Received: 21 April 2020 | Revised: 18 December 2020 | Accepted: 19 December 2020

DOI: 10.1002/dvdy.291

RESEARCH ARTICLE

Developmental Dynamics WILEY

Comparative genomic analysis of human *GLI2* locus using slowly evolving fish revealed the ancestral gnathostome set of early developmental enhancers

Shahid Ali[1] | **Irum Arif**[1] | **Ayesha Iqbal**[1] | **Irfan Hussain**[1] | **Muhammad Abrar**[1] | **Muhammad Ramzan Khan**[2] | **Neil Shubin**[3] | **Amir Ali Abbasi**[1]

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

[2]National Institute for Genomics and Advanced Biotechnology, National Agricultural Research Center, Islamabad, Pakistan

[3]Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois

Correspondence
Amir Ali Abbasi, National Center for Bioinformatics, Program of Comparative and Evolutionary Genomics, Faculty of Biological Sciences, Quaid-i- Azam University, 45320, Islamabad Pakistan.
Email: abbasiam@qau.edu.pk

Neil Shubin, Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois, USA.
Email: nshubin@uchicago.edu

Funding information
Brinson Foundation; NRPU, Grant/Award Number: 20-2085/NRPU/R&D/HEC/12/760

Abstract

Background: The zinc finger-containing transcription factor Gli2, is a key mediator of Hedgehog (Hh) signaling and participates in embryonic patterning of various organs including the central nervous system (CNS) and limbs. Abnormal expression of Gli2 can impede the transcription of Hh target genes through disruption of proper balance between Gli2 and Gli3 functions. Therefore, delineation of enhancers that are required for complementary roles of Glis would allow the interrogation of those pathogenic variants that cause gene dysregulation, and a corresponding abnormal phenotype. Previously, we reported tissue-specific enhancers for Gli family including Gli2 through direct tetrapod-teleost comparisons.

Results: Here, we employed the sequence alignments of slowly evolving spotted gar and elephant shark and have identified six novel conserved noncoding elements in human *GLI2* containing locus. Zebrafish-based transgenic assays revealed that combined action of these autonomous CNEs reflects many aspects of *Gli2* specific endogenous transcriptional activity, including CNS and pectoral fins.

Conclusion: Taken together with our previous findings, this study suggests that Hh-signaling controlled deployment of Gli2 activity in embryonic patterning arose in the common ancestor of gnathostomes. These *GLI2* specific *cis*-regulatory modules will help to identify DNA variants that probably reside outside of coding intervals and are associated with congenital anomalies.

KEYWORDS

CNEs, elephant shark, enhancers, gar, Gli family, Gli2, gnathostomes, Hh signaling, transgenesis, zebrafish

1 | INTRODUCTION

Sonic hedgehog signaling pathway is a signaling cascade that controls a number of key developmental events and embryological processes.[1] The Hh pathway plays a critical role in patterning and growth of several organs including the central nervous system (CNS), limbs and skeleton, and internal organs.[2,3] The effects of Sonic hedgehog (Shh) protein on target cells are transduced through Gli family of proteins, which are zinc finger-containing transcriptional factors.[4] In vertebrates, three *GLI* genes exist which encode for Gli1, Gli2, and Gli3.[5] Biochemical and genetic investigations indicate that Gli2 and Gli3 proteins are primary and immediate mediators to Shh signals, whereas Gli1 appears to play a secondary role in potentiating response to the secreted Shh molecule.[6] Invertebrates, such as the fly (*Drosophila melanogaster*) possesses a single zinc finger-containing homolog of vertebrate Gli genes called *Cubitus interruptus* (*Ci*), which transduces all Hh induced signals.[7] Moreover, additional gene duplication events in teleosts forms the basis for functional diversity in Hh signaling pathway different from that found in tetrapods.[8] For instance, molecular evolutionary analysis of Gli family in vertebrates revealed that *Gli2* underwent teleost specific duplication which resulted in the birth of two copies of this gene (*gli2a* and *gli2b*) in teleosts.[4,9] In vertebrates the Gli2 and Gli3 proteins preserved the transcriptional bi-functionality present in their Drosophila homolog Ci.[10] In contrast, the vertebrate Gli1 functions only as transcriptional activator and its roles during embryonic development are widely reported as dispensable.[11]

Gli genes are known to be expressed in broad range of vertebrate tissues and cell types.[12,13] Studies in mice, frog and zebrafish have revealed that a complex Gli code, the fine-tuned expression of *Gli* genes, both temporally and spatially, provides a basic tool that is used at many stages during early embryogenesis[13] including patterning of CNS, morphogenesis of paired appendages, muscle fiber formation, and the induction of sclerotome, prostate gland, and craniofacial structures.[13-18] These data suggest that *Gli* paralogs are vital for normal morphogenesis, and their aberrant expression could result in developmental defects.

Among Gli proteins, Gli2 has unique characteristic features during early embryonic development and postnatally. For instance, Gli2 has both a repressor protein domain and the two activator protein domains and could in principle function as both a transcriptional repressor and an activator.[5] However, biochemical and molecular data suggest that in most cases Gli2 protein activates transcription of target genes, because posttranslational processing of Gli2 protein into Gli2R protein is inefficient. In contrast, Gli3 which is the duplicated counter-part of Gli2 acts primarily as a transcriptional repressor in most of the cellular contexts and its loss results in gain-of-function phenotypes of Hh signaling.[19] Moreover, genetic analysis in mice revealed that Gli2 null mutants die before birth with defects in several tissues including brain, craniofacial skeleton, vertebrae, long bones of the limb (stylopod and zeugopod), sternum, lungs, and other internal organs.[20] Gene mis-expression studies in mice demonstrated that Gli2 protein is a bona fide transcriptional activator during development.[21] Developmental events regulated by Gli2 protein become evident in diseases associated with mutations of human *GLI2* gene. Nonsynonymous heterozygous *GLI2* mutations underlie a significant fraction of congenital hypopituitarism and also cause holoprosencephaly-like malformations, including abnormal craniofacial development, growth retardation, limb deformities and microcephaly.[22] Over expression of Gli2 in human and mice causes various kinds of cancers including basal cell carcinoma, hepatocellular carcinoma, medulloblastoma, prostate, and breast cancer.[23-26]

Comparative functional analysis demonstrates the divergent requirements of Gli2 in fish and mammals.[27] Duplicated Gli2 copies of zebrafish (*gli2a*/*gli2b*) mediate Hh-dependent activating role and also Hh-independent repressor role during embryogenesis and play redundant as well as specific roles during zebrafish development.[27] Depending upon the developmental stage, the endogenous expression pattern of *gli2a* and *gli2b* overlaps as well as differ substantially.[8] For instance, at the end of gastrulation period the *gli2b* transcript is detected in the anterior neural keel, but not in the region that corresponds to forebrain-midbrain boundary, whereas in contrast the *gli2a* transcripts are detected broadly in the lateral mesoderm and region in the neural keel that corresponds to future midbrain.[9] At 24 hours postfertilization (hpf), *gli2b* transcripts are detected in telencephalon, optic tectum, posterior thalamus, and hindbrain, including the region that corresponds to cerebellum, whereas the cross-sections analysis revealed the absence of *gli2b* transcript from the ventral neural tube.[9] In contrast, at 24 hpf, the *gli2a* mRNAs are detected broadly in the anterior diencephalon, throughout the midbrain and the midbrain-hindbrain boundary (MHB), but in hindbrain, the reduced level is detected. In MHB, the *gli2b* expression is detected only in the rhombomere-1, whereas *gli2a* mRNAs are found in entire MHB.[27] In the ventral diencephalon, the *gli2a* transcripts are present in the posterior hypothalamus, whereas *gli2b* mRNAs are also detected in posterior hypothalamus. These comparative expression data of Gli2s in zebrafish reveal both redundant and unique roles for paralogs during zebrafish development.

Subfunctionalization of zebrafish Gli2s might involve both sequence divergence as well as the acquisition of novel expression patterns among duplicates.[8]

Biochemical, molecular, animal model, and human disease data demonstrates that correct expression of Gli2 in space and time is crucial to ensure appropriate deployment of Hh signaling levels and normal embryogenesis. Significant work has been conducted to reveal the molecular and developmental functions of Gli2 proteins in vertebrates. However, the genetic regulation of Gli2 expression during development remains to be elucidated. Dynamic expression patterns of Gli2 gene in the vertebrate embryo suggest that multiple independent *cis*-regulatory regions might work in a combinatorial fashion to control spatiotemporal and developmental stage specific aspects of its expression.[28] Consistent with this assumption, a previous study by our research group identified five intronic sequences which are conserved between teleosts and tetrapods.[28] Although *GLI2* associated enhancers regions identified in our previous study drove reporter gene expression in several embryonic domains, including neural tube and pectoral fin, apparently none of them fully reproduce all known aspects of *Gli2* dynamic endogenous expression pattern in vertebrates.[28] Therefore, additional *cis*-acting elements participate in the control of the Gli2 spatiotemporal expression must exist.

