Coding exons function as tissue-specific enhancers of nearby genes

Ramon Y. Birnbaum, E. Josephine Clowney, Orly Agamy, et al.

Genome Res. 2012 22: 1059-1068 originally published online March 22, 2012
Access the most recent version at doi:10.1101/gr.133546.111

Coding exons function as tissue-specific enhancers of nearby genes

Ramon Y. Birnbaum,[1,2] E. Josephine Clowney,[3,4] Orly Agamy,[5] Mee J. Kim,[1,2] Jingjing Zhao,[1,2,6] Takayuki Yamanaka,[1,2] Zachary Pappalardo,[1,2] Shoa L. Clarke,[7] Aaron M. Wenger,[8] Loan Nguyen,[1,2] Fiorella Gurrieri,[9] David B. Everman,[10] Charles E. Schwartz,[10,11] Ohad S. Birk,[5] Gill Bejerano,[8,12] Stavros Lomvardas,[3] and Nadav Ahituv[1,2,13]

[1]Department of Bioengineering and Therapeutic Sciences, [2]Institute for Human Genetics, [3]Department of Anatomy, [4]Program in Biomedical Sciences, University of California, San Francisco, California 94143, USA; [5]The Morris Kahn Laboratory of Human Genetics, NIBN, Ben-Gurion University, Beer-Sheva 84105, Israel; [6]Key Laboratory of Advanced Control and Optimization for Chemical Processes of the Ministry of Education, East China University of Science and Technology, Shanghai 200237, China; [7]Department of Genetics, [8]Department of Computer Science, Stanford University, Stanford, California 94305-5329, USA; [9]Istituto di Genetica Medica, Università Cattolica S. Cuore, Rome 00168, Italy; [10]JC Self Research Institute, Greenwood Genetic Center, Greenwood, South Carolina 29646, USA; [11]Department of Genetics and Biochemistry, Clemson University, Clemson, South Carolina 29634, USA; [12]Department of Developmental Biology, Stanford University, Stanford, California 94305-5329, USA

Enhancers are essential gene regulatory elements whose alteration can lead to morphological differences between species, developmental abnormalities, and human disease. Current strategies to identify enhancers focus primarily on noncoding sequences and tend to exclude protein coding sequences. Here, we analyzed 25 available ChIP-seq data sets that identify enhancers in an unbiased manner (H3K4me1, H3K27ac, and EP300) for peaks that overlap exons. We find that, on average, 7% of all ChIP-seq peaks overlap coding exons (after excluding for peaks that overlap with first exons). By using mouse and zebrafish enhancer assays, we demonstrate that several of these exonic enhancer (eExons) candidates can function as enhancers of their neighboring genes and that the exonic sequence is necessary for enhancer activity. Using ChIP, 3C, and DNA FISH, we further show that one of these exonic limb enhancers, *Dynclil* exon 15, has active enhancer marks and physically interacts with *Dlx5/6* promoter regions 900 kb away. In addition, its removal by chromosomal abnormalities in humans could cause split hand and foot malformation 1 (SHFM1), a disorder associated with *DLX5/6*. These results demonstrate that DNA sequences can have a dual function, operating as coding exons in one tissue and enhancers of nearby gene(s) in another tissue, suggesting that phenotypes resulting from coding mutations could be caused not only by protein alteration but also by disrupting the regulation of another gene.

[Supplemental material is available for this article.]

Precise temporal, spatial, and quantitative regulation of gene expression is essential for proper development. This tight transcriptional regulation is mediated in part by DNA sequences called enhancers, which regulate gene promoters. By use of comparative genomics or chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq), candidate enhancer sequences can now be identified in a relatively high-throughput manner (Heintzman and Ren 2009; Visel et al. 2009b). These sequences can then be assayed for enhancer activity using various in vitro and in vivo assays (Woolfe et al. 2005; Pennacchio et al. 2006; Heintzman et al. 2009). However, the majority of these experiments remove coding sequences from their analyses under the assumption that they do not function as enhancers, due to their protein coding role.

Previous exonic enhancers (eExons) have been reported in vertebrates (Neznanov et al. 1997; Lampe et al. 2008; Tumpel et al. 2008; Dong et al. 2010; Eichenlaub and Ettwiller 2011; Ritter et al. 2012). In addition, a recent study scanning for synonymous constraint in protein coding regions (Lin et al. 2011) found an overlap between two of these eExons (Lampe et al. 2008; Tumpel et al. 2008) and synonymous constraint elements. Here, we analyzed 25 available ChIP-seq data sets of enhancer marks (H3K4me1, H3K27ac, and EP300, also known as p300) for their overlap with coding exons. Following this analysis, we wanted to specifically determine whether eExons could regulate their neighboring genes and not the gene they reside in. This was of interest to us due to the phenotypic implications that coding mutations could have on nearby genes. For this purpose, we analyzed a specific EP300 ChIP-seq data set from mouse embryonic day (E) 11.5 limb tissue (Visel et al. 2009a), due to its ability to identify active enhancers with high accuracy (88%) and tissue specificity (80%) in vivo. At E11.5, mouse limb development progresses along three axes: proximal-distal (P-D), anterior-posterior (A-P), and dorsal-ventral (D-V). Specific signaling centers in the limb bud create gradients and feedback loops that determine these axes (Gilbert 2000; Nissim and Tabin 2004; Zeller et al. 2009), and their alteration could lead to morphological differences. In this study, we focused on identifying limb eExons involved in the development along both the P-D and the A-P axes.

[13]Corresponding author.
E-mail nadav.ahituv@ucsf.edu.
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.133546.111.

