HHS Public Access

Author manuscript

Dev Growth Differ. Author manuscript; available in PMC 2025 January 01.

Published in final edited form as:

Dev Growth Differ. 2024 January ; 66(1): 75–88. doi:10.1111/dgd.12901.

Identification of ancestral gnathostome *Gli3* enhancers with activity in mammals

Shahid Ali[1,7], **Muhammad Abrar**[1], **Irfan Hussain**[1], **Fatima Batool**[1], **Rabail Zehra Raza**[2], **Hizran Khatoon**[1], **Matteo Zoia**[3], **Axel Visel**[4,5,6], **Neil H. Shubin**[7,*], **Marco Osterwalder**[3,8,*], **Amir Ali Abbasi**[1,*]

[1]National Center for Bioinformatics, Program of Comparative and Evolutionary Genomics, Faculty of Biological Sciences, Quaid-i-Azam University, 45320, Islamabad Pakistan.

[2]Department of Biological Sciences, Faculty of Multidisciplinary Studies, National University of Medical Sciences Rawalpindi, Pakistan.

[3]Department for Biomedical Research (DBMR), University of Bern, Bern, Switzerland.

[4]Environmental Genomics and System Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA.

[5]U.S. Department of Energy Joint Genome Institute, 1 Cyclotron Road, Berkeley, CA 94720, USA

[6]School of Natural Sciences, University of California, Merced, Merced, CA 95343, USA.

[7]Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, IL 60637, USA.

[8]Department of Cardiology, Bern University Hospital, Bern, Switzerland.

Abstract

Abnormal expression of the transcriptional regulator and hedgehog (Hh) signaling pathway effector Gli3 is known to trigger congenital disease, most frequently affecting the central nervous system (CNS) and the limbs. Accurate delineation of the genomic *cis*-regulatory landscape controlling *Gli3* transcription during embryonic development is critical for the interpretation of non-coding variants associated with congenital defects. Here we employed a comparative genomic analysis on fish species with a slow rate of molecular evolution to identify seven previously unknown conserved noncoding elements (CNEs) in *Gli3* intronic intervals (CNE15-21). Transgenic assays in zebrafish revealed that most of these elements drive activities in *Gli3* expressing tissues, predominantly the fins, CNS, and the heart. Intersection of these CNEs with human disease associated SNPs identified CNE15 as a putative mammalian craniofacial enhancer with conserved activity in vertebrates and potentially affected by mutation associated with human

[*]Corresponding Authors: abbasiam@qau.edu.pk, Tel Office : +92-51-9064-4109, Tel Lab : +92-51-9064-4114, Cell: +92-334-8959-354; nshubin@uchicago.edu, Tel: +1 773-834-7472, Fax No : +1 773-702-0037; marco.osterwalder@unibe.ch.

Authors Contributions

A. A. A., N. S. and M.O. conceived the study. A. A. A., N.S., S. A. and M.O. designed the experiments. Bioinformatics analyses were performed by S. A., F. B., M. A., R. Z. R. I. H., H.K and M. Z. Wet Lab experiments were performed by S. A, M. O., M. A. and I. H. A.V. provided reagents and infrastructure for mouse transgenic experiments. A.A.A., S. A., N.S., M.O. and R. Z. R. wrote the paper.

Competing interest

Authors declare that they have no competing interest.

craniofacial morphology. Finally, comparative functional dissection of an appendage-specific CNE conserved in slowly evolving fish (elephant shark), but not in teleost (CNE14/hs1586) indicates co-option of limb specificity from other tissues prior to the divergence of amniotes and lobe-finned fish. These results uncover a novel subset of intronic *Gli3* enhancers which arose in the common ancestor of gnathostomes and whose sequence components were likely gradually modified in other species during the process of evolutionary diversification.

Keywords

Gli family; Gli3; CNEs; Enhancers; Gar; Elephant shark; Zebrafish; Gnathostomes; Transgenesis

Introduction

The Hedgehog (Hh) pathway represents a key signaling network essential for a variety of cellular processes underlying embryonic development in metazoans (Choudhry et al., 2014). Components of the HH pathway are conserved across vertebrates and invertebrates, and play critical roles during embryonic morphogenesis linked to patterning and polarization of the neural tube and the antero-posterior axis of extremities and other developing organs (Anderson et al., 2012; Jiang & Hui, 2008). In invertebrates, Hh pathway activity is mediated by the single bi-functional zinc finger transcription factor (TF) Cubitus interruptus (Ci) (Abbasi et al., 2009; Méthot & Basler, 2001). In the absence of Hh signaling, Ci undergoes proteolytic cleavage and is converted to a repressor, while in the presence of Hh, Ci is modified to promote activator functions, as cleavage is inhibited (Aza-Blanc & Kornberg, 1999). In vertebrates, gene duplication events resulted in three copies of *Ci*-like ancestral Hh pathway effector genes: *Gli1*, *Gli2* and *Gli3* (Abbasi et al., 2009; Niewiadomski et al., 2019). Gli2 and Gli3 proteins have retained their potential for dual function, while Gli1 exclusively acts as transcriptional activator and its activity is mostly dispensable during development (Coy et al., 2011; Huangfu & Anderson, 2006).

Gli family members are known to be expressed in a wide variety of vertebrate tissues and cell types (Abbasi et al., 2007; Ruiz i Altaba, 2008). Studies in vertebrate model and non-model animals have demonstrated that highly orchestrated spatial and temporal expression of *Gli* genes is critical for the progression of embryonic development (Ruiz i Altaba, 2008). Among *Gli* paralogs, *Gli3* shows a particularly dynamic spatiotemporal expression pattern during development and is known as a key player in the manifestation of early embryological processes (Tyurina et al., 2005). Genetic studies have shown that *Gli3* acts as the main repressor of the *Shh* pathway, while *Gli1*/*Gli2* predominantly execute activator roles (Hu et al., 2006). For instance, $Gli3^{xt}$ mice, lacking a functional copy of *Gli3*, exhibit polydactyly and dorsal central nervous system defects associated with ectopic *Shh* expression (Haddad-Tóvolli et al., 2015; Hui & Joyner, 1993). In addition, *Gli3* null individuals manifest multiple defects affecting the CNS, musculo-skeletal system, craniofacial features, lungs and other internal organs, resulting in embryonic lethality prior to birth (Hui & Joyner, 1993; Lopez-Rios et al., 2012). Mutations in the human *GLI3* gene are associated with various related developmental disorders which are summarized as "*GLI3* morphopathologies", including Greig cephalopolysyndactyly (GCPS) (Vortkamp et

al., 1992), Pallister Hall syndrome (PHS) (Böse et al., 2002), preaxial polydactyly type IV (PPD-IV) and postaxial polydactyly type A (PAP A/B) (Memi et al., 2018). In addition to these disorders, and related to the role of *Gli3* in regulation of proliferation, mutations in *Gli3* can result in various types of tumors including glioblastoma, hypothalamic hamartoma, oral squamous cell carcinoma, colon cancer, gastric cancer and pancreatic cancer (Matissek & Elsawa, 2020; Rodrigues et al., 2018).

Transcriptional enhancers are a major class of *cis*-acting elements in metazoan genomes and harbor TF motifs to integrate signaling cues for orchestration of target gene expression (Long, Prescott, & Wysocka, 2016; Zehra & Abbasi, 2018). Genome-wide association studies and whole genome sequencing have revealed that non-coding variants associated with human disease phenotypes often map to enhancer sequences (Perenthaler et al.,2019). Developmental regulator genes such as *Gli3* are frequently embedded in gene-poor topologically associated domains (TADs) known to ensure accurate spatiotemporal regulation through confinement of long range enhancer-promoter interactions (Robson et al., 2019; Schoenfelder & Fraser, 2019). Individual enhancers can have critical developmental functions that can be affected by loss- or gain-of-function mutations. For example, while the ZRS (zone of polarizing activity regulatory sequence) is essentially required for limb-specific *Shh* expression and distal limb morphogenesis, point mutations in this element results in ectopic *Shh* expression leading to polydactyly (Kvon et al., 2020; Lettice et al., 2003). Often however, enhancer landscapes of key developmental regulator genes exhibit a regulatory architecture hallmarked by the presence of multiple enhancers with overlapping activities to confer transcriptional robustness, and only combinatorial deletion of such enhancers leads to discernible phenotypes, as shown for a pair of *Gli3* limb enhancers in mouse embryos (Kvon et al., 2021; Osterwalder et al., 2018). Also, it has been shown recently that development of digit 1 (thumb) is dependent on negative regulation of Gli3 repressor (Gli3R) activity in the anterior limb mesoderm through direct binding of Hoxa13 and Hoxd13 TFs with multiple *Gli3*-associated enhancer sequences (Bastida et al., 2020).

Transgenic reporter assays have been widely used as a method of choice to define tissue-specific transcriptional enhancer activities *in vivo* (Kvon, 2015) and multiple enhancer elements in the genomic vicinity of *Gli3* have been identified in a number of species (Anwar et al., 2015; Coy, et al., 2011; Hussain et al., 2021; Osterwalder et al., 2018; Mannion et al., 2022). In accordance with hotspots of *Gli3* expression in vertebrate embryos, these enhancer activities were found enriched in the CNS, limb/fins, craniofacial regions, and internal organs (Abbasi et al., 2010, 2013; Anwar et al., 2015; Coy et al., 2011; Hussain et al., 2021; Osterwalder et al., 2018; Osterwalder et al., 2014). However, the mammalian *Gli3*-associated enhancer activities identified so far are not associated with the full spectrum of *Gli3* expression domains. This observation led us to speculate that additional *cis*-acting elements might be in place to regulate spatiotemporal *Gli3* expression. Previous attempts for *Gli3* enhancer discovery made use of fugu and zebrafish genomes (as teleost representatives) to define evolutionary conserved enhancers in intronic and intergenic intervals of the human *GLI3* gene (Minhas et al., 2019; Parker., 2011). It has been reported that a large subset of mammalian tissue-specific enhancers might not be detected by direct mammalian-zebrafish/fugu comparisons, presumably due to large-scale duplications and rapid evolution of these representative teleost genomes (Braasch et al., 2016). Here, we

used a comparative genomics approach for *Gli3* enhancer discovery, including the genomes of fish species exhibiting a slow rate of molecular evolution, such as the spotted gar (*Lepisosteus oculatus*), elephant shark (*Callorhinchus milli*), and coelacanth (*Latimeria chalumanae*) (Ali et al., 2021; Amemiya et al., 2013; Braasch et al., 2016; Venkatesh et al., 2006). Intriguingly, human centric alignments have identified a set of novel CNEs that were not detected by direct human fugu/zebrafish comparisons. Novel CNEs predicted by this comparative analysis were then subjected to functional analysis by employing transgenic reporter assays in zebrafish and mouse embryos. Indeed, the majority of the elements tested acted as transcriptional enhancers driving expression in known compartments of *Gli3* expression in zebrafish and mouse embryos, including the neural tube and pectoral appendages/limbs. Together, our results demonstrate that efficient comparative sequence analysis of mammalian and fish (with slow rate of molecular evolution) genomes represent an effective approach for the discovery of developmental enhancers implicated in vertebrate morphogenesis and organ formation.