Given the assumption that additional unreported *cis*-acting elements for *Gli2* expression exist, it becomes necessary to introduce a finer resolution to enhancer discovery than has been applied thus far.[28] Previously published data was derived from comparison of human/mouse genomes to those of teleost fish such as fugu and zebrafish to search for CNEs. CNEs often function as *cis*-acting regulators,[28,29] however many appear to be undetected in fugu and zebrafish, presumably because of whole genome duplication of teleost fish and their rapid subsequent evolution in some lineages.[30] To enhance the resolution of phylogenetic footprinting, we performed comparative analysis of mammalian *Gli2* locus by including genome sequence of slowly evolving spotted gar (*Lepisosteus oculatus*), elephant shark (*Callorhinchus milli*), and coelacanth (*Latimeria chalumanae*).[31-33] Intriguingly, inspections of human-centric alignments have uncovered a set of novel CNEs that were undetectable in direct human-teleost comparisons. These analyses revealed that large majority of *Gli2*-associated CNEs also aligned to elephant shark, highlighting their presence in gnathostome ancestor. Novel CNEs identified in the present study are subjected to functional analysis by employing transgenic zebrafish assay. Majority of them induced reporter gene expression in known endogenous sites of *gli2a* transcription in zebrafish, including neural tube and pectoral fins.

2 | RESULTS

2.1 | Multispecies sequence alignment of *GLI2* loci among distantly related vertebrate species

Previously, we reported five *GLI2* intronic CNEs through tetrapod-teleost comparative sequence analysis and transgenic zebrafish enhancer assay.[28] In order to search for previously undetected enhancers, we undertook further investigation of human *GLI2* locus by exploiting the recent availability of slowly evolving fish genomes; coelacanth, spotted gar, and elephant shark.[31-33] A multispecies comparative analysis of human *GLI2* locus encompassing 100 kb intergenic regions with its orthologous counterparts from mouse, chicken, lizard, coelacanth, fugu, spotted gar, and elephant shark revealed six novel human-fish CNEs that were not detected in direct human-teleost comparisons (Figure 1). These elements are named as CNE6, CNE7, CNE8, CNE9, CNE10, and CNE11. CNE6 is embedded in *GLI2* intron-1 whereas rest of CNEs identified in the current study is positioned in downstream interval of *GLI2* (Table 1). Taken together with previously published data, human *GLI2* locus harbors in total 11 human-fish conserved noncoding elements with criteria of at least 50% identity over 50 bp sequence window (Figure 1 and Table 1). In silico analysis of transcription factor binding sites (TFBS) within *GLI2*-associated CNEs revealed binding preferences for well-established developmental regulators (Table 1, Additional file 1: Figure S1). To further verify the regulatory potential of *GLI2*-associated CNEs, their genomic coordinates were overlapped with reported Chip-seq marks from ENCODE.[34] These analyses revealed the presence of histone modification marks H3K27ac and H3K4me1, on CNE6 while rest of CNEs carry H3K4me1 and H3K4me3 marks and low enrichment of H3K27ac (Additional file 2: Table S2) representing the poised state of these CNEs in human embryonic stem cells. Moreover, we estimated the total number of active regions in the *GLI2* locus by overlapping the *GLI2*/*Gli2* coordinates with the ENCODE dataset of H3K4me1, H3K27ac in seven hESCs, human fetal leg muscles, mouse ES cells, limbs, neural tube, and embryonic facial prominence. Within human *GLI2* locus (100 kb on either side of *GLI2* gene) 1022 histone modification peaks (size 100-1000 bp) are enriched with human H3K4me1 marks and 278 peaks (size 100-2000 bp) enriched with human H3K27ac marks (Additional file 3). Mouse *Gli2* locus overlapped with mouse ES Chip-seq data revealed 254 peaks of H3K4me1 and 52 peaks of H3K27ac (Additional file 4). Mouse Chip-seq data revealed 730 H3K4me1 and 550 H3K27ac peaks for

FIGURE 1 Multispecies comparative plot of human *GLI2* locus. Multispecies comparative analysis of human *GLI2* locus and 100 kb flanking region from both directions with orthologous counterparts of tetrapods (mouse, chicken and lizard) and aquatic vertebrates (coelacanth, fugu, spotted gar, and elephant shark). Presented here are visualization tool for alignment (VISTA) graphical output by Shuffle-LAGAN tool using human sequence as the baseline. The black arrow above the plot shows the direction of transcription and length of *GLI2* gene (257 kb). The conserved noncoding elements highlighted in light green and indicated on top as red numbers, are selected for functional analysis in this study. CNEs indicated in black are reported previously by our research group.[28] Criteria of alignment were 50 bp and 50% conservation cutoff. Conserved coding and noncoding sequences are depicted in blue and pink peaks, respectively. Y-axis indicates percent identity and x-axis indicates the length of sequences. kb, kilobase; Ex, exon; CNE, conserved noncoding element

embryonic limb, 431 H3K4me1 and 550 H3K27ac peaks for facial prominence outgrowth and 244 H3kme1 and 76 H3K27ac peaks for neural tube (Additional file 4). Additional analysis involving intersection of Chip-seq peaks from human and mouse regions defined a non-redundant list of 24 possible enhancer regions in the *GLI2/Gli2* locus (Additional file 5). These regions are H3K4me1+ (for 7/7 hESC; human fetal leg muscle) and H3K27ac + (for 5/6 hES cell lines; human fetal leg muscle).

2.2 | Association of selected CNEs with target gene body

Evolutionary conserved vertebrate enhancers are known to maintain synteny with target genes across distantly related vertebrate species.[35] Empirical data suggest that CNE-enhancer regions act on a target gene promoter from large genomic distances and even known to reside in the intronic intervals of neighboring genes. For instance, enhancers have been reported to act at a distance of 2 Mb

TABLE 1 Human-fish *GLI2* conserved noncoding elements selected for zebrafish transgenic enhancer assay

Element	Location	GRCH38-Ch2 coordinates	Amplicon size	Conservation depth 50%; >50 bp	Known TFs
GLI2-CNE5	Intron 1	120805726-120806383	658 bp	Elephant shark	AIRE, EN, RAP1, PBX, TCF11,NRF-2, HOXA7, SOX17, FOXO3, FOXC1, TFII-I, MAZ
GLI2-CNE6*	Intron 1	120 887 508-120 887 906	399 bp	Elephant shark	PBX1, FOXM1, DMRT2, ARF2 RREB1, NKX2, FOXJ2, HOXA9 DMRT5, OCT1, PAX6
GLI2-CNE1	Intron 1	120909974-120910966	993 bp	Elephant shark	TTK, FREM1, SMAD3, SOX10, HMX3, NKX2-5, EN, FTZ, LHX3, PTX1, GATA-3, GATA-2
GLI2-CNE2	Intron 2	120931931-120932461	531 bp	Elephant shark	AML1, PBX-1, HOXA4, CHX10, TCF11, OCT_4, NRSF, E2F, TCF11, SREBP-1
GLI2-CNE3	Intron 4	120962305-120963154	850 bp	Elephant Shark	EBF, RORALPHA1, STAT1, FOXA2, PTX1, POU3F2, APEX1, GATA-1,NF-KAPPAB
GLI2-CNE4	Intron 7	12097 529-120973826	1298 bp	Elephant shark	STAT5B, POU5F1, BRN-2, PAX-5, CEBP, GATA-1, DMRT2, PBX, ALX-4, SOX17, CREB
GLI2-CNE7*	Intergenic	121030842-121031476	635 bp	Elephant shark	OTX1, EVI-1, ARF, OSF2
GLI2-CNE8*	Intergenic	121038993-121039763	771 bp	Coelacanth	HOXA7, PAX5, ATF4,SOX17, HOXA5, HOXA4,NKX3-1, LHX3
GLI2-CNE11*	Intergenic	121058716-121059672	957 bp	Elephant shark	XBP1, FOXD3
GLI2-CNE9*	Intergenic	121058659-121059444	787 bp	Elephant shark	POU3F2, FOXA2, PAX-6 GCNF, C-ETS-2, EGR1
GLI2-CNE10*	Intergenic	121064108-121064879	772 bp	Elephant shark	TAL1, NFASC, TBP, FOXJ1, PLXNA2
hs1790	Intron 1	120804710-120808132	3423 bp	Vista element	—
hs522	Intergenic	121074842-121075741	900 bp	Vista element	—

Note: This table presents the chromosomal location, assembly coordinates (GRCH38), size of amplicon, and conservation depth of CNEs previously reported by Minhas et al 2015 and those identified in the current study. *GLI2*-CNEs identified in the present study and shortlisted for functional testing are marked with asterisk sign (*). In addition, known transcription factors binding on the identified CNEs are indicated.