The apical ectodermal ridge (AER) is the signaling center that keeps the underlying mesenchyme in a proliferative state and allows the limb to grow, thus governing the P-D axis. In the developing mouse limb bud, the distal-less homeobox 5 and 6 (*Dlx5/6*) genes are expressed in the AER (Fig. 1D′,E′), and *Dlx5* is also expressed in the anterior mesenchyme (Fig. 1E′). Disruption of both *Dlx5* and *Dlx6* in mice leads to a split hand and foot malformation (SHFM) phenocopy (Robledo et al. 2002). In humans, chromosomal aberrations in the *DLX5/6* region, some of which do not encompass the coding sequences of *DLX5/6*, cause SHFM1 (MIM 183600) and are associated with incomplete penetrance with intellectual disability, craniofacial malformations and deafness (Elliott and Evans 2006). Other than one family with a *DLX5* missense mutation (Shamseldin et al. 2012), no other coding mutation in either gene has been found in individuals with SHFM1, suggesting that disruption of *DLX5/6* gene regulatory elements could lead to a SHFM1 phenotype.

The A-P axis is controlled by a signaling center called the zone of polarizing activity, located in the posterior mesenchyme and defined by the expression of Sonic Hedgehog (*SHH*). *Twist1*, is a transcription factor that inhibits *Shh* expression in the anterior limb bud by antagonizing *HAND2*, a *Shh*-positive regulator (Firulli et al. 2005). Homozygous *Twist1*-null mice have limb bud developmental defects (Chen and Behringer 1995), and heterozygous mice develop polydactyly (Bourgeois et al. 1998) and show ectopic *Shh* expression (O'Rourke et al. 2002). In humans, mutations in *TWIST1* lead to various syndromes, the majority of which encompass various forms of limb malformations (MIM*601622).

Here, by examining various enhancer-associated ChIP-seq data sets, we characterized the general prevalence of peaks that overlap exons. We then chose seven limb EP300 ChIP-seq exonic sequences and functionally tested them for enhancer activity using a mouse transgenic enhancer assay. Four out of seven tested exonic sequences were shown to be functional limb enhancers in the mouse. Further analysis of one of these enhancers, *Dync1i1* eExon 15, using chromatin conformation capture (3C) and DNA fluorescent in situ hybridization (FISH) showed that it physically interacts with *Dlx5/6* promoter regions in the developing mouse limb. Mutation analysis of individuals with SHFM1 indicated that chromosomal aberrations encompassing this enhancer could be one of the causes of their SHFM1 phenotype. Combined, these findings demonstrate that a DNA sequence can function both as a coding exon in one tissue and as an enhancer in a different tissue and suggest the need to be cautious when assigning a coding mutation phenotype to protein function.

Results

Exon overlap analyses of enhancer-associated ChIP-seq data sets

To determine the genome-wide prevalence of enhancer-associated ChIP-seq peaks that overlap coding exons, we analyzed 25 available ChIP-seq data sets of enhancer marks (H3K4me1, H3K27ac, and EP300) from various human cell lines and mouse E11.5 tissues (Supplemental Table 1; see Methods). Since these enhancer marks could also identify potential promoters, we only looked for overlap with coding exons after excluding for the first exon. In all the analyses described in this study, coding exons are defined here as only those exons that are not first exons. Analysis of the individual histone marks, H3K4me1 and H3K27ac, showed that 7% and 10% of all peaks overlap coding exons, respectively (Supplemental Table 1). It is worth noting that the average peak size in the H3K4me1 and H3K27ac ChIP-seq data sets is 2441 and 3107 bp, and the average size of peaks overlapping coding exons is 3476 and 4195 bp for each mark. Analysis of the average exon size in the human genome shows that it is ~280 bp (see Methods). Compared with the average peak size of the histone marks and those that overlap exons in particular, it is quite possible that the functional entity of the peak does not constitute the exon and that the percentages above are an overestimate. Therefore, we analyzed six different EP300 ChIP-seq data sets from various human cell lines that have shorter peak sizes, the average of which was 426 bp. In these data sets, we found that on average, 4% of the peaks overlapped with coding exons (Supplemental Table 1). To get a better indication if these sequences could be functional enhancers, we examined the overlap between exonic EP300 ChIP-seq peaks and H3K4me1 and H3K27ac peaks. We used ChIP-seq data from two different cell lines, GM12878 and K562, where all three enhancer marks were available. We found that 8% and 5% of the ChIP-seq peaks that had all three enhancer marks overlapped coding exons in GM12878 and K562 cells, respectively (Supplemental Table 1). We next screened coding sequences that had all three enhancer marks for their overlap with a recently published study that scanned the genome for synonymous constraint in protein coding regions (Lin et al. 2011). We found that 9% of coding exons with all three ChIP-seq enhancer marks overlapped with synonymous constraint exons for both GM12878 and K562 cell lines (Sup-