Results

Comparative analysis of *Gli3* loci in slowly evolving fish species uncovers deeply conserved intronic elements

Our previous studies have identified 14 *Gli3* intergenic CNE enhancers (CNE1-14) through human-fugu comparative genomics and transgenic zebrafish/mouse assays (Abbasi et al., 2010; Abbasi et al., 2007; Anwar et al., 2015; Coy et al., 2011; Paparidis et al., 2007). In order to search for previously undetected, deeply conserved enhancers, we performed a modified comparative analysis including the human *GLI3* locus and the “slowly-evolving” fish genomes; coelacanth, spotted gar and elephant shark (Amemiya et al., 2013; Braasch et al., 2016; Nikaido et al., 2013; Venkatesh et al., 2014). Multi-species sequence alignment of the human *GLI3* locus with its orthologous counter parts from mouse, chicken, lizard, coelacanth, fugu, spotted gar and elephant shark confirmed the previously identified CNE1-14 elements, and also identified seven novel CNEs (CNE15-21) that were not detected in direct tetrapod-teleost comparisons (Fig. 1A, S1; Tables 1, S1) (Anwar et al., 2015). Remarkably, all of these conserved noncoding elements (CNEs) are located within the intronic regions of *Gli3*, indicating a potential regulatory association with the *Gli3* promoter (Fig. 1). To investigate this, we visualized chromatin topology and TAD structure at the *Gli3* locus using available Hi-C datasets from mouse embryonic stem cells (mESCs), neural progenitor cells (NPCs) and cortical neurons (CNs) (Bonev et al., 2017) (Fig. S1A). This exploration revealed that while the overall *Gli3* TAD structure is maintained during neuronal differentiation, the *Gli3* gene body itself is embedded in a subdomain with increased interactions during the NPC stage (Fig 1B, Fig. S1A). Thus, these results indicate stronger preference for *Gli3* promoter-enhancer interactions in intronic regions. In summary, integration of our new analysis with previously reported data concludes that the *Gli3* locus harbors in total 21 human-fish CNEs based on selection criteria of 50% sequence identity over at least 50 bp sequence length (Fig. 1A; Table 1). Hereby, except for CNE20 (mm1826) the newly identified enhancers do not overlap previously tested *Gli3*-associated enhancer elements listed in the Vista enhancer database (https://enhancer.lbl.gov) (Fig. S1B).

Additionally, phylogenetic foot-printing analysis in these *Gli3*-associated CNEs revealed conserved binding motifs for several developmental transcription factors (TFs) (Table 1), corroborating the possible functional relevance of these newly identified CNEs. Furthermore, many of the TFs associated with these conserved binding motifs are known to be co-expressed with *Gli3* in various tissues during embryonic development, including limbs, heart, and the central nervous system (CNS). Some examples of these TFs include Pbx1, Foxp3, Foxo1, Nkx2, Oct1, and Hoxa7 (Table 1) (Baldarelli et al., 2021). In addition, CNE15, 19 and 20 also showed general histone modification (H3K27ac) signatures of developmental enhancers (Abascal et al., 2020) (Fig. 1B). Therefore, we concluded that these newly identified CNEs have the potential to act as *Gli3* enhancers.

A majority of intronic CNEs exhibit *Gli3*-associated enhancer activity in zebrafish

To interrogate the *in vivo* enhancer potential of the identified CNEs we performed transgenic reporter assays in zebrafish, a well-established model for the study of human enhancer function (Parisi et al., 2021). To this purpose, corresponding human sequences of the identified CNEs were cloned into a Tol2 reporter vector containing a *cfos* minimal promoter and EGFP cassette (Fisher et al., 2006) (Table S2). Following injection of constructs, zebrafish embryos were monitored for GFP expression at 24 hpf and 48 hpf. We first focused on the elements with highest sequence identity between human and elephant shark (CNE18, CNE19, CNE21) (Fig. 2A). While *Gli3*-CNE18 displays 80% sequence identity with elephant shark over a stretch of 370 bp, CNE19 and CNE21 share 71 % and 79% sequence identity, respectively (Fig. 1A; Table 1). Remarkably, CNE18, CNE19 and CNE21 each drove GFP reporter expression in multiple tissues. Each of these elements drove reporter gene expression in the developing pectoral fins at 48 hpf (Figs. 2A, S2, S3), in 60%, 62% and 53% of transgenic zebrafish embryos, respectively (Table S3). In comparison, the previously characterized and evolutionarily less conserved mm1179 mouse *Gli3* limb enhancer element located 120kb upstream of the *Gli3* transcription start site (TSS) (Osterwalder et al., 2014; Osterwalder et al., 2018) induced reporter gene expression mereley in the anterior-most margin of the developing zebrafish pectoral fin at 48 hpf (Figs. 2B, S2, S3). Apart from driving enhancer activity in the pectoral fin, CNE18 also showed activity in the forebrain at 48 hpf (Figs. 2A, S2) in 56% of transgenic zebrafish embryos (Table S3). CNE19 also triggered GFP expression in the developing forebrain at 48 hpf (Figs. 2A, S2). This forebrain enhancer activity was observed in 64% of CNE19 transgenic embryos (Table S3). In addition to activity in fins and CNS domains, CNE18 and CNE21 also drove GFP expression in the developing heart at 48 hpf (Figs. 2A, S2). While CNE16 and CNE20 did not reveal any reproducible enhancer activities in zebrafish embryos, CNE17 drove GFP reporter expression exclusively in forebrain (52%) at 48 hpf (Figs. 2C, S2; Table S3). As another unique reproducible activity detected among the newly identified subset of CNEs, CNE15 drove activity in craniofacial region (Figs. 3A, B, S2; Table S3). Taken together, these results assign the majority of the newly identified and deeply conserved CNEs to additional critical domains of *Gli3* transcription, with predominant activities in limbs and the CNS. A subset of these CNEs also shows overlap in their spatial domains, likely establishing or contributing to transcriptional robustness of *Gli3* expression in the respective tissues.

Association of a *GLI3* non-coding disease variant with CNE function

Single nucleotide polymorphisms (SNPs) in enhancers are considered important evolutionary drivers of disease and variable human phenotypes (Long et al., 2016; Kvon et al., 2020). Approximately 90% of SNPs with phenotypic associations in GWAS studies are mapped to non-coding regions (Huang & Ovcharenko, 2015), which suggests a major role in gene regulation. To explore a direct relationship of our identified CNE regions with human disease mechanisms, we performed intersection of our CNEs with a collection of human disease-associated SNPs based on a genome-wide association scan (Adhikari et al., 2016). This analysis identified a single SNP (rs17640804) that localized to the core of the CNE15 enhancer and that previously was associated with human nasal morphology specifically with nose wing breadth (Adhikari et al., 2016) (Fig 1; Table 1). Conservation analysis revealed that CNE15 shared 73% sequence homology with elephant shark over a span of 110 bp (Fig. 1; Table 1). Remarkably, this observation coincided well with our functional validation of CNE15 enhancer activity in transgenic zebrafish which revealed GFP reporter expression in the nasal pit (np) at 48 hpf in 68% transgenic embryos (Figs. 3A, B; S3). To investigate conservation of CNE15 craniofacial activity in mammals, we also performed transgenic CNE15-LacZ reporter assays in mouse embryos. These results corroborated the craniofacial activity of the CNE15 enhancer as reproducible signal was observed in mouse embryos at day 11.5 in the facial prominence at the boundary region to the forebrain hemispheres (Figs. 3C, S2). In summary, our findings indicate that *Gli3* intronic CNEs can co-localize with disease-associated SNPs, potentially disturbing enhancer function in a tissue-specific context.