or more from target gene body.[35,36] Furthermore, expression of a typical animal gene is likely to be governed by multiple distinct enhancer elements and in some cases individual enhancers are known to regulate multiple genes.[35,37] Therefore, the association of functionally confirmed enhancers with their target genes have remained a challenging task. To associate the CNEs identified in the present study with their target gene body, we employed comparative genomics and endogenous expression pattern analyses. *Gli2* and its flanking 1 Mb region on either side are comparatively analyzed in human, mouse and spotted gar (Figure 2A). Intriguingly, the comparative synteny data suggest that *Gli2* locus has remained unaltered during the course of vertebrate evolution, with *Gli2* and its flanking genes *Inhbb* and *Tfcp2l1* have remained syntenicaly linked in distantly related vertebrate genomes (Figure 2A). Furthermore, RNA in situ hybridization-based endogenous expression patterns of mouse *Gli2* along with its flanking genes *Tfcp2l1* and *Inhbb* was inspected (Additional file 2: Table S3). The expression data suggest that *Gli2* coexpresses with its neighboring genes. Therefore, combined employment of comparative synteny and endogenous expression pattern analysis failed to precisely assign target gene to the subject CNEs. Distal regulatory elements are known to be in spatial proximity with the target gene bodies.[2] This occurs due to loops in the three-dimensional structure that brings distant genes and regulatory region into physical contacts.[38] Such physical interactions between enhancers and their target gene promoters can be identified by chromatin conformation capture technologies.[39] To solve the puzzle, we next attempted to collect high-resolution Hi-C (chromosome capture by high throughput sequencing) data from human embryonic stem cells line (H1-ESC) available at 3D genome browser

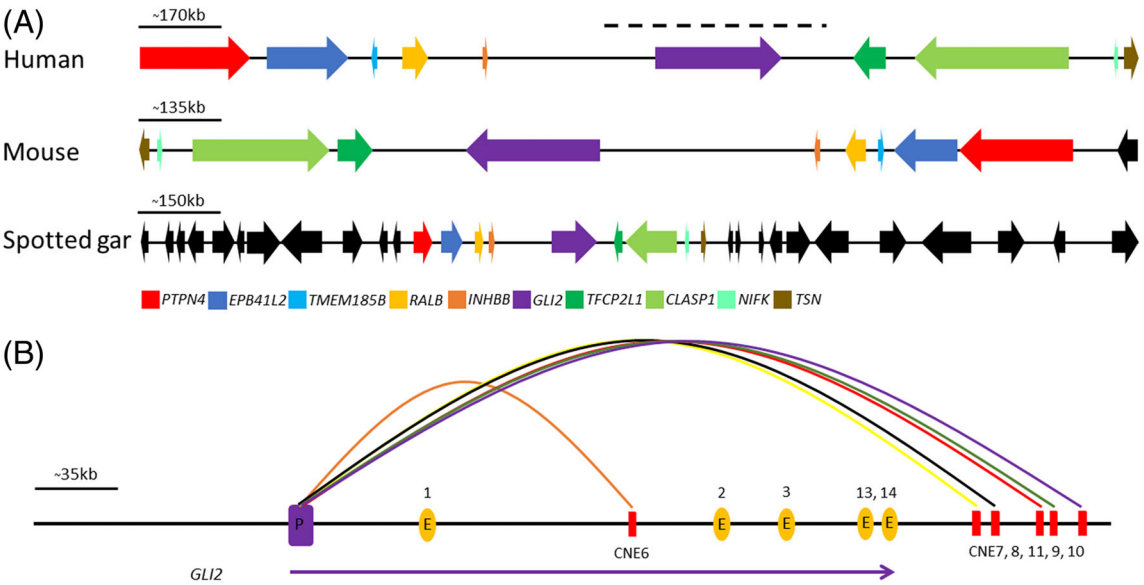

FIGURE 2 Syntenic mapping of *GLI2* locus and *GLI2* promoter physical interaction with intragenic and intergenic CNEs. Panel A: Comparative syntenic investigation of human-mouse and spotted gar *Gli2* containing orthologous loci reveals the deep conservation of this genomic block down to fish. This 2 MB orthologous block consists of conserved genes *GLI2*, *TFCP2L1*, *CLASP1*, *NIFK*, *TSN*, *INHBB*, *RALB*, *TMEM185B*, *EPB41L5*, *PTTPN4*. The endogenous expression pattern analysis and orthology mapping failed to exclusively associate CNEs identified in the present study with a single gene body. Conserved genes are color coded (other than black), whereas non-conserved genes are depicted in black. The genic architecture of loci is drawn according to the approximate scale (blackline). The black dashed line on top of *GLI2* locus showing the region presented in Panel B. Panel B: The virtual 4C data obtained from the 3D genome browser shows the physical interaction of intragenic and intergenic CNEs with *GLI2* promoter. The blue rectangle depicts *GLI2* promoter while CNEs are depicted as red rectangles. The orange ovals are *GLI2* exons. The parabolas direct the CNEs interaction with promoter. The purple arrowhead underneath depicts the direction of *GLI2* transcription

(http://promoter.bx.psu.edu/hi-c/virtual4c.php).[40] The human embryonic stem cell line (H1-ESC)-based virtual 4C data revealed that CNEs identified in present study can potentially interact with *GLI2* promoter (Figure 2B). Physical interactions between CNEs and *GLI2* promoter were further verified in additional cell lines known to express GLI2 endogenously (Additional file 6: Figure S2).

2.3 | Functional analysis of *GLI2*-associated CNEs

To determine the in vivo gene regulatory potential of CNE identified in the present study, we employed a *Tol2*-based transgenic enhancer assay in zebrafish.[41] For this purpose, human CNEs were PCR amplified and cloned in reporter vector containing enhanced green fluorescent protein (EGFP) coding gene (Additional file 7). The recombinant reporter constructs along with transposase mRNA were injected in zebrafish fertilized eggs at the one cell stage. The embryos were raised for 24 hours and then screened for the reporter gene expression at different developmental time points.

CNE6 resides within intronic intervals of *GLI2* and depicts 60% identity over 125 bp between humans and elephant sharks (Figure 1). In transgenic zebrafish, the CNE6 region induced the reporter gene expression specifically in developing pectoral fin bud at 48 hpf (Figure 3). For this particular element, strong reporter expression is observed in ~ 70% of the injected zebrafish embryos (Additional file 8: Table S4). Furthermore, two intergenic CNEs (CNE7 and CNE8) residing in the downstream interval of *GLI2* and depicts ~70% identity over 140 and 85 bp sequence length, respectively, in humans and elephant sharks, induced the reporter gene expression exclusively in the developing pectoral fin of zebrafish at 48 hpf (Figure 1, Table 1 and Additional file 8: Table S4). Therefore, the present study reports three independent enhancer elements directing the reporter gene expression in the pectoral fin of zebrafish embryos.

CNE10 and CNE11 reside within a downstream intergenic region of *GLI2* and depicts 65% and 70% identity, respectively, in human and elephant shark (Table 1). In transgenic zebrafish, human CNE10 was able to induce *GFP* expression exclusively in forebrain at 24 hpf (Figure 4 and Additional file 8: Table S4). Whereas, CNE11 induced expression was detected in developing zebrafish hindbrain and spinal cord at 48 hpf (Figure 4 and Additional file 8: Table S4). The current study identified in total two enhancer elements with CNS specific regulatory activity.