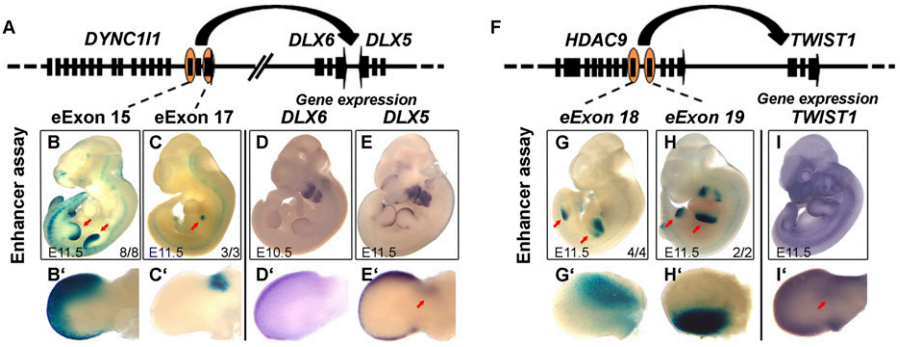

Figure 1. eExons within *DYNC1I1* and *HDAC9* characterized using a mouse transgenic enhancer assay. A schematic of the *DYNC1I1-DLX5/6* (*A*) and *HDAC9-TWIST1* (*F*) genomic regions. Black boxes and orange ovals represent coding exons and positive eExons, respectively. The black arrows point to the genes that are thought to be regulated by the eExons. (*B*–C′) Mouse enhancer assays of *DYNC1I1* eExon 15 and 17 at embryonic day 11.5 (E11.5). (*B*,*B*′) *DYNC1I1* eExon 15 shows apical ectodermal ridge (AER) and limb mesenchyme enhancer activity (red arrows), and (*C*,*C*′) *DYNC1I1*eExon 17 shows anterior limb bud mesenchyme enhancer activity (red arrow). (*D*,*E*) Mouse whole-mount in situ hybridization of *Dlx6* and *Dlx5*. (*D*′,*E*′) *Dlx6* and *Dlx5* limb expression pattern is similar to *DYNC1I1* eExon 15 enhancer activity. In addition, *Dlx5* is also expressed in anterior limb bud as depicted by the red arrow (*E*′), similar to *DYNC1I1* eExon 17 enhancer activity (*C*′). (*G*–*H*′) Mouse enhancer assays of *HDAC9* eExons 18 and 19. (*G*,*G*′) *HDAC9* eExon 18 shows anterior limb bud enhancer activity (red arrows), and (*H*,*H*′) *HDAC9* eExon 19 shows posterior limb bud (red arrows) and branchial arch enhancer activity in E11.5 mice. (*I*) Mouse whole-mount in situ hybridization of *Twist1* at E11.5. (*I*′) *Twist1* limb expression pattern is similar to the *HDAC9* eExon 18 anterior limb bud enhancer activity (*G*′) marked by red arrow and *HDAC9* eExon 19 posterior limb bud enhancer activity (*H*′). For *B, C, G*, and *H,* the numbers in the *bottom right* corner indicate the number of embryos showing this limb expression pattern/total LacZ stained embryos.

plemental Table 2). In summary, we found that on average, 7% of the peaks in the 25 enhancer-associated ChIP-seq data sets that we analyzed overlapped coding exons after removing first exons.

Analysis of a EP300 limb ChIP-seq data set and eExon candidate selection

Given that several eExons were previously discovered to regulate the gene they reside in (Neznanov et al. 1997; Lampe et al. 2008; Tumpel et al. 2008; Ritter et al. 2012), we explicitly set out to search for coding eExons that do not autoregulate but rather could regulate their nearby genes. This was of interest to us due to the phenotypic consequences that coding mutations could have on their nearby genes. In order to do this, we needed a tissue-specific ChIP-seq enhancer data set. We thus focused our analysis on EP300 ChIP-seq data sets that were shown to predict functional enhancers in three different mouse E11.5 tissues (forebrain, midbrain, limb) with high accuracy and tissue specificity (Visel et al. 2009a). In this data set, we observed that 4% of EP300 ChIP-seq peaks from all three tissues overlap with coding exons after excluding the first exon (Supplemental Table 1). These lower percentages could be due to experimental differences such as cell line versus tissue. For our functional assays, we next focused on the limb EP300 ChIP-seq data set. We scanned this data set for exonic sequences that reside in genes that are not known to be expressed in the limb but are located in the vicinity (up to 1 Mb on either side) of known limb-associated genes (see Methods). From the 252 limb EP300 ChIP-seq peaks that overlaid exons, 152 sequences overlapped coding exons and 134 were in a gene that is not expressed in the limb (Supplemental Table 3). Out of those 134 sequences, 90 had at least one limb expressed gene up to 1 Mb away on either side of the gene. We chose seven exons near important limb developmental genes (*C14orf49* exon 16 [near *DICER1*], *CDC14B* exon 13 [near *PTCH1*], *DYNC1I1* exon 15 and exon 17 [near *DLX5/6*], *HDAC9* exon 18 and exon 19 [near *TWIST1*], *STX18* exons 4–5 [near *MSX1*]) for subsequent mouse enhancer assays (Supplemental Table 3).

Mouse enhancer assays

To test whether these exonic sequences function as enhancers, we tested all seven sequences for their enhancer activity in mice. The human sequences were cloned into the *Hsp68*-LacZ vector that contains the heat shock protein 68 minimal promoter followed by a LacZ reporter gene (Kothary et al. 1988). Transgenic mice were generated, and embryos were harvested at E11.5 and stained for LacZ. We found that four out of the seven sequences showed limb enhancer activity in mice. *DYNC1I1* eExon 15 drove specific LacZ expression in the limb mesenchyme and AER (Fig. 1B,B′; Supplemental Fig. 1A), and *DYNC1I1* eExon 17 drove specific LacZ expression in the anterior limb mesenchyme (Fig. 1C,C′; Supplemental Fig. 1C). *HDAC9* eExon 18 showed enhancer activity in the anterior limb bud (Fig. 1G,G′; Supplemental Fig. 2A) and *HDAC9* eExon 19 in the posterior limb bud (Fig. 1H,H′; Supplemental Fig. 2B).