CNEs as an evolutionary platform for modulation of tissue-specific enhancer activity

Given the deep sequence conservation of CNEs, in last step we focused on exploring the involvement of CNEs in evolutionary diversification. Hereby, we selected CNE14 (Anwar et al., 2015) as this represents a CNE with well-defined activity and function in the mammalian limb. CNE14 corresponds to the hs1586 enhancer (Vista Enhancer Browser, https://enhancer.lbl.gov) known to regulate *Gli3* expression in a partially redundant manner with mm1179 (Anwar et al., 2015; Osterwalder et al., 2018). While hs1586 drives expression in the *Gli3* overlapping domain in the anterior limb mesenchyme of mouse embryos, CNE14 was shown to be active in the pectoral fin of zebrafish (Abbasi et al., 2007; Anwar et al., 2015; Osterwalder et al., 2018) (Fig. S2). Unlike the previous study by Anwar et al., 2015, which reported conservation of CNE14/hs1586 down to the lizard species, the present study uncovers deeper conservation of the CNE14/hs1586 core sequence, exhibiting 73% sequence identity over a span of 230 base pairs (bps) when compared to coelacanth (lobe finned fish) and 66.9% over 127 bps when compared to elephant shark (cartilaginous fish) (Fig. 1). To elucidate the functional requirement of the deeply conserved CNE14/hs1586 core for fin- and limb-specific *Gli3* expression, we subjected the full-length enhancer sequence to reduction analysis and compared the activity of the respective elements in zebrafish and mouse embryos (Fig. 4). To this purpose, we considered the evolutionary conservation depth of core-flanking regions and validated three different extensions of the CNE14/hs1586 enhancer using transgenic reporter assays in mouse and zebrafish (Fig. 4A-C; Tables 1). Hereby, the CNE14-full-length (CNE14-L) element corresponded to the human-mouse conserved sequence, the CNE14-intermediate

version (CNE14-M) represented the human-chicken conserved element and the CNE14-core (CNE14-C) was restricted to the sequence block conserved in human and coelacanth (Fig. 4A). While CNE14-L (hs1586) was previously shown to be active in the anterior limb mesenchyme of mouse embryos at E11.5 (Fig. 4C; Vista Enhancer Browser) (Osterwalder et al., 2018), transgenic analysis in zebrafish embryos revealed reporter expression in the pectoral fin at 48 hpf (n=58% of transgenic embryos) (Figs. 4B, S2). The human-chicken CNE14-M conserved block (958 bp) was found to retain activity in the pectoral zebrafish fin (n=48% of transgenic zebrafish embryos), however, the anterior portion of developing mouse forelimbs exhibited reduced LacZ staining at E11.5 (n=4/6) when compared to CNE14-L transgenics (Fig. 4B, C; Table S2). Remarkably, the 304 bp human-coelacanth element (CNE14-C) was largely insufficient to drive reproducible reporter gene expression in both developing zebrafish fins or anterior mouse limbs (Fig. 4B, C; Table S2). Together these results indicate that the hs1586/CNE14 genomic element acquired limb-specificity after the emergence of lobe-finned fish and later in mammals gained robustness via anterior forelimb activity.

Discussion

Over the last couple of decades comparative genomic approaches have revealed numerous evolutionarily conserved *cis*-acting regulatory modules in vertebrates (CRMs) (Ali et al., 2016; Minhas et al., 2015; Paparidis et al., 2007). CNEs are enriched for CRMs with developmental enhancer activities (Boffelli et al., 2004; Pennacchio et al., 2006; Woolfe et al., 2005) and vertebrate genomes hold thousands of anciently conserved CNEs which frequently cluster near trans-developmental genes (Parveen et al., 2013; Woolfe et al., 2005). In particular, multi-species comparative analysis of tetrapod-teleost orthologous loci have identified a plethora of conserved CRMs in genomic domains of developmental genes including *Shh*, *Gli2*, *Gli3*, *Sim1*, *Sox* genes, *Hoxa* and *Hoxd* clusters (Abbasi et al., 2010; Ali et al., 2021; Gehrke et al., 2015; Kim et al., 2014). Human and teleost lineages diverged approximately 450 million years ago (Mya) and it was assumed that CNEs conserved between them reflect strong selection pressure, most likely due to critical function (Elgar et al., 1996). Indeed, genome-wide comparisons of humans with fugu and zebrafish have identified hundreds of CNEs with suggested functional roles as tissue-specific enhancers (Gehrke et al., 2015; Paparidis et al., 2007; Woolfe et al., 2005). Therefore, teleost genomes and genomes of fish with slow rates of molecular evolution are being considered instrumental and ideal models for the identification of evolutionary conserved transcriptional enhancers of human developmental genes (Christoffels et al., 2004; Woolfe & Elgar, 2007).

Gli3 is one of the key transcriptional effectors of the hedgehog signaling pathway (Hui & Angers, 2011) and genetic analyses in mice, chicken or zebrafish revealed an essential requirement of Gli3 for the development of the CNS, limbs/fins and various other organs during embryogenesis (Hui & Angers, 2011). Therefore, conserved CRMs located within intronic intervals of *Gli3* are expected to contribute to these processes by orchestrating accurate spatiotemporal *Gli3* expression. In our previous studies, comparative analysis of the human *GLI3* locus with orthologous counterparts from teleost fish (e.g. fugu and zebrafish) has identified 14 intronic CNEs (CNE1-14) acting as tissue-specific developmental enhancers in transgenic mouse/zebrafish assays (Schmidt et al., 2013; Abbasi

et al., 2010; Abbasi et al.,2009; Ali et al., 2021; Anwar et al., 2015; Coy et al., 2011; Paparidis et al., 2007; Tanaka, 2016). Given that a large proportion of regulatory elements arose in the common ancestors of jawed vertebrates and have been either lost or have diverged beyond recognition in teleost fish, relevant ancestral CNEs might have been missed by using human-teleost comparisons. Therefore, in the current study we re-evaluated the *Gli3* CNE landscapes by including not only tetrapod (human, mouse, chicken, and lizard) and teleost genomes, but also genomic sequences from fish species with a slower rate of molecular evolution, such as spotted gar and elephant shark (Amemiya et al., 2013; Braasch et al., 2016). The novel elements identified by this method (CNE15-21) share deep homology with elephant shark, except CNE17 which is conserved only in spotted gar. As the majority of the newly identified human CNEs (n=5/7) drove tissue-specific transgenic reporter expression in known *Gli3* expressing tissues in zebrafish embryos, our findings demonstrate the significance of slowly evolving fish genomes for the identification of ancestral CNEs exhibiting mammalian enhancer activity.

Interestingly, all the identified CNEs in this study are located within intronic regions of *Gli3*. A BLAST-algorithm based similarity search of the shortlisted CNEs confirmed their exclusive presence within the human *GLI3* locus. Accordingly, our analysis revealed the presence of binding motifs for distinct TFs with well-established roles in embryonic development and that are known to be co-expressed with *Gli3* (Cobb & Duboule, 2005; Reymond et al., 2002; Visel et al., 2004). *Gli3* exerts multiple essential roles during limb morphogenesis and is a main determinant of anterior-posterior axis specification in early limb/fin buds (Galli et al., 2010; Osterwalder et al., 2015). Tight spatio-temporal control of *Gli3* is required for early restriction of transcriptional regulators and induction of the zone of polarizing activity, while later *Gli3* controls cell cycle regulators and the exit of proliferating progenitors in the anterior limb mesenchyme, preventing over-proliferation that leads to polydactyly (Lopez-Rios et al., 2012). This transgenic analysis revealed that the majority of newly identified CNEs (CNE18, CNE19, CNE21) drive reporter expression in multiple *Gli3*-expressing tissues such as the developing pectoral zebrafish fin, thereby increasing the number of presently known elements with *Gli3*-associated enhancer activities in paired appendages of vertebrates to eight (Abbasi et al., 2010; Anwar et al., 2015; Osterwalder et al., 2018). Recent assessment of the enhancer landscape within the *Gli3* TAD based on a 5kb-tiling approach of genomic elements also uncovered additional putative limb enhancer segments in regions upstream of the *Gli3* gene body (Mannion et al., 2022). This diversity, in addition with the newly discovered intronic CNEs with limb activity in zebrafish, suggests an increasing number of *Gli3* enhancers, potentially with redundant activities and providing an explanation for the lack of drastic *Gli3* transcript reduction in mouse limb buds following the combined deletion of mm1179 and hs1586 (Osterwalder et al., 2018). However, more work is needed to uncover the functional relationships of these enhancers and how they regulate early or later roles of *Gli3* during limb development. *Gli3* also plays an important role in the development and morphogenesis of vertebrate CNS (Matissek & Elsawa, 2020). For instance, *Gli3* participates in the development of various CNS domains including dentate gyrus, hippocampus, telencephalon, di-cephalon, cerebral cortex, neocortex, corpus callosum and others (Amaniti et al., 2013; Kuschel et al., 2003; Matissek & Elsawa, 2020; Theil et al., 1999). Concordantly, mutations in the human

GLI3 gene are associated with macrocephaly, macroencephaly and hypothalamic hamartoma (Craig et al., 2008; Matissek & Elsawa, 2020). In addition, the dorsal telencephalon in *Gli3* mouse mutants fails to develop normally (Craig et al., 2008; Kuschel et al., 2003; Matissek & Elsawa, 2020). Endogenous *Gli3* expression dynamics during early CNS patterning and morphogenesis reveals the genes' highly conserved function (Aoto et al., 2002; Baldarelli et al., 2021; Tyurina et al., 2005). Three of the CNEs identified in this study drove reporter expression in forebrain (CNE17, CNE18, and CNE19), indicating involvement in early CNS regulation. Adding to our previously published studies, these results expand the number of *in vivo* validated *Gli3* brain enhancers to twelve (Abbasi et al., 2010; Hussain et al., 2021) indicating that a complex and conserved *Gli3* regulatory architecture is in place to drive *Gli3* expression in brain subregions of mammals and fish (Hussain et al., 2021). *Gli3* also acts as a mediator of Shh pathway activity in posterior second heart field cells during cardiac development, and GLI3-A has been recently identified as an important component for intraventricular septum (IVS) formation through cilia-mediated PDGFR alpha signaling during mouse heart development (Wiegering et al., 2020). While CNE18 and CNE21 due to their heart-specific activity in zebrafish appear as candidates for the regulation of such cardiac *Gli3* functions, analysis of these elements in mice will be required to establish a role in the mammalian heart. Our transgenic reporter analysis in zebrafish embryos did not reveal reproducible tissue-specific activities for CNE16 and CNE20 (Figure 1; Table 1). It is however possible that these elements are active at earlier or later embryonic stages. Another potential explanation for their lack of activity could be the absence of specific trans-acting factors in zebrafish (Gehrke & Shubin, 2016). However, without further evaluating their functions in a native trans-environment, no definite conclusion can be made about the regulatory potential of these two CNEs in mammals (Gehrke & Shubin, 2016). Furthermore, CNEs might also function as transcriptional repressors, which is not evaluated in the applied transgenic reporter framework (Anwar et al., 2015; Segert et al, 2021).