CNE9 resides in the downstream intergenic intervals of human *GLI2* and depicts 70% identity over 85 bp

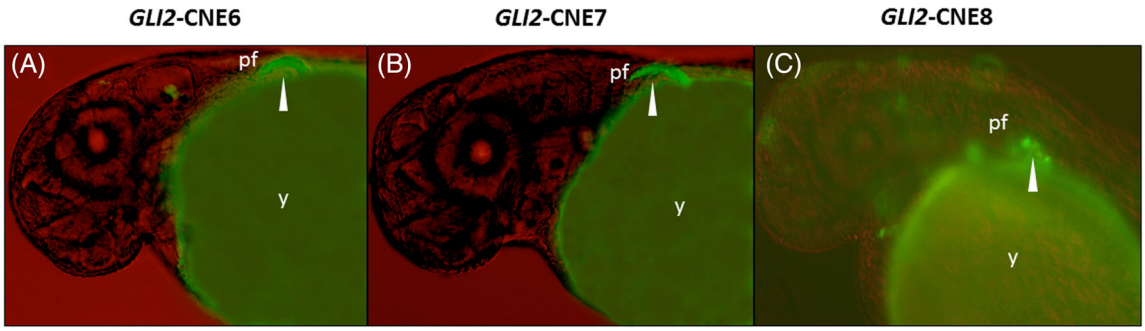

FIGURE 3 Subset of *GLI2*-associated CNEs depict seemingly redundant enhancer activity in the pectoral fin-bud of zebrafish. Representative images of live zebrafish transgenic embryos are shown (merged bright-field and fluorescent microscopy), orientation of the embryos is anterior to left, dorsal to top with lateral views. White arrowheads point to the GFP signals in transgenic zebrafish expressing cells in the developing pectoral fin at 48 hpf, A-C, induced by CNE6, CNE7, and CNE8, respectively. CNE6 is *GLI2* intragenic element while CNE7 and CNE8 are intergenic elements. hpf, hours postfertilization; pf, pectoral fin; y, yolk sac

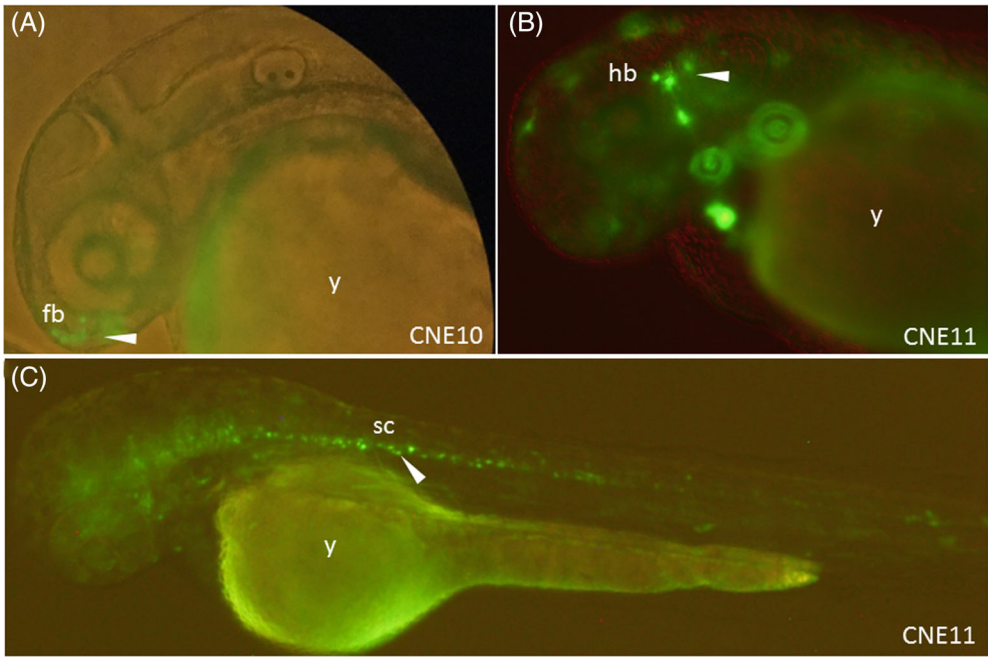

FIGURE 4 CNE10 and CNE11 induced the reporter gene expression in the brain and spinal cord. Representative images of live zebrafish transgenic embryos are shown (merged bright-field and fluorescent microscopy), orientation of the embryo is anterior to left, dorsal to top with lateral views. White arrowheads point to the GFP signals in transgenic zebrafish expressing cells in the forebrain at 24 hpf, A, and hindbrain and spinal cord at 48 hpf, B,C. hpf, hours postfertilization; fb, forebrain; hb, hindbrain; sc, spinal cord; y, yolk sac

sequence length in human-elephant shark comparison. In transgenic zebrafish, human CNE9 drove the *GFP* expression exclusively in pumping heart at 24 and 48 hpf (Figure 5) in 55% of injected embryos (Additional file 8: Table S4). In addition to pectoral fin, the CNE6 (aforementioned) triggered *GFP* expression was also detected in pumping heart at 48 hpf in 42% of the injected embryos (Figure 5 and Additional file 8: Table S4).

2.4 | Endogenous expression of *gli2a* in zebrafish embryos

In order to determine the endogenous expression patterns of *gli2a* in zebrafish; riboprobe-based whole mount in situ hybridization was performed. The in situ hybridization results showed that *gli2a* expression is widespread and is not spatially restricted at 24 hpf. At 48 and 50 hpf, staining is lost in most of the trunk region and tail but still stronger expression is observed in epidermis, forebrain, hindbrain, retinal choroid fissure, branchial arches, heart, and pectoral fins (Figure 6).

3 | DISCUSSION

Identification of *cis*-regulatory enhancers for human developmental genes can be accomplished by employing various approaches including biochemical analysis, chromatin conformation capture assays, epigenomic profiling,

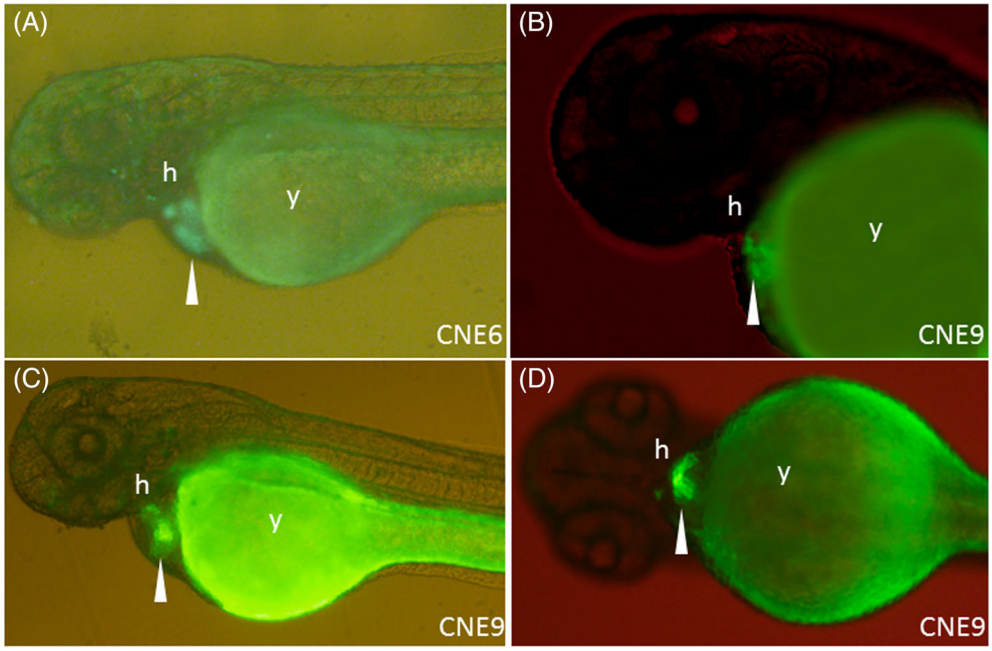

FIGURE 5 Heart specific cis-regulatory activity was observed for CNE6 and CNE9. Transgenic zebrafish embryos expressing the reporter gene are shown (merged bright-field and fluorescent microscopy), lateral view in panels A, B, C and ventral view in panel D. The white arrowhead points to the reporter gene signal of transgenic zebrafish expressing cells in developing heart at 48 hpf, A, induced by CNE6, 24 hpf, B, and 48 hpf, C,D, induced by CNE9. hpf, hours postfertilization; h, heart; y, yolk sac