The exonic sequence is necessary for enhancer activity

The sequences tested in the mouse enhancer assays had some intronic regions due to the ChIP-seq peak overlapping part of the intron (Supplemental Table 3). We thus wanted to assess whether the exonic sequence is necessary for enhancer activity. Since human limb and zebrafish fin development are considered highly comparable on the molecular level (Hall 2007; Iovine 2007; Mercader 2007) and since zebrafish enhancer assays are rapid and cost-efficient, we carried out a deletion series analyses using this assay. We first characterized whether our functional mouse limb enhancers were positive for fin enhancer expression in zebrafish. The four limb enhancers (Supplemental Table 3) were cloned from human genomic DNA into a zebrafish enhancer assay vector, containing an *E1b* minimal promoter followed by the green fluorescent protein (*GFP*) reporter gene (Li et al. 2009). These vectors were microinjected into one-cell-stage zebrafish embryos along with the *Tol2* transposase to facilitate genomic integration. GFP expression was monitored at 48 and 72 h post-fertilization (hpf), both time points when the pectoral fin can be observed. Two of our four functional mouse limb enhancers, *DYNC1I1* eExon 15 and *HDAC9* eExon 19, were found to be functional fin enhancers in zebrafish (Supplemental Table 4). At 72 hpf, *DYNC1I1* eExon 15 drove GFP expression in the pectoral fin, caudal fin, and somitic muscles (Supplemental Fig. 1B), and *HDAC9* eExon 19 exhibited enhancer activity in the pectoral fin and branchial arch (Supplemental Fig. 2C).

In order to determine whether the actual exonic sequences are necessary for enhancer activity, we used these two fin enhancers, *DYNC1I1* eExon 15 and *HDAC9* eExon 19, for deletion series analyses. *DYNC1I1* eExon 15 was divided into three segments—5′ intron, exon, and 3′ intron (Fig. 2A)—and *HDAC9* eExon 19 was divided into the following segments: 5′ distal intron, 5′ proximal intron, and exon (Fig. 2C). We found that in both cases the exon and 5′ intron sequence adjacent to the exon had lower enhancer activity by themselves, but when combined, their enhancer activity was substantially increased and comparable to that of the previously injected longer version of *DYNC1I1* eExon 15 and *HDAC9* eExon 19, respectively (Fig. 2B,D). These results demonstrate that the exonic sequences are necessary but not sufficient for full enhancer activity.

Limb genes associated with eExons enhancer function

In order to identify the limb expressed genes that could be regulated by our characterized eExons, we analyzed the RNA expression of nearby genes and carried out synteny block analysis (Ahituv et al. 2005). Whole-mount in situ hybridization of neighboring genes found that *Dlx5/6* have similar limb expression patterns to *DYNC1I1* eExons 15 and17 (Fig. 1B–E′; Supplemental Figs. 1,3), and *Twist1* has a limb expression pattern that is similar to *HDAC9* eExons 18 and 19 (Fig. 1G–I′; Supplemental Figs. 2,3). We also extracted RNA from E11.5 mouse limbs and adult mouse cortex and heart and performed quantitative PCR (qPCR) to validate the tissue specific expression of these genes. *Dlx5*, *Dlx6*, and *Twist1* were expressed in E11.5 limbs (Supplemental Fig. 3F,G). However, *Dync1i1* and *Hdac9* were not detected in mouse E11.5 limbs but expressed in the mouse adult cortex and heart, respectively (Supplemental Fig. 3F,G). In addition, examination of the genomic location of *DYNC1I1-DLX5/6* and *HDAC9-TWIST1* in various vertebrate genomes shows that they remain adjacent to each other from human to fish (Supplemental Fig. 4). Based on a previous analysis (Ahituv et al. 2005), the human–mouse–chicken *DYNC1I1-DLX5/6* block is 1.37 Mb in size and the *HDAC9-TWIST1* is 2.52 Mb, both above the 1.02 Mb average length (N50) of a human–mouse–chicken synteny block in that study. These results further suggest that these eExons could be important for *DLX5/6* and *TWIST1* regulation.

Dynclil eExon 15 is marked in the limb by an enhancer chromatin signature

To examine the dual role of these DNA sequences, we chose *DYNC1I1* eExon 15 for further functional analysis. We analyzed this eExon