Over the last few years, it has become increasingly clear that the majority of disease-associated mutations reside within non-coding enhancers regions (Kvon et al., 2020; Moyon et al., 2022). For instance, based on GWAS data, a SNP rs17640804 (C>T) in the *GLI3* non-coding region (7q13, intron-3) has been associated with human face morphology, more specifically, with determining the nose wing breadth (Adhikari et al., 2016). Here we showed that rs17640804 (C>T) is positioned within the CNE15 core located in intron 3 of *GLI3* (Fig. 1A). In line with the GWAS results, our comparative analysis in mouse and zebrafish embryos revealed conserved enhancer activity of the human CNE15 element in the nasal prominence (Fig. 3). These findings illustrate that intersection of genetic variants and functionally validated *GLI3* enhancers can provide insight into the mechanistic basis of *GLI3* morphopathies which include GCPS (MIM175700) (Kalff-Suske et al., 1999; Wild et al., 1997) PHS (MIM146510) (Kang et al, 1997) PPD-IV (MIM 174700) (Radhakrishna et al., 1999) and PAP A/B (MIM174200) (Radhakrishna et al, 1997).

A key aspect in evolutionary developmental biology is to understand how the distinct features of different organisms evolved at the genomic level. For instance, among the vertebrates the genetic basis of fin-to-limb transition has been a central focus of research in studies focused on evolutionary aspects (Clack, 2009; Abbasi, 2011). Of particular importance to evolution of morphological traits such as limbs are enhancers that recruit

combinations of TFs to short binding sequence motifs that collectively determine when, where, and how genes are transcribed during development, thereby defining the physical properties of distinct cell populations (Yousaf et al., 2015). These properties render enhancers a fundamental tool in fine-tuning the evolution of traits, for example in developing limbs (Kvon et al., 2016). However, despite the well-established roles of enhancers in evolution, many questions remain unresolved regarding the overall sequence makeup and transcriptional output of individual enhancers. Here we focused on the coelacanth-conserved hs1586/CNE14 enhancer shown to contribute to robust spatiotemporal expression of *Gli3* in the developing mouse limb (Osterwalder et al., 2018). While there is no detectable sequence conservation in teleost species (fugu/zebrafish), we considered hs1586 as an ideal candidate to address the evolutionary history of the enhancer core element (Fig. 4) (Abbasi et al., 2007; Osterwalder et al., 2018). Our comparative transgenic *in vivo* analyses in mouse and zebrafish support the hypothesis that robust anterior limb activity of hs1586/CNE14-L in tetrapods has evolved after divergence from lobe-finned fish (Fig. 4). In accordance with limb autopod diversification, our results indicate that forelimb-specific activity is likely encoded in the core-flanking regions of hs1586/CNE14-L which from a mammalian perspective are less conserved in chicken (CNE14-M) and nearly absent in coelacanth (CNE14-C) (Fig. 4). In this regard, it can be speculated that limb activity of hs1586/CNE14-L has been progressively acquired since the divergence of amniotes and lobe-finned fish. Such an increase in regulatory *Gli3* activity in the anterior mesenchyme of mammalian embryonic limbs might have contributed to stabilization of Gli3R levels in order to counteract the Shh agonist and to restrict the limb to pentadactyly (Galli et al., 2010; Lopez-Rios et al., 2012). Such an effect might have arisen through progressive gain of regulatory motifs in the enhancer throughout evolutionary diversification (Cotney et al., 2013; Rebeiz & Tsiantis, 2017).

Overall, our results suggest that many of the CNEs identified in the *Gli3* locus are involved in regulating tissue-specific aspects of *Gli3* expression and can serve as a resource to investigate vertebrate diversity and morphological variability in human population.

Materials and Methods

Multispecies comparative sequence analysis

Human *GLI3* sequence together with its orthologous counterparts from mouse, chicken, lizard, coelacanth, fugu, spotted gar and elephant shark were obtained from ensemble genome browser (https://asia.ensembl.org). These orthologous sequences were submitted to shuffle lagan tool for comparative analysis (Brudno et al., 2003). Human sequence was used as baseline. The alignment results were visualized by using the vista visualization tool (Brudno et al., 2003). The selection criteria to identify anciently conserved fish-tetrapod CNEs included 50% sequence identity over at least 50 bp sequence length.

The orthologous CNEs sequences were subjected to the MEME (Multiple Em for Motif Elicitation) tool for the identification of conserved transcription factor binding sites (TFBSs) (Bailey et al., 2006). The MEME tool is based on position weight metrices algorithm which is used to scan the orthologous sequences for over-represented motifs. The conserved transcription factor binding motifs identified by MEME tool were investigated for binding

preferences for known TFs by STAMP tool using TRANSFAC library (Mahony & Benos, 2007).

Hi-C 3D chromatin mapping

Hi-C maps of the extended *Gli3* TAD in mESCs, NPCs and CNs were generated from valid read pairs (MAPQ30) based on the datasets published by (Bonev et al., 2017) (GSE96107) and available from a third-party re-analysis using HiCUP v.0.6.1 (GSE161259). The following code was utilized for the representation of the HiC maps: https://github.com/lldelisle/Hi-C_reanalysis_Bonev_2017. Raw .cool format Hi-C maps were produced using the 'cooler cload tabix' tool (Cooler v0.8.11), loading validated read pairs of a *Gli3*-containing native chromosome 13 genomic interval into a matrix of fixed bins (5kb resolution). Further normalization and diagonal filtering included us of the Cooler matrix balancing tool (Abdennur & Mirny, 2020) with the options '--mad-max 5 --min-nnz 10 --min-count 0 --ignore-diags 2 --tol 1e-05 --max-iters 200 --cis-only' which resulted in balanced 'cool' maps as final output. Heatmap matrices shown in Figs. 1B and S1A were aligned with other genomic datasets and plotted using pygenometracks (Lopez-Delisle et al., 2021).

Zebrafish transgenic enhancer assay

Genomic DNA from human whole blood was extracted by using DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer instructions. The selected CNE elements were amplified from human genomic DNA with primers listed in Table S2. PCR products (~500ng/µl) were ligated with the PCR8/GW2/TOPO TA cloning vector system to generate entry clones according to manufacturer instructions (Invitrogen, Life Technologies). Insert identity was confirmed by sequencing. CNE fragments were moved to destination vector pGW-*cfos*-*EGFP* via gateway cloning technology (Fisher et al., 2006). The LR (attL and attR) recombination reaction between entry and destination vectors (150ng/µl each) was performed using the LR clonase enzyme. The resulting destination vectors were sequence verified by Sanger sequencing.

Zebrafish were bred and raised according to the standard protocol. Fertilized eggs were collected from natural spawning of the wild type zebrafish. Transposase encoding mRNA was synthesized by *in vitro* transcription from linearized pCS-TP plasmid using the Sp6 mMessage mMachine kit (Ambion). The mRNA was precipitated in 100 % molecular grade ethanol and lithium chloride, followed by phenol-chloroform purification. A protocol devised by Fisher et al. 2006 was used for the preparation of injection solution containing 1µl reporter vector pGW-CNE-*cfos*-EGFP (125ng/µl), 0.5µl transposase mRNA (300ng/µl) and 0.5µl phenol red, and adjusted to 5µl by adding molecular grade water (Fisher et al., 2006). Approximately 2nl of injection solution were injected into cytoplasm of fertilized embryos at one-to-two cell stages. The embryos were kept at 28.5 °C in 1X E3 media containing 0.003% phenylthiourea.

The injected embryos were raised in E3 media. At 24 hpf embryos were dechorionated manually and anesthetized in tricaine. We relied on mosaic transgenesis and the F0 embryos were screened for reporter gene expression (*GFP*) using a fluorescent inverted microscope

IX71 (Olympus, Japan). Live zebrafish embryos were imaged using a DP72 camera with monochrome software.

Mouse LacZ transgenic reporter analysis

Mouse experiments were performed, reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee (Pennacchio et al., 2006). All mice were monitored daily for food and water intake, and animals were inspected weekly. For enhancer-reporter analysis in transgenic mouse embryos, CNE versions were inserted into the pHsp68-lacZ vector (#170102) using Gibson cloning as described (Kotharym et al., 1989; Osterwalder et al., 2022). The reporter vector was linearized using NotI, microinjected into fertilized mouse oocytes and implanted into pseudopregnant mouse females (Osterwalder et al., 2022). Transgenic embryos were collected at e11.5 and stained with X-gal to visualize LacZ reporter activity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are thankful to Higher Education Commission of Pakistan for NRPU grant (No. 20-2085/NRPU/R&D/HEC/12/760). MO was funded by Swiss National Science Foundation (SNSF) grant PCEFP3_186993. A.V. was supported by U.S. National Institutes of Health (NIH) grants R01DE028599, UM1HG009421, and R01HG003988. We thank the research staff of the transgenic core in the Visel/Pennacchio group at LBNL for pro-nuclear injection of reporter constructs and mouse founder embryo generation. Research conducted at the E.O. Lawrence Berkeley National Laboratory was performed under U.S. Department of Energy Contract DE-AC02-05CH11231, University of California (UC).

Sponsors:

This work was funded by Higher Education Commission of Pakistan grant (No. 20-2085/NRPU/RandD/HEC/12/760) to A.A.A. MO was supported by Swiss National Science Foundation (SNSF) grant PCEFP3_186993. S. A holds a fellowship under International Research Support Initiative Program (IRSIP) of Higher Education Commission of Pakistan.