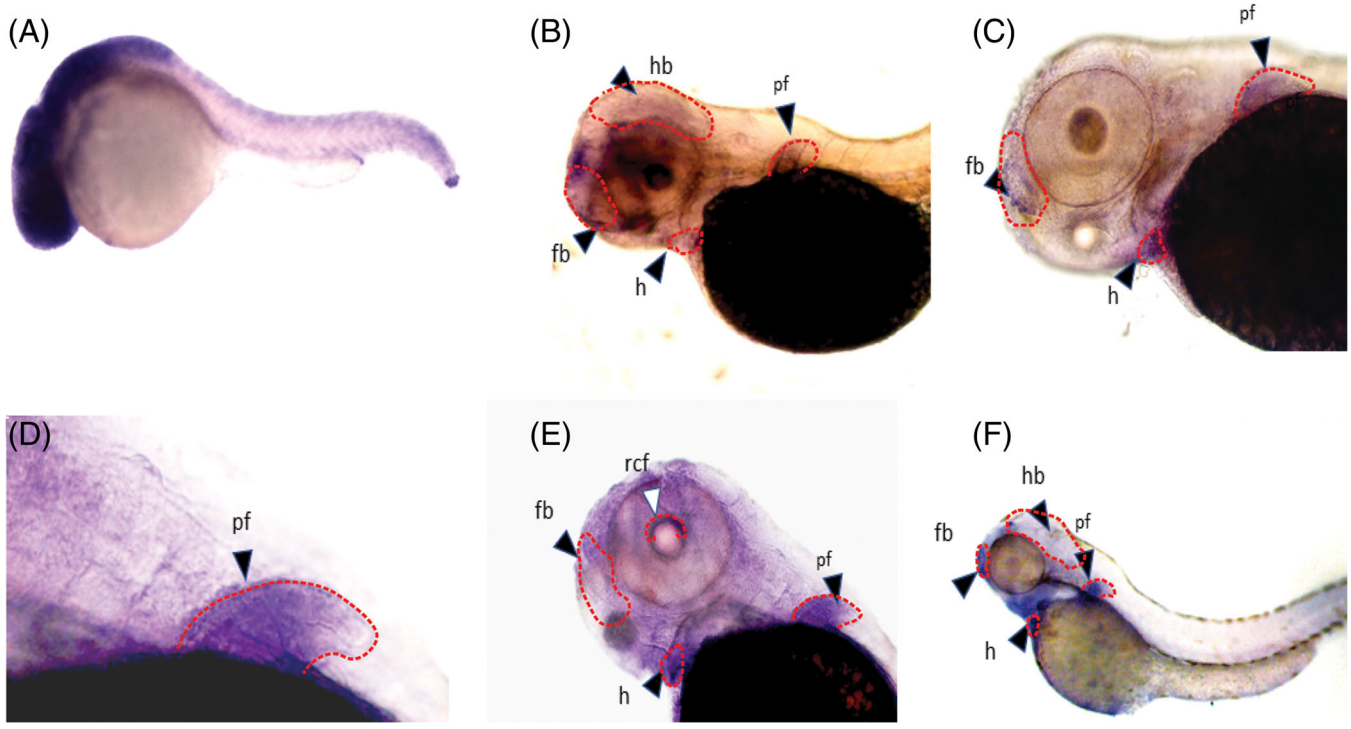

FIGURE 6 Whole mount in situ hybridization of *gli2a* in developing zebrafish embryos. Expression of Zebrafish *gli2a* mRNA during embryogenesis determined by whole mount in situ hybridization. A, 24 hpf lateral view. B,C, 50 hpf lateral view. D, 48 hpf lateral view of pectoral fin with high resolution. E,F, 48 hpf lateral view. At 24 hpf, the expression is not spatially restricted as *gli2a* is ubiquitously expressed, A. At 48 and 50 hpf, staining is lost from most of the trunk and tail but is still strong in the epidermis, dorsal CNS, pectoral fin, branchial arches, and retina choroid fissure. The black and white arrowheads point to the expression domain. hpf, hours postfertilization; fb, forebrain; hb, hindbrain; pf, pectoral fin; h, heart; rcf, retina choroid fissure

phylogenetic footprinting, and transgenic reporters assays.[42] For instance, biochemical methods; like Chip-seq and DNase-seq have been instrumental in characterizing certain genomic regions that act as *cis*-acting regulatory enhancers for gene expression.[43] Furthermore, comparative genomics-based approaches have

facilitated the identification of *cis*-acting gene regulatory networks by employing sequence comparison among distantly related vertebrate species.[44-46] For instance, comparative analysis of human-teleost fish orthologous loci has pinpointed the *cis*-acting gene regulatory network for important developmental genes like; *SHH*, *GLI3*, *SIM1*, *HOXD*, and *HOXA*.[12,29,47-49]

Gli2 is one of the transcription factors for signal transduction of the hedgehog pathway.[50] Functional investigation of Gli2 in mouse and zebrafish reveals that Gli2/gli2a play an essential role in the embryonic patterning of limbs/fins, CNS, and internal organs.[8,20] Therefore, *cis*-acting regulatory control is indispensable for the spatiotemporal expression regulation of *Gli2*.

Comparative analyses of vertebrate genomes have revealed that a large proportion of conserved sequences that reside in noncoding portions of the genomes rather than protein-coding intervals.[51] Vertebrate genomes hold thousands of evolutionary conserved noncoding regions. Several reports have shown that such conserved noncoding genomic regions, are often clustered in the neighborhood of genes essential for early embryonic development and transcriptional regulation.[46] Functional assays by employing cell lines and transgenic animal models have revealed that a substantial proportion of vertebrate CNEs act as *cis*-regulatory elements, such as enhancers that regulate the tissue-specific expression of the associated gene body during early stages of embryonic development.[46] Human and teleost fish lineages diverged approximately 450 Mya and it was assumed that noncoding genomic sequences conserved between them are under selective constraint and thus more likely to be functional.[52] Indeed, whole-genome comparisons of human with teleost fish like fugu and zebrafish, have identified several hundreds of CNEs, and the functional assay of a subset of them revealed that many human-teleost fish CNEs are tissue specific enhancers.[12,46] Therefore, teleost genomes are considered an ideal model for the identification of evolutionarily conserved tissue specific enhancers for human developmental genes.

Recently, the genome of a cartilaginous fish, the elephant shark was generated.[53] Cartilaginous fishes are the sister group of bony vertebrates (such as teleost fish and mammals) and therefore serve as a valuable outgroup for predicting the ancestral jawed vertebrate genomic features, as well as for inferring the shared and derived aspects of bony vertebrate lineages.[53] In addition, the genome of a basal nonteleost ray-finned fish, the spotted gar has been sequenced.[33] Both of these taxa lack the whole genome duplication and rapid rate of genome evolution seen in teleosts.

Comparative analyses have shown that the teleost fish (including fugu and zebrafish) have lost a substantial proportion of the ancestral jawed vertebrate CNEs that are still conserved in other jawed vertebrates including human, gar and elephant shark.[30] The nucleotide substitution rate of CNEs retained in teleosts showed that the teleost CNEs are evolving faster than their counterparts in humans, gar and elephant shark, suggesting that the fast nucleotide substitution rate of noncoding sequences is responsible for the absence of ancestral jawed vertebrate CNEs in teleosts.[54] Furthermore, genome-wide comparisons showed that a large majority of teleost CNEs are diverged beyond recognition despite their putative target genes still being present in genomes, suggesting that the disappearance of CNEs was not due to the loss of their target gene in teleost fish.[53] Comparisons of human (and other mammals) and teleost fish genomes together with spotted gar and/or elephant shark genomes have been much more useful in identifying CNEs that have evolved in the common ancestor of bony vertebrates or in the common ancestor of jawed vertebrates and have diverged beyond recognition in teleost fish.

Previously, comparative analysis of human *GLI2* containing locus with orthologous loci from teleost fish like fugu and zebrafish has identified five intronic CNEs with tissue specific regulatory potential in transgenic zebrafish embryos.[28] Given the fact that a large majority of gene regulatory elements that might have evolved in the common ancestor of jawed vertebrates have largely been lost or diverged beyond recognition in teleost fish, our previous work could have missed many ancestral CNEs in human *GLI2* locus. Therefore, in the present study we again perform the comparisons of human (other tetrapods) and teleosts *Gli2* orthologous loci together with comparatively slowly evolving loci of elephant shark and spotted gar. These human centric comparative data identified 11 CNEs, six of them are novel that were not detected in previously published direct human-teleost alignments of *Gli2* locus.[28]These data signifies the importance of slowly evolving fish genomes in the identification of gnathostome CNEs (conserved among bony vertebrates and elephant shark) and bony vertebrates CNEs (missing from elephant shark). Furthermore, it appears that majority of CNEs in and around vertebrate *Gli2* containing locus are evolved in the gnathostome ancestral genome (Figure 1 and Table 1).

The majority of novel CNEs identified in the present study reside in immediate downstream intervals of human *GLI2*, whereas only one CNE is located within first intronic interval of this gene (Figure 1 and Table 1). The blast-based similarity search of predicted CNEs validated the uniqueness of these sequences to the human *GLI2* loci as no obvious paralogous copies are found in *GLI1* and *GLI3* containing loci or elsewhere in the human genome. In silico analysis reveals multiple

conserved TFBS for various developmental regulators that are known to be coexpressed with Gli2 during embryonic patterning (Table 1). To evaluate the gene regulatory potential of selected subset of Gli2-associated CNEs, the present study employed the transgenic zebrafish assay and RNA in situ hybridization of *gli2a* in zebrafish.