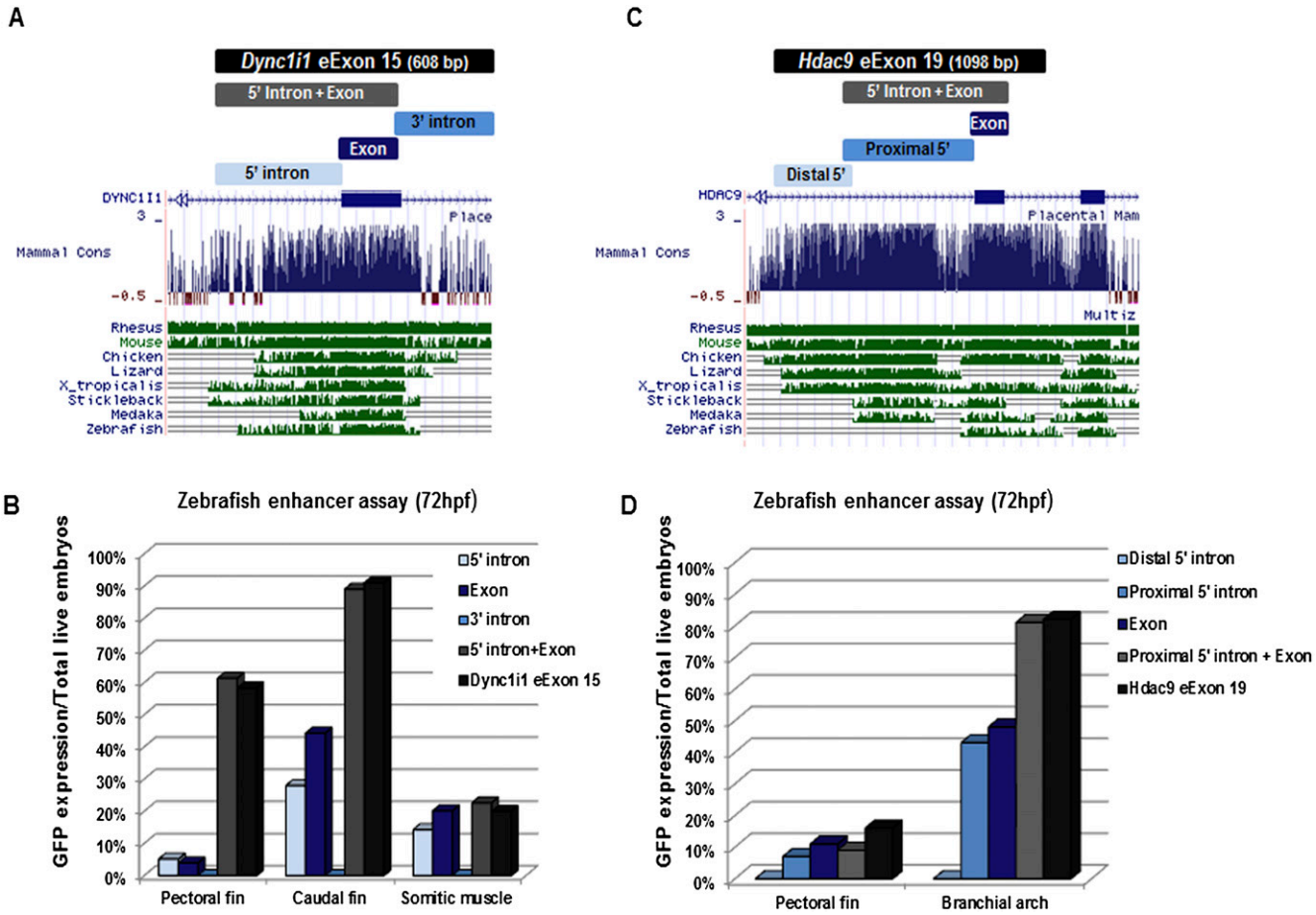

Figure 2. Segmental analysis of *DYNC1I1* eExon 15 and *HDAC9* eExon 19 enhancer function in zebrafish. (*A*) *DYNC1I1* eExon 15 was divided into three overlapping segments: 5′ intron, exon, 3′ intron. (*Below*) The UCSC Genome Browser (http://genome.ucsc.edu) conservation track shows that only the 5′ intron and exon are conserved between human and fish. (*B*) Zebrafish enhancer assay results for the different *DYNC1I1* eExon 15 segments. While the 5′ intron and exon show enhancer activity in the fins and somitic muscles, only the combination of both gives comparable enhancer expression to the 608-bp originally injected fragment of *DYNC1I1* eExon 15. The 3′ intron segment did not show enhancer activity. (*C*) *HDAC9* eExon 19 was divided into three overlapping segments: distal 5′ intron, proximal 5′ intron, and exon. (*Below*) The UCSC Genome Browser conservation track shows that the proximal 5′ intron and exon are conserved between human and fish. (*D*) Zebrafish enhancer assay results for the different *HDAC9* eExon 19 segments. While the proximal 5′ intron and exon show enhancer activity in the pectoral fin and branchial arches, only the combination of both gave comparable enhancer expression to the previously injected 1098-bp *HDAC9* eExon 19 sequence. The distal 5′ intron segment did not show enhancer activity. Enhancer function is plotted as percentage of GFP expression/total live embryos. Each of these segments was injected into at least 100 zebrafish embryos.

for histone modification signatures during limb development. We carried out ChIP followed by qPCR (ChIP-qPCR) on *Dync1i1* eExon 15 for enhancer (H3K4me1, H3K27ac), promoter (H3K4me3), and transcribed gene (H3K36me3) chromatin signatures (Hon et al. 2009). We found that in the mouse E11.5 limb, *Dync1i1* eExon 15 is marked by H3K4me1 and H3K27ac (Fig. 3B,C) but not by H3K4me3 or H3K36me3 (Fig. 3D,E). In contrast, *Dync1i1* exon 6 was not marked by H3K4me1 or H3K27ac (Fig. 3B,C) in the limb, and *Dlx5/6* coding exons were marked by H3K36me3 (Fig. 3E). Thus, the chromatin status correlates with the proposed limb enhancer activity of *Dync1i1* eExon 15.

3C and DNA FISH show that *Dync1i1* eExon 15 physically interacts with the promoter regions of *DLX5/6*

To determine whether *Dync1i1* eExon 15 physically interacts with the *Dlx5/6* promoter regions, we carried out 3C on mouse E11.5 heart and limb (AER enriched; see Methods) tissues. The mouse heart tissue served as a negative control, as *Dlx5/6* are not expressed in the heart during that stage (Fig. 1D,E). We observed an increased interaction frequency between *Dync1i1* eExon 15 and the *Dlx5/6* promoters in the limb tissue compared with the heart, indicating a physical interaction between them in the limb (Fig. 4B). These results suggest that *Dync1i1* eExon 15 functions as an enhancer in the AER through enhancer–promoter DNA looping.