References

Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B, … Zimmerman J (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature, 583(7818), 699–710. 10.1038/s41586-020-2493-4 [PubMed: 32728249]

Abbasi Amir A., Minhas R, Schmidt A, Koch S & Grzeschik KH (2013). Cis-regulatory underpinnings of human GLI3 expression in embryonic craniofacial structures and internal organs. Development Growth and Differentiation, 55(8), 699–709. 10.1111/dgd.12076

Abbasi Amir A., Paparidis Z, Malik S, Bangs F, Schmidt A, Koch S, … Grzeschik KH (2010). Human intronic enhancers control distinct sub-domains of Gli3 expression during mouse CNS and limb development. BMC Developmental Biology, 10. 10.1186/1471-213X-10-44

Abbasi Amir Ali. (2011). Evolution of vertebrate appendicular structures: Insight from genetic and palaeontological data. Developmental Dynamics, 240(5), 1005–1016. 10.1002/dvdy.22572 [PubMed: 21337665]

Abbasi Amir Ali, Goode DK, Amir S & Grzeschik KH (2009). Evolution and functional diversification of the GLI family of transcription factors in vertebrates. Evolutionary Bioinformatics, 2009(5), 5–13. 10.4137/EBO.S2322

Abbasi Amir Ali, Paparidis Z, Malik S, Goode DK, Callaway H, Elgar G& Grzeschik KH (2007). Human GL13 intragenic conserved non-coding sequences are tissue-specific enhancers. PLoS ONE, 2(4). 10.1371/journal.pone.0000366

Abdennur N, & Mirny LA (2020). Cooler: Scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics, 36(1), 311–316. 10.1093/bioinformatics/btz540 [PubMed: 31290943]

Adhikari K, Fuentes-Guajardo M, Quinto-Sánchez M, Mendoza-Revilla J, Camilo Chacón-Duque J, Acuña-Alonzo V, … Ruiz-Linares A (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nature Communications, 7(May), 1–11. 10.1038/ncomms11616

Ali S, Amina B, Anwar S, Minhas R, Parveen N, Nawaz U, … Abbasi AA (2016). Genomic features of human limb specific enhancers. Genomics, 108(3–4), 143–150. 10.1016/j.ygeno.2016.08.003 [PubMed: 27580967]

Ali S, Arif I, Iqbal A, Abrar M, Abbasi AA, & Hussain I (2021). Comparative genomic analysis of human GLI2 locus using slowly evolving fish revealed the ancestral gnathostome set of early developmental enhancers. (December 2020), 669–683. 10.1002/dvdy.291

Amaniti EM, Hasenpusch-Theil K, Li Z, Magnani D, Kessaris N, Mason JO, & Theil T (2013). Gli3 is required in Emx1+ progenitors for the development of the corpus callosum. Developmental Biology, 376(2), 113–124. 10.1016/j.ydbio.2013.02.001 [PubMed: 23396189]

Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, MacCallum I, … Lindblad-Toh K (2013). The African coelacanth genome provides insights into tetrapod evolution. Nature, 496(7445), 311–316. 10.1038/nature12027 [PubMed: 23598338]

Anderson E, Peluso S, Lettice LA, & Hill RE (2012). Human limb abnormalities caused by disruption of hedgehog signaling. Trends in Genetics, 28(8), 364–373. 10.1016/j.tig.2012.03.012 [PubMed: 22534646]

Anwar S, Minhas R, Ali S, Lambert N, Kawakami Y, Elgar G, … Abbasi AA (2015). Identification and functional characterization of novel transcriptional enhancers involved in regulating human GLI3 expression during early development. Development Growth and Differentiation, 57(8), 570–580. 10.1111/dgd.12239

Aoto K, Nishimura T, Eto K, & Motoyama J (2002). Mouse GLI3 regulates Fgf8 expression and apoptosis in the developing neural tube, face, and limb bud. Developmental Biology, 251(2), 320–332. 10.1006/dbio.2002.0811 [PubMed: 12435361]

Aza-Blanc P, & Kornberg TB (1999). Ci: A complex transducer of the Hedgehog signal. Trends in Genetics, 15(11), 458–462. 10.1016/S0168-9525(99)01869-7 [PubMed: 10529809]

Bailey TL, Williams N, Misleh C, & Li WW (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research, 34(WEB. SERV. ISS.), 369–373. 10.1093/nar/gkl198

Baldarelli RM, Smith CM, Finger JH, Hayamizu TF, McCright IJ, Xu J, … Ringwald M (2021). The mouse Gene Expression Database (GXD): 2021 update. Nucleic Acids Research, 49(D1), D924–D931. 10.1093/nar/gkaa914 [PubMed: 33104772]

Bastida MF, Pérez-Gómez R, Trofka A, Zhu J, Rada-Iglesias A, Sheth R, … Ros MA (2020). The formation of the thumb requires direct modulation of Gli3 transcription by Hoxa13. Proceedings of the National Academy of Sciences of the United States of America, 117(2), 1090–1096. 10.1073/pnas.1919470117 [PubMed: 31896583]

Boffelli D, Nobrega MA, & Rubin EM (2004). Comparative genomics at the vertebrate extremes. Nature Reviews Genetics, 5(6), 456–465. 10.1038/nrg1350

Bonev B, Cohen NM, Szabo Q, Hugnot J, Tanay A, Cavalli G, … Lubling Y (2017). Multiscale 3D Genome Rewiring during Mouse Article Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell, 171(3), 557.e1–557.e24. 10.1016/j.cell.2017.09.043 [PubMed: 29053968]

Böse J, Grotewold L, & Rüther U (2002). Pallister-Hall syndrome phenotype in mice mutant for Gli3. Human Molecular Genetics, 11(9), 1129–1135. 10.1093/hmg/11.9.1129 [PubMed: 11978771]

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, … Postlethwait JH (2016). The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nature Genetics, 48(4), 427–437. 10.1038/ng.3526 [PubMed: 26950095]

Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, & Batzoglou S (2003a). Glocal alignment: Finding rearrangements during alignment. Bioinformatics, 19(SUPPL. 1). 10.1093/bioinformatics/btg1005

Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, & Batzoglou S (2003b). Glocal alignment: Finding rearrangements during alignment. Bioinformatics, 19(SUPPL. 1). 10.1093/bioinformatics/btg1005

Choudhry Z, Rikani AA, Choudhry AM, Tariq S, Zakaria F, Asghar MW, … Mobassarah NJ (2014). Sonic hedgehog signalling pathway: A complex network. Annals of Neurosciences, 21(1), 28–31. 10.5214/ans.0972.7531.210109 [PubMed: 25206052]

Christoffels A, Koh EGL, Chia JM, Brenner S, Aparicio S, & Venkatesh B (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Molecular Biology and Evolution, 21(6), 1146–1151. 10.1093/molbev/msh114 [PubMed: 15014147]

Clack JA (2009). The fin to limb transition: New data, interpretations, and hypotheses from paleontology and developmental biology. Annual Review of Earth and Planetary Sciences, 37, 163–179. 10.1146/ANNUREV.EARTH.36.031207.124146

Cobb J, & Duboule D (2005). Comparative analysis of genes downstream of the Hoxd cluster in developing digits and external genitalia. Development, 132(13), 3055–3067. 10.1242/DEV.01885 [PubMed: 15944189]

Cotney J, Leng J, Yin J, Reilly SK, Demare LE, Emera D, … Noonan JP (2013). XThe evolution of lineage-specific regulatory activities in the human embryonic limb. Cell, 154(1), 185. 10.1016/j.cell.2013.05.056 [PubMed: 23827682]

Coy S, Caamano JH, Carvajal J, Cleary ML, & Borycki A-G (2011). A Novel Gli3 Enhancer Controls the Gli3 Spatiotemporal Expression Pattern through a TALE Homeodomain Protein Binding Site. Molecular and Cellular Biology, 31(7), 1432–1443. 10.1128/mcb.00451-10 [PubMed: 21262763]

Craig DW, Itty A, Panganiban C, Szelinger S, Kruer MC, Sekar A, … Kerrigan JF (2008). Identification of Somatic Chromosomal Abnormalities in Hypothalamic Hamartoma Tissue at the GLI3 Locus. American Journal of Human Genetics, 82(2), 366–374. 10.1016/j.ajhg.2007.10.006 [PubMed: 18252217]

Elgar G, Sandford R, Aparicio S, Macrae A, Venkatesh B, & Brenner S (1996). Small is beautiful: Comparative genomics with the pufferfish (Fugu rubripes). Trends in Genetics, 12(4), 145–150. 10.1016/0168-9525(96)10018-4 [PubMed: 8901419]

Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, Kawakami K, & McCallion AS (2006). Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nature Protocols, 1(3), 1297–1305. 10.1038/nprot.2006.230 [PubMed: 17406414]

Galli A, Robay D, Osterwalder M, Bao X, Bénazet JD, Tariq M, … Zeller R (2010). Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development. PLoS Genetics, 6(4). 10.1371/journal.pgen.1000901

Gehrke AR, Schneider I, De Calle-mustienes E, Tena JJ& Gomez-marin C (2015). Deep conservation of wrist and digit enhancers in fish. 112(3), 803–808. 10.1073/pnas.1420208112

Gehrke AR, & Shubin NH (2016). Cis-regulatory programs in the development and evolution of vertebrate paired appendages. Seminars in Cell and Developmental Biology, 57, 31–39. 10.1016/j.semcdb.2016.01.015 [PubMed: 26783722]

Haddad-Tóvolli R, Paul FA, Zhang Y, Zhou X, Theil T, Puelles L, … Alvarez-Bolado G (2015). Differential requirements for Gli2 and Gli3 in the regional specification of the mouse hypothalamus. Frontiers in Neuroanatomy, 9(MAR), 1–17. 10.3389/fnana.2015.00034 [PubMed: 25657619]

Hu MC, Mo R, Bhella S, Wilson CW, Chuang PT, Hui CC, & Rosenblum ND (2006). GLI3-dependent transcriptional repression of Gli1, Gli2 and kidney patterning genes disrupts renal morphogenesis. Development, 133(3), 569–578. 10.1242/dev.02220 [PubMed: 16396903]

Huang D, & Ovcharenko I (2015). Identifying causal regulatory SNPs in ChIP-seq enhancers. Nucleic Acids Research, 43(1), 225–236. 10.1093/nar/gku1318 [PubMed: 25520196]

Huangfu D, & Anderson KV (2006). Signaling from Smo to Ci/Gli: Conservation and divergence of Hedgehog pathways from Drosophila to vertebrates. Development, 133(1), 3–14. 10.1242/dev.02169 [PubMed: 16339192]

