar on the right indicates the relative intensity of H3K27ac signal, with higher enrichment in red and lower enrichment in blue.