To further analyze the chromosomal conformation around the *Dlx5/6* locus during limb development, we performed DNA FISH using *Dlx5/6* and *Dync1i1* eExon 15 probes on mouse E11.5 limb buds and heart. After capturing images of the two fluorescent signals, the physical distance between *Dync1i1* eExon 15 and the *Dlx5/6* coding region was calculated (Fig. 4C–J). Frequency distribution patterns of the physical distance between *Dync1i1* eExon 15 and *Dlx5/6* for the AER compared with the heart were measured (Fig. 4K,L). In the AER, 35% of the *Dync1i1* eExon 15 signals were in close proximity to the *Dlx5/6* signals (<0.2 μm) (Fig. 4K), with a mean distance of 0.32 ± 0.06 μm. In contrast, the frequency of colocalized signals in the heart was greatly reduced (12%; $P < 0.01$, *t*-test) (Fig. 4L), and the overall frequency of separated signals was higher compared to the AER

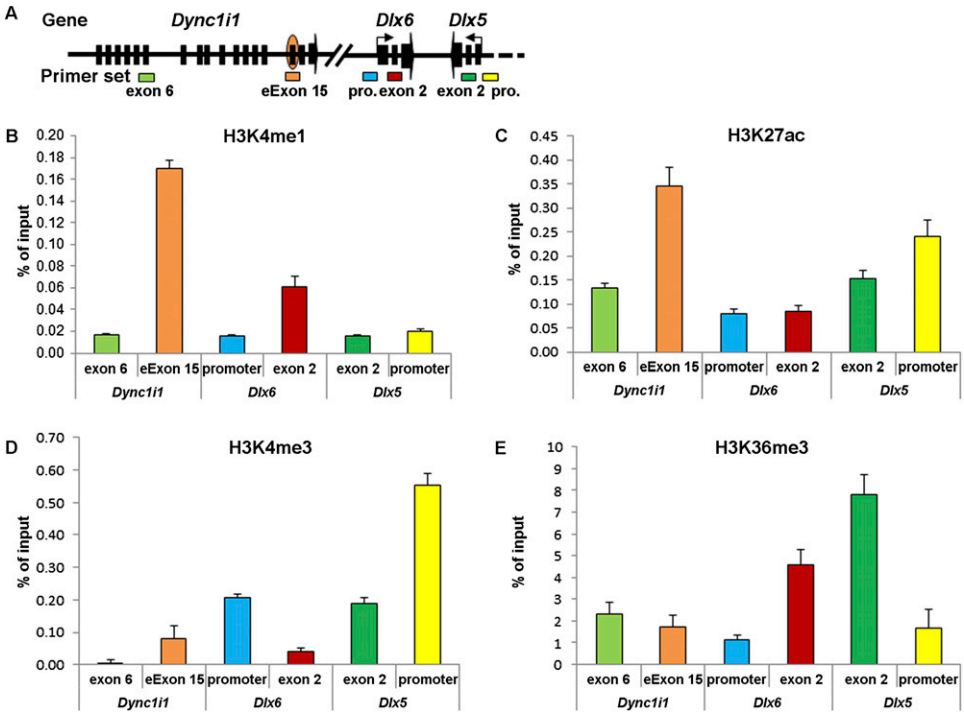

Figure 3. Histone modification signatures of the *Dync1i1* eExon 15 in the mouse E11.5 limb bud. (*A*) Schematic representation of the *Dync1i1-Dlx5/6* locus, showing the relative positions of primer sets used for ChIP-qPCR analyses: *Dync1i1* exon 6 and eExon 15; *Dlx6* promoter (pro.) and exon 2; *Dlx5* exon 2 and promoter (pro.). (*B*) ChIP–qPCR analyses of H3K4me1, an enhancer histone mark. (*C*) ChIP-qPCR analyses of H3K27ac, an active enhancer histone mark. (*D*) ChIP-qPCR analyses of H3K4me3, a promoter histone mark. (*E*) ChIP-qPCR analyses of H3K36me3, a transcribed gene histone mark. (*X*-axis) Primer pairs; (*y*-axis) percentage of input recovery. (Error bars) SE from three technical replicates of a representative experiment.

($P < 0.01$, *t*-test) (Fig. 4L), with a mean distance of 0.47 ± 0.3 μm. These results show that *Dync1i1* eExon 15 is in close proximity with *Dlx5/6* promoter regions in the developing AER at E11.5, supporting its proposed role as an enhancer during limb development.

Human chromosomal aberrations encompassing *DYNC1I1* eExons 15 and 17 are associated with SHFM1

To test whether alterations of *DYNC1I1* eExon 15 and 17 could be associated with a limb phenotype, we analyzed available individuals and previously reported cases with SHFM1 (Fig. 5). We mapped a family (GK) with SHFM1 that has a 46,XY,t(7;20)(q22;p13) translocation (Fig. 5). In addition, we mapped the inversion breakpoints of a previously published SHFM1 family (Tackels-Horne et al. 2001) to be within chr 7: 96,219,611 and 109,486,136 (K6200 family) (Everman et al. 2005; Everman et al. 2006). In addition, we referred to two recently reported SHFM1 cases: an individual with SHFM1 who has a de novo pericentric inversion of chromosome 7: 46, XY, inv(7) (p22q21.3), with the breakpoint mapped to chr 7: 95.53–95.72 Mb (van Silfhout et al. 2009), and another individual with a split foot phenotype who has an 880-kb microdeletion of 95.39–96.27Mb (Fig. 5; Kouwenhoven et al. 2010). It is worth noting that an AER enhancer named BS1 was recently identified 300 kb centromeric to *DLX5/6* (Kouwenhoven et al. 2010). However, at least two individuals with SHFM1 that are described here have chromosomal aberrations that do not include BS1 (Fig. 5), suggesting that additional limb enhancers, such as *DYNC1I1* eExon 15 and 17, could lead to SHFM1. All of the chromosomal abnormalities described above overlap *DYNC1I1* eExon 15 and 17 and suggest that their removal could disrupt the transcriptional regulation of *DLX5/6* and be one of the causes of these human limb malformations.