Hui CC, & Angers S (2011). Gli proteins in development and disease. Annual Review of Cell and Developmental Biology, 27, 513–537. 10.1146/annurev-cellbio-092910-154048

Hui CC, & Joyner AL (1993). A mouse model of Greig cephalo–polysyndactyly syndrome: The extra–toesJ mutation contains an intragenic deletion of the Gli3 gene. Nature Genetics, 3(3), 241–246. 10.1038/ng0393-241 [PubMed: 8387379]

Hussain I, Raza RZ, Ali S, Abrar M, & Abbasi AA (2021). Molecular signatures of selection on the human GLI3 associated central nervous system specific enhancers. Development Genes and Evolution, 231(1–2), 21–32. 10.1007/s00427-021-00672-1 [PubMed: 33655411]

Jiang J, & Hui C. chung. (2008). Hedgehog Signaling in Development and Cancer. Developmental Cell, 15(6), 801–812. 10.1016/j.devcel.2008.11.010 [PubMed: 19081070]

Kalff-Suske M, Wild A, Topp J, Wessling M, Jacobsen EM, Bornholdt D, … Grzeschik KH (1999). Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. Human Molecular Genetics, 8(9), 1769–1777. 10.1093/hmg/8.9.1769 [PubMed: 10441342]

Kang S Jr, J. MG, Olney AH, & Biesecker LG (1997). GLI3 frameshif mutations cause autosomal dominant Pallister-Hall syndrome. 15(march 1997), 266–268.

Kim MJ, Oksenberg N, Hoffmann TJ, Vaisse C, & Ahituv N (2014). Functional characterization of SIM1 -associated enhancers. 23(7), 1700–1708. 10.1093/hmg/ddt559

Kotharym R, Clapoff S, Darling S, Perry MD, Moran LA, & Ant JR (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. In Development (Vol. 105).

Kuschel S, Rüther U, & Theil T (2003). A disrupted balance between Bmp/Wnt and Fgf signaling underlies the ventralization of the Gli3 mutant telencephalon. Developmental Biology, 260(2), 484–495. 10.1016/S0012-1606(03)00252-5 [PubMed: 12921747]

Kvon EZ (2015). Using transgenic reporter assays to functionally characterize enhancers in animals. Genomics, 106(3), 185–192. 10.1016/j.ygeno.2015.06.007 [PubMed: 26072435]

Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, … Visel A (2016). Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell, 167(3), 633–642.e11. 10.1016/j.cell.2016.09.028 [PubMed: 27768887]

Kvon EZ, Waymack R, Gad M, & Wunderlich Z (2021). Enhancer redundancy in development and disease. Nature Reviews. Genetics, 22(5), 324–336. 10.1038/S41576-020-00311-X

Kvon EZ, Zhu Y, Kelman G, Novak CS, Plajzer-Frick I, Kato M, … Pennacchio LA (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. Cell, 180(6), 1262–1271.e15. 10.1016/j.cell.2020.02.031 [PubMed: 32169219]

Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, … de Graaff E (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human Molecular Genetics, 12(14), 1725–1735. 10.1093/hmg/ddg180 [PubMed: 12837695]

Long HK, Prescott SL, & Wysocka J (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell, 167(5), 1170–1187. 10.1016/J.CELL.2016.09.018 [PubMed: 27863239]

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, … Manke T (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics, 37(3), 422–423. 10.1093/bioinformatics/btaa692 [PubMed: 32745185]

Lopez-Rios J, Speziale D, Robay D, Scotti M, Osterwalder M, Nusspaumer G, … Zeller R (2012). GLI3 constrains digit number by controlling both progenitor proliferation and BMP-dependent exit to chondrogenesis. Developmental Cell, 22(4), 837–848. 10.1016/J.DEVCEL.2012.01.006 [PubMed: 22465667]

Mahony S, & Benos PV (2007). STAMP: A web tool for exploring DNA-binding motif similarities. Nucleic Acids Research, 35(SUPPL.2), 253–258. 10.1093/nar/gkm272

Mannion BJ, Osterwalder M, Tran S, Plajzer-Frick I, Novak CS, Afzal V, … Pennacchio LA (2022). Uncovering Hidden Enhancers Through Unbiased In Vivo Testing. BioRxiv, 2022.05.29.493901.

Matissek SJ, & Elsawa SF (2020). GLI3: A mediator of genetic diseases, development and cancer. Cell Communication and Signaling, 18(1), 1–20. 10.1186/s12964-020-00540-x [PubMed: 31900175]

Memi F, Zecevic N, & Radonjić N (2018). Multiple roles of Sonic Hedgehog in the developing human cortex are suggested by its widespread distribution. Brain Structure and Function, 223(5), 2361–2375. 10.1007/s00429-018-1621-5 [PubMed: 29492654]

Méthot N, & Basler K (2001). An absolute requirement for Cubitus interruptus in Hedgehog signaling. Development, 128(5), 733–742. 10.1242/dev.128.5.733 [PubMed: 11171398]

Minhas R, Paterek A, Łapiński M, Bazała M, Korzh V, & Winata CL (2019). A novel conserved enhancer at zebrafish zic3 and zic6 loci drives neural expression. Developmental Dynamics, 248(9), 837–849. 10.1002/dvdy.69 [PubMed: 31194899]

Minhas R, Pauls S, Ali S, Doglio L, Khan MR, Elgar G, & Abbasi AA (2015). Cis-regulatory control of human GLI2 expression in the developing neural tube and limb bud. Developmental Dynamics, 244(5), 681–692. 10.1002/dvdy.24266 [PubMed: 25715918]

Moyon L, Berthelot C, Louis A, Nguyen NTT, & Crollius HR (2022). Classification of non-coding variants with high pathogenic impact. PLoS Genetics, 18(4), 1–20. 10.1371/journal.pgen.1010191

Niewiadomski P, Niedziółka SM, Markiewicz Ł, Uśpieński T, Baran B, & Chojnowska K (2019). Gli Proteins: Regulation in Development and Cancer. Cells, 8(2), 147. 10.3390/cells8020147 [PubMed: 30754706]

Nikaido M, Noguchi H, Nishihara H, Toyoda A, Suzuki Y, Kajitani R, … Okada N (2013). Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome Research, 23(10), 1740–1748. 10.1101/gr.158105.113 [PubMed: 23878157]

Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, … Pickle CS (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature, 554(7691), 239–243. [PubMed: 29420474]

Osterwalder M, Speziale D, Shoukry M, Mohan R, Ivanek R, Kohler M, … Zeller R (2014). HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. Developmental Cell, 31(3), 345–357. 10.1016/J.DEVCEL.2014.09.018 [PubMed: 25453830]

Osterwalder M, Speziale D, Shoukry M, Mohan R, Ivanek R, Kohler M, … Zeller R (2015). HHS Public Access. 31(3), 345–357. 10.1016/j.devcel.2014.09.018.HAND2

Osterwalder M, Tran S, Hunter RD, Meky EM, von Maydell K, Harrington AN, … Visel A (2022). Characterization of Mammalian In Vivo Enhancers Using Mouse Transgenesis and CRISPR Genome Editing. Methods in Molecular Biology (Clifton, N.J.), 2403, 147–186. 10.1007/978-1-0716-1847-9_11

Paparidis Z, Abbasi AA, Malik S, Goode DK, Callaway H, Elgar G, … Grzeschik KH (2007). Ultraconserved non-coding sequence element controls a subset of spatiotemporal GLI3 expression. Development Growth and Differentiation, 49(6), 543–553. 10.1111/j.1440-169X.2007.00954.x

Parisi C, Vashisht S, & Winata CL (2021). Fish-Ing for Enhancers in the Heart. International Journal of Molecular Sciences, 22(8). 10.3390/IJMS22083914

Parker HJ, Piccinelli P, Sauka-Spengler T, Bronner M, & Elgar G (2011). Ancient Pbx-Hox signatures define hundreds of vertebrate developmental enhancers. BMC Genomics, 12(1), 637. 10.1186/1471-2164-12-637 [PubMed: 22208168]

Parveen N, Masood A, Iftikhar N, Minhas BF, Minhas R, Nawaz U, & Abbasi AA (2013). Comparative genomics using teleost fish helps to systematically identify target gene bodies of functionally defined human enhancers. BMC Genomics, 14(1). 10.1186/1471-2164-14-122

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, … Rubin EM (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature, 444(7118), 499–502. 10.1038/nature05295 [PubMed: 17086198]

Perenthaler E, Yousefi S, Niggl E, & Barakat TS (2019). Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. Frontiers in Cellular Neuroscience, 13(July), 1–21. 10.3389/fncel.2019.00352 [PubMed: 30723396]

Radhakrishna U, Bornholdt D, Scott HS, Patel UC, Rossier C, Engel H, … Antonarakis SE (1999). The phenotypic spectrum of GLI3 morphopathies includes autosomal dominant

preaxial polydactyly type-IV and postaxial polydactyly type-A/B; no phenotype prediction from the position of GLI3 mutations. American Journal of Human Genetics, 65(3), 645–655. 10.1086/302557 [PubMed: 10441570]

Radhakrishna U, Wild A, Grzeschik KH, & Antonarakis SE (1997). Mutation in GLI3 in postaxial polydactyly type a. Nature Genetics, 17(3), 269–271. 10.1038/ng1197-269 [PubMed: 9354785]

Rebeiz M, & Tsiantis M (2017). Enhancer evolution and the origins of morphological novelty. Current Opinion in Genetics \& Development, 45, 115–123. [PubMed: 28527813]

Reymond A, Marigo V, Yaylaoglu MB, Leoni A, Ucla C, Scamuffa N, … Ballabio A (2002). Human chromosome 21 gene expression atlas in the mouse. Nature 2002 420:6915, 420(6915), 582–586. 10.1038/nature01178