Anterio-posterior patterning of limbs requires precise temporal and spatial expression of *GLI2* [14]. RNA in situ hybridization studies in mice have shown that *Gli2* is broadly expressed in undifferentiated mesenchyme of the early limb bud.[20,21] *Gli2* knockout mice have reduced ossification of limb bones including stylopod, autopod, and zeugopod.[20] In addition, limb disorders like postaxial and preaxial polydactyly have been associated with *GLI2* mutations.[22,55] In accordance with reported functional activity/expression patterns of *Gli2* in mice, the present study reports the widespread presence of *gli2a* transcript in developing pectoral fin of zebrafish at 48 hpf (Figure 6). Consistent with widespread endogenous expression of *gli2a* in pectoral fin, our transgenic analysis revealed that subset of CNEs (CNE6, CNE7, and CNE8) drove the reporter gene expression in developing pectoral fin of zebrafish embryos (Figure 3). Therefore, taking into consideration previously reported elements, in total five *cis*-acting enhancer regions contributes to expression of *Gli2* in developing paired appendages of vertebrates (Figure 7). This apparent redundancy in reporter gene expression derived by *GLI2*-associated CNEs reflects the complex role of *cis*-acting regulatory repertoire in limb and fin morphogenesis. Similar functional redundancy in limb specific *cis*-regulators has been reported previously for other crucial limb developmental genes.[56]

Gli2 is also known to play multiple essential roles in development and morphogenesis of vertebrate CNS. Functional data from knockout mice have revealed defects in patterning of hindbrain and spinal cord with severely affected floorplate.[15,57] In harmony with known roles of Gli2 in vertebrate CNS development, the RNA in situ hybridization analysis of *gli2a* in the present study revealed robust expression of *gli2a* in developing CNS of zebrafish at 24 and 48 hpf (Figure 6). Consistent with endogenous expression patterns of *gli2a* activity, transgenic analysis of two novel *Gli2*-associated CNEs (CNE10 and CNE11) indicated their ability to up-regulate reporter gene expression in various domains of CNS (Figure 4). For instance, human CNE10 induced the *GFP* expression exclusively in forebrain, whereas CNE11 induced reporter activity was observed in hindbrain and spinal cord (Figure 4). Taken together with previously reported *GLI2*-associated CNE-enhancers, it appears that in total five distinct *cis*-acting regions in and around *GLI2* locus regulate its expression in developing CNS (Figure 7).[28]

In addition to limbs and CNS, Gli2 is known to play a prominent role in the developmental patterning of heart in mammals.[58] Similarly, in the present study, riboprobe-based endogenous expression pattern analysis of *gli2a*, revealed its strong presence in developing heart of zebrafish at 24 and 48 hpf (Figure 6). Among the *GLI2*-associated CNEs identified in the current investigation, CNE6 and CNE9 drove the reporter gene expression prominently in the developing heart of zebrafish at 24 and 48 hpf (Figure 5). Intriguingly, these two CNEs

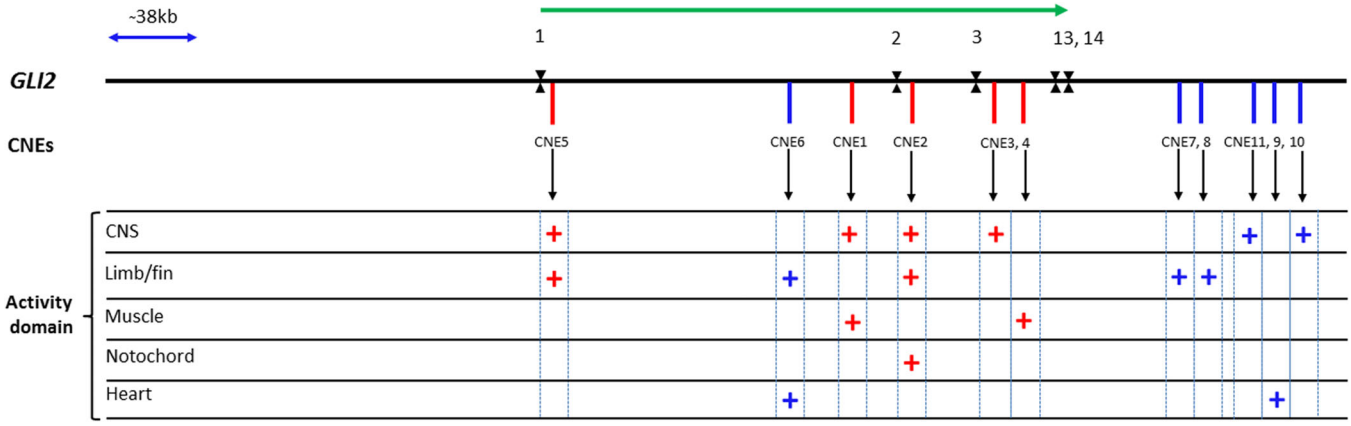

FIGURE 7 Schematic overview of the inventory of *cis*-regulatory elements regulating *GLI2* expression in various developmental domains. Schematic view of human *GLI2*, encompassing 100 kb flanking region from both directions showing exons in the black collate and conserved CNEs in red and blue lines. CNEs represented in red lines were reported previously[28] whereas, those presented as blue vertical lines are detected in the present study. The black arrow underneath the CNE elements points to their activity in respective tissue domain. CNEs drove the reporter gene expression in various domains (including heart, fin, and CNS) are marked as plus sign (+). Genic architecture is drawn according to scale. Green arrowhead depicts the direction of *GLI2* transcription. Ex, exon; kb, kilobase

share multiple anciently conserved binding motifs for transcription factors that are known to coexpress with *Gli2* in mouse heart (Table 1).

4 | CONCLUSION

The present study employs the relatively slowly evolving and unduplicated genomes of basal, nonteleost ray-finned fish and a cartilaginous fish to discover six novel *GLI2*-associated CNEs that were not predicted by direct human-teleost comparisons. Majority of CNEs in and around vertebrate *Gli2* containing locus arose in the gnathostome ancestral genome, evidenced by human-centric alignments for seven vertebrates (including gar, elephant shark, one teleost fish, one lobe-finned fish, and three tetrapods). Functional analyses of these anciently conserved *GLI2*-associated CNEs through transgenic zebrafish assay revealed their redundant ability to drive expression of reporter gene in known sites of endogenous Gli2/gli2 activity, including paired appendages, CNS and heart. Description of a catalog for *GLI2* specific *cis*-regulatory genomic regions offers a new perspective for investigating the novel molecular mechanisms by which the downstream effectors of Hh signaling pathway might themselves be deployed at proper place and precise time to direct pattern formation during embryogenesis. For instance, inventory of *Gli* enhancers can be used to elucidate the *cis*-regulatory basis of the complementary roles of *Gli2* and *Gli3* during the patterning of neural tube, limbs, internal organs and the bone development and repair. Identifying the code of *trans*-acting molecules that confers the tissue specificity on *Gli* associated enhancers may help to understand the molecular mechanisms by which a proper balance between Hh and Gli transcripts is established in complementary developmental domains of the limbs and CNS. In humans, functional aberrations of *GLI2* are associated with holoprosencephaly, cranial and midline facial deformities, abnormalities in limb development (preaxial and postaxial polydactyly), congenital growth hormone deficiency, pituitary anomalies (hypopituitarism), and carcinogenesis. Mutations in enhancer modules directing *GLI2* expression in affected tissue domains can potentially affect the normal availability of its transcript during embryogenesis. In this case, though the gene body and the protein product is expected to normal, DNA variants within concerned enhancer elements can cause abnormal phenotypes through anomalous gene expression. Therefore, *cis*-regulatory inventory of *GLI2* provides a novel target for clinical diagnostic sequencing to determine the genetic etiology of many cases of *GLI2* associated congenital anomalies which cannot be attributed to protein coding sequence of this gene.

5 | EXPERIMENTAL PROCEDURES

5.1 | Multispecies comparative analysis and transcription factor binding motifs searching

Human *GLI2* sequence along with its orthologous counterparts from mouse, chicken, lizard, coelacanth, fugu, spotted gar and elephant shark was obtained from Ensemble Genome Browser (https://asia.ensembl.org). These orthologous sequences were submitted to Shuffle Lagan tool for multispecies comparative analysis.[59] Human sequence was kept as baseline. The Shuffle lagan alignment results were visualized by the vista visualization tool.[60] The selection criteria for the identification of CNEs were defined as 50% identity over 50 bp sequence length down to fish.

The orthologous sequences of CNEs were subjected to the MEME tool for putative transcription factor binding motifs discovery.[61] The MEME program is based on a position weight matrices algorithm that scans the subjected sequences for over-represented motifs in the query dataset. The minimum size of TFBS was set to 6 to 12 bp. The MEME output files for predicted TFBSs were further proceeded with the STAMP program to find known transcription factors from TRNASFAC library.[62]

5.2 | Genomics and epigenomics analysis

Association of selected CNEs with their target gene is based on two filters such as comparative genomics and endogenous expression pattern analysis.[35,36] For this purpose human synteny map containing predicted CNEs loci was drafted using UCSC (http://genome.ucsc.edu/) and Ensemble Genome Browser (https://asia.ensembl.org). These maps depict flanking genes located at a radius of 1 Mb. Through multispecies sequence alignment, the conservation depth of CNEs was estimated (Table 1). The sequence conservation of CNEs between human and distantly related vertebrate species implies functional conservation. It would be appropriate then to speculate that the target gene would also be the same between compared species.[35,36] Under these assumptions, a comparative picture of CNEs bearing human synteny map was comparatively analyzed in mice and spotted gar using multispecies view option in the Ensembl genome browser. Evolutionarily conserved physical linkage between CNE elements and one or more neighboring genes was taken as an evidence of functional association. To further confirm these associations, for one or more genes depicting evolutionary conserved physical association with selected CNEs, the RNA in situ hybridization-

based endogenous expression pattern of mouse orthologs was obtained from MGI (http://www.informatics.jax.org/).