Discussion

Studies aimed at discovering gene regulatory elements usually concentrate on noncoding DNA sequences as potential candidates and ignore coding sequences. However, several studies have shown that protein coding sequences may have additional encrypted information in their sequence (Chamary et al. 2006; Itzkovitz and Alon 2007; Lin et al. 2011). Here, by analyzing ChIP-seq data sets for enhancer marks from various cell lines and tissues, we found that on average 7% of peaks overlap with coding exons after excluding the first exon. With only ~1.6% of the human or mouse genomes encoding for protein, eExons could be overrepresented in these ChIP-seq enhancer-associated data sets. To test whether exons are enriched in ChIP-seq enhancer data sets, we generated a random data set from all mappable sequences that are used for any whole-genome sequencing alignment from the UCSC Genome Browser (http://moma.ki.au.dk/genome-mirror/cgi-bin/hgTrackUi?hgsid=148&c=chrX&g=wgEncodeMapability) containing an identical number of peaks as in the EP300 ChIP-seq data sets of GM12787 and K562 (51,260 and 17,883 peaks, respectively) and tested how many peaks overlap exons compared to the ChIP-seq data sets. We found a significantly higher percentage of peaks overlapping with coding exons (after excluding the first exon) in the EP300 ChIP-seq data sets of both GM12787 ($P < 0.014$; Fisher exact test) and K562 ($P < 8.44 \times 10^{-8}$; Fisher exact test) cell lines. In addition, using a random sampling approach, we randomly sampled 1000 peaks (from all mappable sequences) having an equal distribution to that of the two EP300 ChIP-seq data sets 1000 times and found that a significantly higher fraction of peaks overlapped coding exons in the ChIP-seq data sets versus our random samples ($P < 2.2 \times 10^{-16}$). Combined, these assays suggest an overrepresentation of