Robson MI, Ringel AR, & Mundlos S (2019). Review Regulatory Landscaping : How Enhancer-Promoter Communication Is Sculpted in 3D. Molecular Cell, 74(6), 1110–1122. 10.1016/j.molcel.2019.05.032 [PubMed: 31226276]

Rodrigues MFSD, Miguita L, De Andrade NP, Heguedusch D, Rodini CO, Moyses RA, … Nunes FD (2018). GLI3 knockdown decreases stemness, cell proliferation and invasion in oral squamous cell carcinoma. International Journal of Oncology, 53(6), 2458–2472. 10.3892/ijo.2018.4572 [PubMed: 30272273]

Ruiz i Altaba A (2008). The Gli code. 17(9), 438–447. 10.1016/j.tcb.2007.06.007.The

Schoenfelder S, & Fraser P (2019). Long-range enhancer-promoter contacts in gene expression control. Nature Reviews. Genetics, 20(8), 437–455. 10.1038/S41576-019-0128-0

Segert JA, Gisselbrecht SS, & Bulyk ML (2021). Transcriptional Silencers: Driving Gene Expression with the Brakes On. Trends in Genetics, 37(6), 514–527. 10.1016/j.tig.2021.02.002 [PubMed: 33712326]

Tanaka M (2016). Fins into limbs: Autopod acquisition and anterior elements reduction by modifying gene networks involving 5'Hox, Gli3, and Shh. Developmental Biology, 413(1), 1–7. 10.1016/j.ydbio.2016.03.007 [PubMed: 26992366]

Theil T, Alvarez-Bolado G, Walter A, & Rüther U (1999). Gli3 is required for Emx gene expression during dorsal telencephalon development. Development, 126(16), 3561–3571. 10.1242/dev.126.16.3561 [PubMed: 10409502]

Tyurina OV, Guner B, Popova E, Feng J, Schier AF, Kohtz JD, & Karlstrom RO (2005). Zebrafish Gli3 functions as both an activator and a repressor in Hedgehog signaling. Developmental Biology, 277(2), 537–556. 10.1016/j.ydbio.2004.10.003 [PubMed: 15617692]

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, … Brenner S (2006). Ancient noncoding elements conserved in the human genome. Science, 314(5807), 1892. 10.1126/science.1130708 [PubMed: 17185593]

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, … Warren WC (2014). Elephant shark genome provides unique insights into gnathostome evolution. Nature, 505(7482), 174–179. 10.1038/nature12826 [PubMed: 24402279]

Visel A, Thaller C, & Eichele G (2004). GenePaint.org: an atlas of gene expression patterns in the mouse embryo. 10.1093/nar/gkh029

Vortkamp A, Franz T, Gessler M, & Grzeschik K-H (1992). Deletion of GLI3 supports the homology of the human Greig cephalopolysyndactyly syndrome (GCPS) and the mouse mutant extra toes (XI). Mammalian Genome, 3, 461–463. [PubMed: 1322743]

Wiegering A, Adibi P, Ruether U, & Gerhardt C (2020). Gain-of-function mutation in Gli3 causes ventricular septal defects. 10.1101/2020.02.10.942144

Wild A, Kalff-Suske M, Vortkam A, Bornholdt D, König R, & Grzeschik KH (1997). Point mutations in human GLI3 cause Greig syndrome. Human Molecular Genetics, 6(11), 1979–1984. 10.1093/hmg/6.11.1979 [PubMed: 9302279]

Woolfe A, & Elgar G (2007). Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. Genome Biology, 8(4). 10.1186/gb-2007-8-4-r53

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, … Elgar G (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biology, 3(1). 10.1371/journal.pbio.0030007

Yousaf A, Raza MS, & Abbasi AA (2015). The evolution of bony vertebrate enhancers at odds with their coding sequence landscape. Genome Biology and Evolution, 7(8), 2333–2343. 10.1093/gbe/evv146 [PubMed: 26253316]

Zehra R, & Abbasi AA (2018). Homo sapiens-specific binding site variants within brain exclusive enhancers are subject to accelerated divergence across human population. Genome Biology and Evolution, 10(3), 956–966. 10.1093/gbe/evy052 [PubMed: 29608725]

Major findings:

- Fish species with a slow rate of molecular evolution enable the identification of novel *Gli3* enhancers.
- Fins/limbs and the central nervous system are the major expression domains of these newly identified *Gli3* enhancers.
- A *Gli3* intronic enhancer (CNE14/hs1586) gained full limb specific-enhancer activity during the course of mammalian evolution.
- The delineated *cis*-regulatory catalogue provides novel targets for mutational analysis of *GLI3* associated developmental anomalies.
- Taken together, this work provides new understanding of the cis-regulatory complexity governing *Gli3* expression.

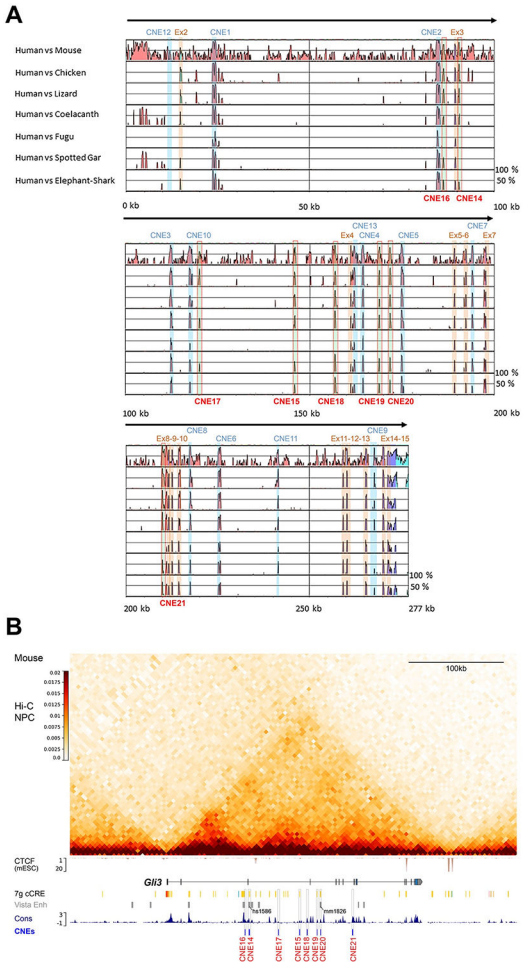

Figure 1. Comparative genomic analysis reveals deeply conserved *GLI3* intronic CNEs with associated enhancer identity

A) Multi-species sequence alignment of the genomic interval containing the human *GLI3* locus (ENSG00000106571) with orthologous counterparts of tetrapod (mouse, chicken and lizard) and aquatic vertebrates (coelacanth, fugu, spotted gar and elephant shark). Alignments are shown in VISTA (Visualization Tool for Alignment) graphical output by the Shuffle-LAGAN tool, using human sequence as a baseline. The horizontal black arrow on top indicates the direction of *GLI3* transcription and genomic extension of the human *GLI3* gene body (277 kb). The newly identified coelacanth and elephant shark-conserved non-coding elements (CNE 15-21) show absence of conservation in teleost fish (fugu) and framed in red (red labels). CNE14 shares this signature (also marked red). CNEs highlighted in light blue were previously identified. *Gli3* protein coding exons (Ex2-15) are marked brown. Criteria of alignment were 50bp and 50% conservation cut-off. Conserved coding and non-coding sequences are depicted as blue and pink peaks, respectively. The y-axis indicates percent identity and the x-axis informs about the extension of elements. Genomic location of newly and previously identified CNEs is listed in Table 1 and S1, respectively. kb, kilobase; Ex, exon CNE, conserved non-coding element. **B**) 3D chromatin interaction (Hi-C) heatmap spanning the *Gli3* gene body in neuronal progenitor cells (NPCs) from (Bonev et al., 2017). Chromatin interaction profiles across the entire *Gli3* TAD in

mouse embryonic stem cells (mESCs), NPCs and cortical neurons (CNs) is shown in Fig. S1A. CTCF profiles from mESCs (Bonev et al., 2017), general candidate cis-regulatory elements (cCREs) from ENCODE (Abascal et al., 2020) and the vertebrate conservation (phyloP60way) track are shown below along with the location of tested Vista Enhancer (Enh) Elements (Vista Enhancer browser) and newly identified CNEs. cCREs include predicted promoter-like elements (red), enhancer-like sequences (ELS, yellow) and CTCF-only sequences (blue) (Abascal et al., 2020).

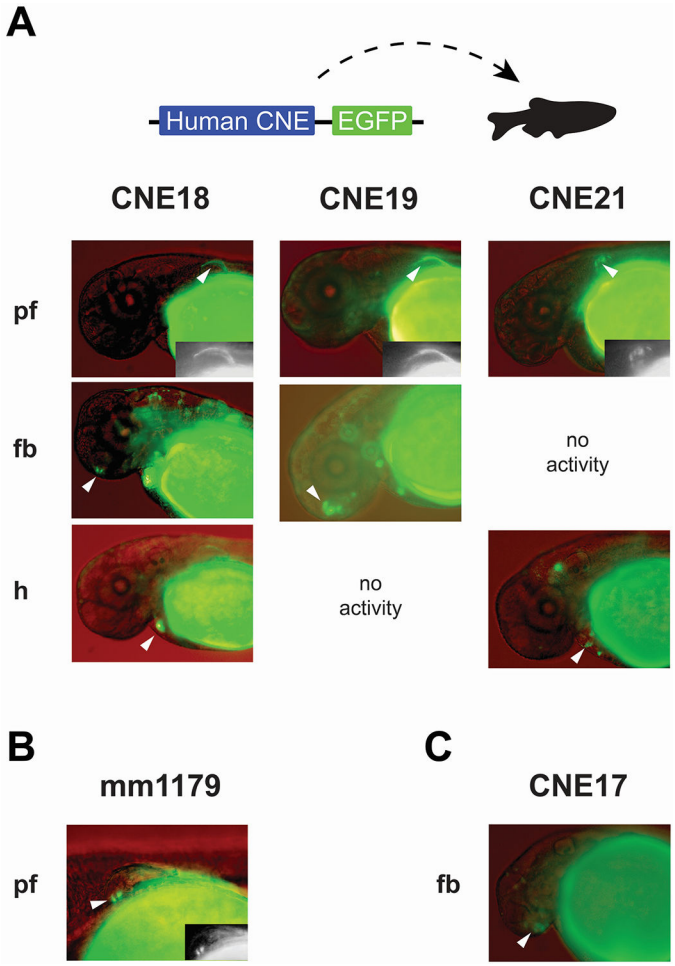

Figure 2. *Gli3*-CNEs conserved in slowly evolving fish genomes exhibit enhancer activity in *Gli3*-related tissues.