The 3D genome browser provides a Virtual 4C tool to analyze spatial interactions between a pair of genomic intervals.[63] We used Hi-C derived Virtual 4C data from H1 human embryonic stem cell lines, mesenchymal, fetal brain, and cardiac cell lines (H1-ESC; Dixon2015-raw).[40] The genomic coordinates from hg19 assembly of *GLI2* loci was subjected for finding the physical binding between promoter of *GLI2* and predicted CNE elements. The schematics depicting enhancer regions interacting with *GLI2* promoter was drawn manually (Figure 2B).

Further, presence of three types of histone modification marks; H3K27ac, H3K4me1, and H3k4me2 was investigated in selected CNEs. For this purpose, histone modification ChIP-seq datasets for seven human embryonic stem cell lines were fetched from Encode experimental matrix and overlapped with genomic coordinates of respective CNEs using Bedtools (v2.17.0) https://genome.ucsc.edu/encode/dataMatrix/encodeChipMatrixHuman.html). Moreover, encode dataset of H3K4me1 and H3K27ac for human embryonic stem cell lines; human fetal leg muscles, mouse ES cells, limbs, neural tube, and embryonic facial prominence was overlapped with *GLI2*/Gli2 genomic coordinates.

5.3 | Zebrafish transgenic enhancer assay

Human genomic DNA was purified from whole blood by phenol chloroform method. The shortlisted CNE elements were amplified from human genomic DNA using primers listed in Additional file 2: Table S1. The amplified products (~500 ng/μL) were directly ligated with PCR8/GW/TOPO vector to prepare the entry clone by following the manufacturer instructions (Invitrogen, Life Technologies). The orientation of the insert (CNE elements) was confirmed by Sanger sequencing. The CNE fragments were then moved to destination vector pGW-*cfos-EGPF* via gateway cloning system. Using LR clonase enzyme; LR (attL and attR) recombination reaction between entry clone (150 ng/μL) and destination vector (150 ng/μL) were performed. The resulted destination clones were confirmed by restriction digestion and orientation of insert was checked by Sanger sequencing.

Zebrafish was bred and raised according to the standard procedures. Zebrafish fertilized eggs were collected from natural spawning of the wild type fish. Transposase encoding mRNA was synthesized by in vitro transcription from linearized pCS-TP plasmid, using SP6 mMessage mMachine kit (Ambian). The synthesized mRNA was precipitated in molecular grade 100% ethanol and lithium chloride, followed by phenol-chloroform purification. Protocol devised by Fisher et al 2006 was followed to make injection solution containing 1 μL purified reporter vector (pGW-CNE-*cfos*-EGFP) (125 ng/μL), 0.5 μL Transposase mRNA (350 ng/μL), 0.5 μL phenol red and volume was raised to 5 μL by adding nuclease free water.[41] Approximately 2 nL 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 embryos raising media containing 0.003% phenylthiourea.

The injected embryos were raised up to 24 hpf, then dechorionated manually and anesthetized in tricaine. The embryos were screened for reporter gene expression (*GFP*) under fluorescent inverted microscope IX71 (Olympus, Japan). The images of live transgenic embryos were captured by DP72 camera using monochrome software. Elements are annotated based on the observed spatial expression patterns. To be defined as positive enhancer, an element has to show reproducible expression in the same developmental domain in at least 25% of transgenic embryos.

5.4 | Whole mount in situ hybridization

To describe the endogenous expression patterns, in situ hybridization for *gli2a* gene were carried out following standard protocol.[64] Briefly, primers were designed to amplify a coding region of *gli2a* from zebrafish wild type cDNA. To synthesize antisense probe, the PCR amplified DNA fragment was cloned in PCRII Topo vector. The RNA probe was synthesized by in vitro transcription with the mMessage mMachine kit using SP6 polymerase. Zebrafish dechorionated embryos at 24 hpf and 48 to 50 hpf were fixed overnight in 4% paraformaldehyde at 4°C. Embryos were overnight hybridized at 70°C with *gli2a* antisense probe in a 1000 μL hybridization mix containing 250 ng probe concentration. On the following day, embryos were washed followed by blocking at room temperature for 1 hour. Next, embryos were overnight incubated in 1000 μL blocking solution containing 1 μL anti DIG antibody at 4°C with gentle agitation. Next day, embryos were washed in NTMT buffer and stained in purple AP substrate. The color reaction was periodically monitored under phase contrast microscope. The stained samples were visualized by Leica M20FA microscope.

ACKNOWLEDGMENTS

The authors are thankful to Higher Education Commission of Pakistan for NRPU grant (No. 20-2085/NRPU/R&D/HEC/12/760). The authors are also grateful to Dr Tetsuya Nakamura Rutgers University and Dr Thomas

A. Stewart University of Chicago for useful discussion and suggestions. S. A holds a fellowship under International Research Support Initiative Program (IRSIP) of Higher Education Commission of Pakistan. N. S. was supported by a grant from the Brinson Foundation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Shahid Ali: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; validation; visualization; writing-original draft; writing-review and editing. **Irum Arif**: Methodology; software. **Ayesha Iqbal**: Methodology; software. **Irfan Hussain**: Data curation; formal analysis; methodology; software. **Muhammad Ramzan Abrar**: Methodology; software. **Muhammad Khan**: Investigation; methodology; resources. **Neil Shubin**: Conceptualization; data curation; formal analysis; funding acquisition; investigation; project administration; resources; supervision; validation; writing-original draft; writing-review and editing. **Amir Ali Abbasi**: Conceptualization; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing-original draft; writing-review and editing.

ORCID

Amir Ali Abbasi https://orcid.org/0000-0003-4556-8129

REFERENCES

1. Choudhry Z, Rikani AA, Choudhry AM, et al. Sonic hedgehog signalling pathway: a complex network. *Ann Neurosci*. 2014;21(1):28-31.
2. Anderson E, Peluso S, Lettice LA, Hill RE. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet*. 2012;28(8):364-373.
3. Huizar F, Soundarrajan D, Paravitorghabeh R, Zartman J. Interplay between morphogen-directed positional information systems and physiological signaling. *Dev Dyn*. 2020;249(3):328-341.
4. Abbasi AA, Goode DK, Amir S, Grzeschik K-H. Evolution and functional diversification of the GLI family of transcription factors in vertebrates. *Evol Bioinform Online*. 2009;5:S2322.
5. Niewiadomski P, Niedziółka SM, Markiewicz Ł, Uśpieński T, Baran B, Chojnowska K. Gli proteins: regulation in development and cancer. *Cell*. 2019;8(2):147.
6. Kinzler KW, Ruppert JM, Bigner SH, Vogelstein B. The GLI gene is a member of the Kruppel family of zinc finger proteins. *Nature*. 1988;332(6162):371-374.
7. Aza-Blanc P, Kornberg TB. Ci: a complex transducer of the hedgehog signal. *Trends Genet*. 1999;15(11):458-462.
8. Ke Z, Kondrichin I, Gong Z, Korzh V. Combined activity of the two Gli2 genes of zebrafish play a major role in Hedgehog signaling during zebrafish neurodevelopment. *Mol Cell Neurosci*. 2008;37(2):388-401.
9. Ke Z, Emelyanov A, Lim SES, Korzh V, Gong Z. Expression of a novel zebrafish zinc finger gene, gli2b, is affected in Hedgehog and Notch signaling related mutants during embryonic development. *Dev Dyn*. 2005;232(2):479-486.
10. Huangfu D, Anderson KV. Signaling from Smo to Ci/Gli: conservation and divergence of Hedgehog pathways from Drosophila to vertebrates. *Development*. 2006;133(1):3-14.
11. Bai CB, Auerbach W, Lee JS, Stephen D, Joyner AL. Gli2, but not Gli1, is required for initial Shh signaling and ectopic activation of the Shh pathway. *Development*. 2002;129(20):4753-4761.
12. Abbasi AA, Paparidis Z, Malik S, et al. Human GLI3 intragenic conserved non-coding sequences are tissue-specific enhancers. *PLoS One*. 2007;2(4):e366.
13. i Altaba AR, Mas C, Stecca B. The Gli code: an information nexus regulating cell fate, stemness and cancer. *Trends Cell Biol*. 2007;17(9):438-447.
14. Bowers M, Eng L, Lao Z, et al. Limb anterior–posterior polarity integrates activator and repressor functions of GLI2 as well as GLI3. *Dev Biol*. 2012;370(1):110-124.
15. Lebel M, Mo R, Shimamura K, C-c H. Gli2 and Gli3 play distinct roles in the dorsoventral patterning of the mouse hindbrain. *Dev Biol*. 2007;302(1):345-355.
16. McDermott A, Gustafsson M, Elsam T, Hui C-C, Emerson CP, Borycki A-G. Gli2 and Gli3 have redundant and context-dependent function in skeletal muscle formation. *Development*. 2005;132(2):345-357.
17. Onimaru K, Kuraku S, Takagi W, Hyodo S, Sharpe J, Tanaka M. A shift in anterior–posterior positional information underlies the fin-to-limb evolution. *Elife*. 2015;4:e07048.
18. Sasaki H, Nishizaki Y, Hui C-C, Nakafuku M, Kondoh H. Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling. *Development*. 1999;126(17):3915-3924.
19. Veistinen L, Takatalo M, Kesper DA, Vortkamp A, Rice DP. Deletion of Gli3 in mice causes abnormal frontal bone morphology and premature synostosis of the interfrontal suture. *Front Physiol*. 2012;3:121.
20. Mo R, Freer AM, Zinyk DL, et al. Specific and redundant functions of Gli2 and Gli3 zinc finger genes in skeletal patterning and development. *Development*. 1997;124(1):113-123.
21. Bai CB, Joyner AL. Gli1 can rescue the in vivo function of Gli2. *Development*. 2001;128(24):5161-5172.
22. Bertolacini CDP, Ribeiro-Bicudo LA, Petrin A, Richieri-Costa A, Murray JC. Clinical findings in patients with GLI2 mutations–phenotypic variability. *Clin Genet*. 2012;81(1):70-75.
23. Cheng A-L, Kang Y-K, Chen Z, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol*. 2009;10(1):25-34.
24. Fulda S, Wick W, Weller M, Debatin K-M. Smac agonists sensitize for Apo2L/TRAIL-or anticancer drug-induced apoptosis and induce regression of malignant glioma in vivo. *Nat Med*. 2002;8(8):808-815.
25. Grachtchouk M, Mo R, Yu S, et al. Basal cell carcinomas in mice overexpressing Gli2 in skin. *Nat Genet*. 2000;24(3):216-217.