A) CNEs with highest sequence identity between human and elephant shark drive reproducible GFP transgenic reporter activity in multiple zebrafish embryonic tissues including the pectoral fin (pf), forebrain (fb) and heart (h). **B**) The mm1179 mouse embryonic *Gli3* limb enhancer has been validated in zebrafish for comparison and drives restricted activity in the pf. **C**) CNE17 promotes transcriptional activity exclusively in the fb. Representative images (merged bright field and fluorescent signals) of live zebrafish transgenic embryos are shown and White arrowheads indicate reproducible GFP reporter activities (see also Fig. S2 and Table S3). Inlets for elements active in the pf show GFP-only signals (in gray scale). Orientation of embryos is anterior to the left and dorsal to the top, with a lateral view.

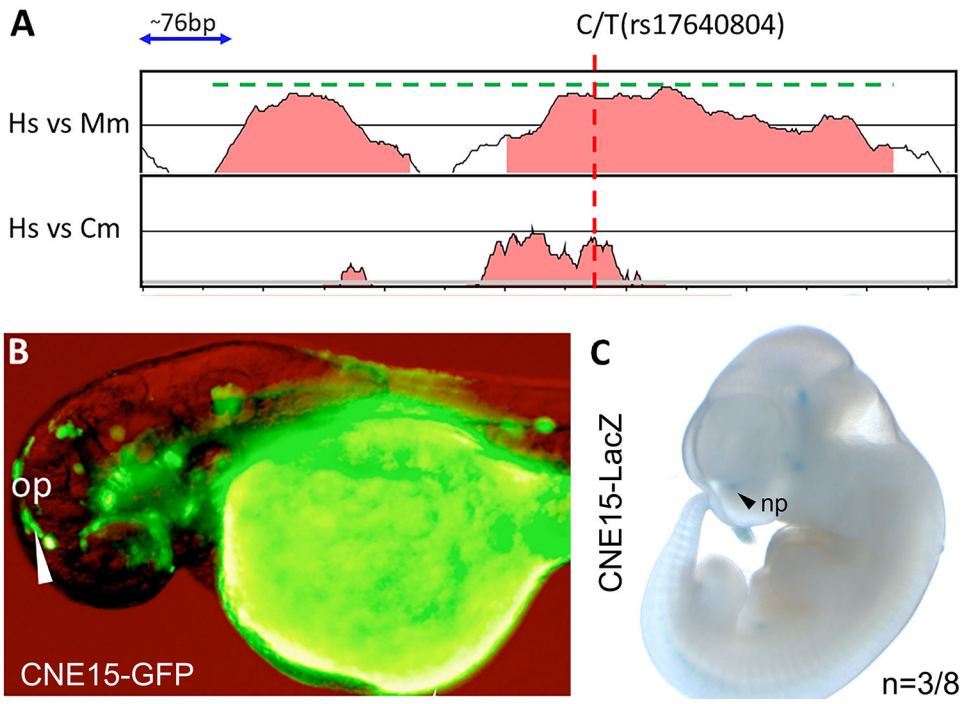

Figure 3. Implication of a *GLI3*-CNE in disease-related sequence variation.

A) Vista plot showing CNE15 deep conservation down to elephant shark (cm; Callorhinchus Milii) using human as a baseline. Human-mouse conserved sequence is shown underneath the dashed green line (566 bp) and is extended by ~50 bp in both directions to cover the human-mouse conserved peak included for transgenesis. **B**) CNE15 enhancer induced transgenic GFP reporter expression in the nasal pit of zebrafish embryos at 48 hpf (white arrowhead). **C**) CNE15 drives LacZ reporter activity in the presumptive upper nasal process (np) at the boundary to the forebrain in mouse embryos at E11.5 (black arrowhead). “n” indicates the number of transgenic mouse embryos with LacZ activity in the np. np, nasal pit; np, nasal process.

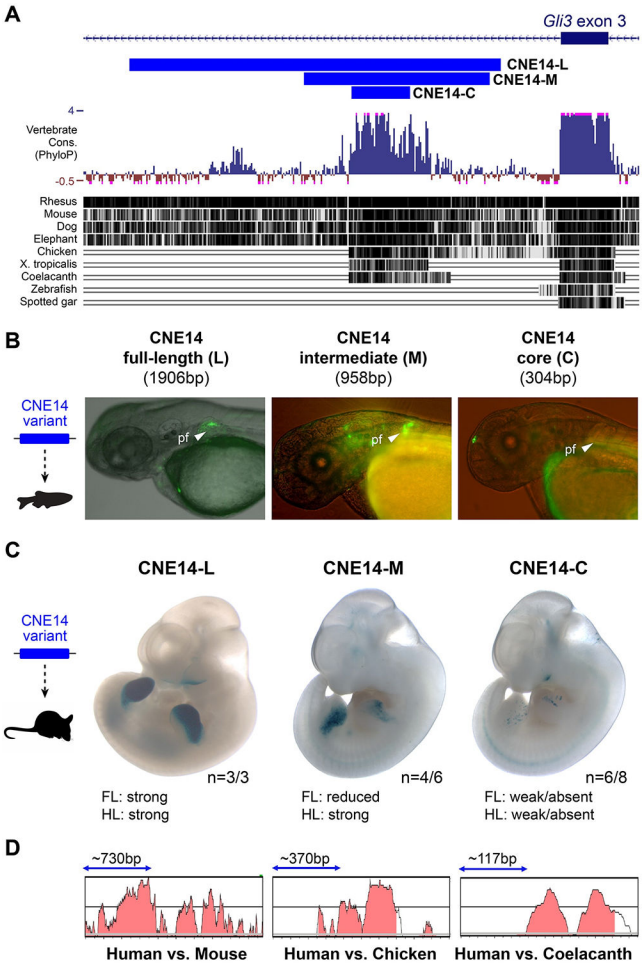

Figure 4. CNE extension correlates with evolutionary gain of enhancer function.

A) Genomic location and conservation of CNE14 variants based on UCSC PhyloP and MultiZ alignment tools. CNE14 human-mouse conserved region (CNE-L, 1906 bp), CNE14 intermediate human-chicken region (CNE-M, 958 bp) and the core conserved human-coelacanth sequence (CNE-C, 304 bp) are compared. **B**) CNE14-L and CNE14-M induce GFP expression in the developing zebrafish fish pectoral fin. **C**) Transgenic LacZ reporter activities of CNE14 versions in mouse embryos at E11.5. While CNE14-L induced strong LacZ expression in the anterior limb mesenchyme, activity driven by CNE14-M and CNE14-C was progressively reduced. “n” indicates the number of embryos with reproducible transgenic reporter activity versus the total number of transgenic embryos showing any LacZ signal. **D**) Pairwise sequence comparison of CNE14 by Vista plots. Blue double arrow indicates scale. FL, forelimb; HL, hindlimb.

Table 1.

Human-fish *GLI3* conserved non-coding elements selected for zebrafish/mice transgenic assay

Element	Location	GRCH38-Ch7 coordinates	Amplicon Size	Conservation Depth 50%;>50 bp	Predicted TFs
***GLI3*-CNE14-L**	Intron 3	42147909-42146003	1907 bp	Elephant Shark	Gata1, Rest, Creb1, Hnfa4, Nrf2f1, Pax6, Ppara, Hsf, Cdc5, Hsf1, Nfkb5
***GLI3*-CNE14-M**	Intron 3	42146901-42147858	958 bp	Elephant Shark	= =
***GLI3*-CNE14-S**	Intron 3	42147140-42147443	306 bp	Elephant Shark	= =
***GLI3*-CNE15**	Intron 3	42091574-42092239	666 bp	Elephant Shark	Nkx2-5, Foxj1, Foxd3, Foxi1, Foxo3, Foxo1, Nr2f1, Runx1
***GLI3*-CNE16**	Intron 2	42151204-42151889	686 bp	Elephant Shark	Dmrtc2, Tcf3, Foxo3, Abf1, Rfx1, Foxp3, Smad3, Foxo4, Foxj1, Foxo1, Hand1, Tal1
***GLI3*-CNE17**	Intron 3	42117459-42117998	540 bp	Spotted gar	Irf1, Atf1, Ebf1, Plzf, Evi1, Foxd3, Pou6f1
***GLI3*-CNE18**	Intron 3	42080686-42081557	872bp	Elephant Shark	Tal1, Tfap4, Pbx1, Mrf2, Nr1h3, Tgif1
***GLI3*-CNE19**	Intron 4	42068709-42069418	710 bp	Elephant Shark	Tfap4, Zbtb18, Foxa2, Pax6, Ets1, Elk1, Gzf1, Mafk, Dmrtc2, Tcf3
***GLI3*-CNE20**	Intron 4	42065743-42066301	559 bp	Elephant Shark	Rfx1, Hoxa4, Pou1f1, Foxq1, Freac3, Foxd3, Oct1
***GLI3*-CNE21**	Intron 7	42027388-42028300	913 bp	Elephant Shark	Zfp74, Myb, Oct1, Hoxa7, Pbx1, Lmo2
mm1179	Intergenic	42342873-42343965	1093 bp	Chicken	Pbx1, Elk1, Foxd3, Nrf2f1

This table presents chromosomal location, genome assembly coordinates (GRCH38), size of amplicon and conservation depth of CNEs identified in the current study. In addition, the mm1179 Vista enhancer and predicted transcription factor (TF) motifs in each element are also listed.