26. Sicklick JK, Li Y-X, Choi SS, et al. Role for hedgehog signaling in hepatic stellate cell activation and viability. *Lab Invest*. 2005; 85(11):1368-1380.
27. Karlstrom RO, Tyurina OV, Kawakami A, et al. Genetic analysis of zebrafish gli1 and gli2 reveals divergent requirements for gli genes in vertebrate development. *Development*. 2003;130(8): 1549-1564.
28. Minhas R, Pauls S, Ali S, et al. Cis-regulatory control of human GLI2 expression in the developing neural tube and limb bud. *Dev Dyn*. 2015;244(5):681-692.
29. Abbasi AA, Paparidis Z, Malik S, et al. Human intronic enhancers control distinct sub-domains of Gli3 expression during mouse CNS and limb development. *BMC Dev Biol*. 2010;10 (1):44.
30. Ravi V, Venkatesh B. The divergent genomes of teleosts. *Annu Rev Anim Biosci*. 2018;6:47-68.
31. Venkatesh B, Kirkness EF, Loh Y-H, et al. Ancient noncoding elements conserved in the human genome. *Science*. 2006;314 (5807):1892-1892.
32. Amemiya CT, Alföldi J, Lee AP, et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature*. 2013; 496(7445):311-316.
33. Braasch I, Gehrke AR, Smith JJ, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 2016;48(4):427-437.
34. Consortium EP. The ENCODE (ENCyclopedia of DNA elements) project. *Science*. 2004;306(5696):636-640.
35. Parveen N, Masood A, Iftikhar N, et al. Comparative genomics using teleost fish helps to systematically identify target gene bodies of functionally defined human enhancers. *BMC Genomics*. 2013;14(1):122.
36. Ali S, Amina B, Anwar S, et al. Genomic features of human limb specific enhancers. *Genomics*. 2016;108(3–4):143-150.
37. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. *Nat Rev Genet*. 2013;14(4):288-295.
38. Eres IE, Luo K, Hsiao CJ, Blake LE, Gilad Y. Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet*. 2019;15(7):e1008278.
39. Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev*. 2016;30 (12):1357-1382.
40. Dixon JR, Jung I, Selvaraj S, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518 (7539):331-336.
41. Fisher S, Grice EA, Vinton RM, et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc*. 2006;1(3):1297-1305.
42. Coppola CJ, Ramaker RC, Mendenhall EM. Identification and function of enhancers in the human genome. *Hum Mol Genet*. 2016;25(R2):R190-R197.
43. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108-112.
44. Paparidis Z, Abbasi AA, Malik S, et al. Ultraconserved noncoding sequence element controls a subset of spatiotemporal GLI3 expression. *Dev Growth Differ*. 2007;49(6):543-553.
45. Parker HJ, Piccinelli P, Sauka-Spengler T, Bronner M, Elgar G. Ancient Pbx-Hox signatures define hundreds of vertebrate developmental enhancers. *BMC Genomics*. 2011; 12(1):637.
46. Woolfe A, Goodson M, Goode DK, et al. Highly conserved noncoding sequences are associated with vertebrate development. *PLoS Biol*. 2004;3(1):e7.
47. Gehrke AR, Schneider I, de la Calle-Mustienes E, et al. Deep conservation of wrist and digit enhancers in fish. *Proc Natl Acad Sci U S A*. 2015;112(3):803-808.
48. Gehrke AR, Shubin NH. Cis-regulatory programs in the development and evolution of vertebrate paired appendages. *Semin cell Dev Biol*. 2016;57(31-39).
49. Kim MJ, Oksenberg N, Hoffmann TJ, Vaisse C, Ahituv N. Functional characterization of SIM1-associated enhancers. *Hum Mol Genet*. 2014;23(7):1700-1708.
50. Hui CC, Angers S. Gli proteins in development and disease. *Annu Rev Cell Dev Biol*. 2011;27:513-537.
51. Boffelli D, Nobrega MA, Rubin EM. Comparative genomics at the vertebrate extremes. *Nat Rev Genet*. 2004;5(6):456-465.
52. Elgar G, Sandford R, Aparicio S, Macrae A, Venkatesh B, Brenner S. Small is beautiful: comparative genomics with the pufferfish (Fugu rubripes). *Trends Genet*. 1996;12(4):145-150.
53. Venkatesh B, Lee AP, Swann JB, et al. Venkatesh et al. reply. *Nature*. 2014;511(7508):E9-E10.
54. Yousaf A, Sohail Raza M, Ali Abbasi A. The evolution of bony vertebrate enhancers at odds with their coding sequence landscape. *Genome Biol Evol*. 2015;7(8):2333-2343.
55. França MM, Jorge AA, Carvalho LR, et al. Relatively high frequency of non-synonymous GLI 2 variants in patients with congenital hypopituitarism without holoprosencephaly. *Clin Endocrinol (Oxf)*. 2013;78(4):551-557.
56. Letelier J, de la Calle-Mustienes E, Pieretti J, et al. A conserved Shh cis-regulatory module highlights a common developmental origin of unpaired and paired fins. *Nat Genet*. 2018;50(4): 504-509.
57. Ding Q, Motoyama J, Gasca S, et al. Diminished Sonic hedgehog signaling and lack of floor plate differentiation in Gli2 mutant mice. *Development*. 1998;125(14):2533-2543.
58. Voronova A, al Madhoun A, Fischer A, Shelton M, Karamboulas C, Skerjanc IS. Gli2 and MEF2C activate each other's expression and function synergistically during cardiomyogenesis in vitro. *Nucleic Acids Res*. 2011;40(8):3329-3347.
59. Brudno M, Do CB, Cooper GM, et al. LAGAN and multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res*. 2003;13(4):721-731.
60. Mayor C, Brudno M, Schwartz JR, et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*. 2000;16(11):1046-1047.
61. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res*. 2006;34(suppl_2):W369-W373.
62. Mahony S, Benos PV. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res*. 2007;35(suppl 2): W253-W258.
63. Wang Y, Song F, Zhang B, et al. The 3D genome browser: a web-based browser for visualizing 3D genome organization

and long-range chromatin interactions. *Genome Biol*. 2018;19 (1):151.
64. Thisse C, Thisse B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc*. 2008;3(1):59-69.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ali S, Arif I, Iqbal A, et al. Comparative genomic analysis of human *GLI2* locus using slowly evolving fish revealed the ancestral gnathostome set of early developmental enhancers. *Developmental Dynamics*. 2021;250: 669–683. https://doi.org/10.1002/dvdy.291