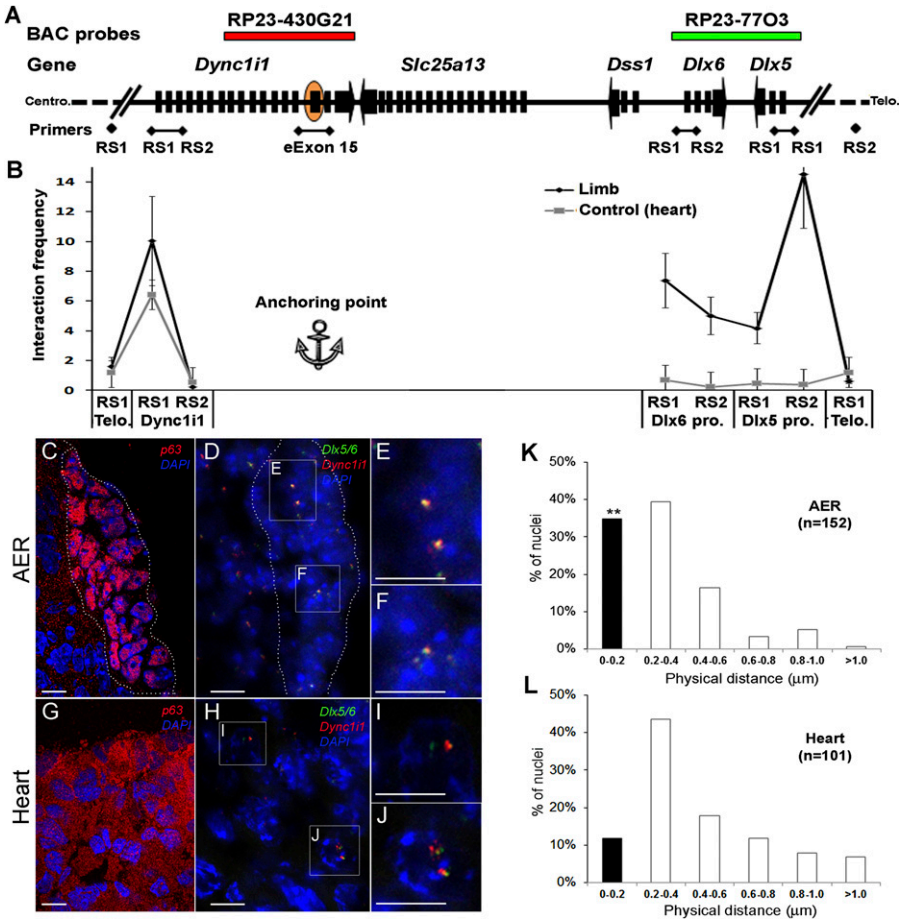

Figure 4. 3C and DNA-FISH show a physical interaction between *Dync1i1* eExon 15 and *Dlx5/6* promoter regions in the mouse E11.5 limb. (*A*) Schematic of the *Dync1i1-Dlx5/6* locus, showing the relative location of the primers used for 3C and the BAC probes used for DNA-FISH. (*B*) Chromatin looping events detected using 3C between *Dync1i1* eExon 15 (orange oval) and promoters within the *Dync1i1-Dlx5/6* locus. The closest HindIII restriction sites (RS1 and RS2) of each promoter were used to analyze the interaction frequencies to *Dync1i1* eExon 15 (anchoring point). In the limb, the interaction frequencies between *Dync1i1*eExon 15 and *Dlx6* and *Dlx5* promoter regions were significantly higher compared to the heart negative control (more than 10- and 15-fold, respectively). No significant interaction differences were found between *Dync1i1*eExon 15 and the *Dync1i1* promoter, the closest tested site to the anchoring point, or the two control regions (~900 kb away from the *Dync1i1-Dlx5/6* locus) in limb versus heart tissues. (Error bars) SE of the average of three independent PCR reactions. (*C–L*) DNA-FISH results with BAC probe RP23-430G21, which covers the *Dync1i1*eExon 15 region (red), and BAC probe RP23-77O3, which covers the *Dlx5/6* gene regions (green). (C) E11.5 limb section with the dotted line highlighting the AER, as depicted by p63 staining in the nucleus. (*D*) BAC probes and DAPI staining of E11.5 limbs. (Squares) Magnified regions in *E* and *F* that highlight the colocalization of *Dync1i1*eExon 15 and *Dlx5/6* signals. (*G*) E11.5 heart section shows p63 staining in the cytoplasm. (*H*) BAC probes and DAPI staining of E11.5 heart. (Squares) Magnified regions in *I* and *J* that show a separation of *Dync1i1*eExon 15 and *Dlx5/6* signals. The white scale bars represent 5 μm length. (*K,L*) Calculated frequencies for every 0.2 μm distance interval in mouse E11.5 AER (*K*) and heart (*L*) tissues. (Black columns) Fraction of colocalized signals (0–0.2 mm). The number (n) of loci observed in this experiment indicates a significant difference between the frequencies of the colocalized signals in the AER and heart tissues (**$P < 0.01$; Student's *t*-test).

coding exons in ChIP-seq data sets. However, it is worth noting that the technical variability of the ChIP-seq assay due to differences in antibodies, cross-linking, pull down, sequencing depth, and others along with sequence mappability are not taken into account in these analyses.

Using a mouse transgenic enhancer assay for seven mouse E11.5 limb EP300 ChIP-seq peaks, we show that four eExons are functional limb enhancers and could regulate their neighboring genes. The observed 57% (4/7) success rate does not imply that ~57% of exons overlapping enhancer-associated ChIP-seq peaks are bona fide enhancers and further functional assays will be needed in order to determine this.

It is worth noting that *Dlx5/6* and *Twist1* expression as detected by whole-mount in situ hybridization is restricted to the AER and the A-P domains of the developing mouse limb, respectively. However, the mouse limb enhancer expression pattern of *DYNC1I1* eExon 15 extends into the limb bud mesenchyme and the *HDAC9* eExon 18 extends into the posterior limb bud mesenchyme. These expression pattern discrepancies could be due to an inability of the whole-mount in situ assay to detect low RNA expression levels versus the more robust staining seen through LacZ enhancer assays. One such example is the ability to study the role and expression of the myocyte enhancer factor 2C (*MEF2C*) gene in the neural crest due to its enhancer function, which was previously confounded and could not be observed through in situ hybridization (Agarwal et al. 2011). Alternatively, these limb enhancers could potentially be regulating other limb-associated genes. However, our results for *DYNC1I1* eExon 15 demonstrate a physical interaction between this eExon and the *Dlx5/6* promoter regions, suggesting that it regulates *Dlx5/6* expression in the limb. Another possibility for the discrepancy in expression patterns could be associated with the "artificial" nature of the transgenic enhancer assay, which could lead to different results due to the use of a minimal promoter instead of the promoter of the regulated gene, the site of transgene integration, the variation in transgene copy number between transgenic animals, and/or other complications.

Two of our characterized limb enhancers, *DYNC1I1* eExon 15 and 17, reside in the coding exons of *DYNC1I1*, a subunit of the cytoplasmic Dynein 1 motor protein complex that is not expressed in the limb during development (Supplemental Fig. 3A,F; Crackower et al. 1999). Both *DYNC1I1* eExon 15 and 17 are present in all three *DYNC1I1* splice isoforms (RefSeq: NM_004411.4, NM_001135557.1, NM_001135556.1), suggesting that alternative splicing does not occur at these exons. *DYNC1I1* eExon15 is also marked by an enhancer chromatin signature and physically interacts with the *Dlx5/6* promoter regions specifically in the limb. Both eExons encode protein domains important for the cargo binding and specificity of this Dynein (Kardon and Vale 2009). Cytoplasmic Dynein 1 is involved in neuronal migration during brain development by interacting with Dynactin and platelet-activating factor acetylhydrolase 1b (*LIS1*). Defects in neuronal migration can lead to brain malfor-

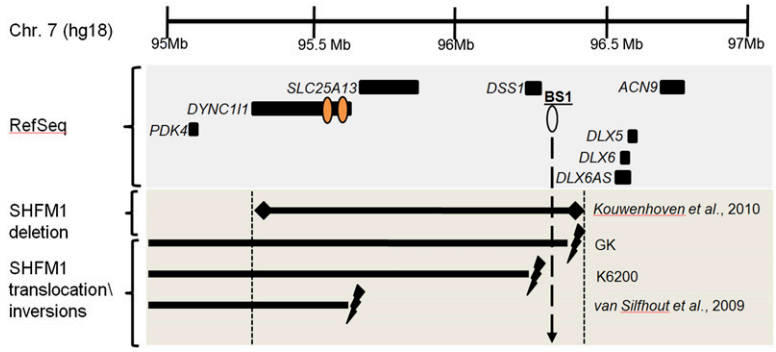

Figure 5. Chromosomal abnormalities at chromosome 7q21-23 associated with SHFM1. A schematic representation of the genomic positions of breakpoints from chromosomal rearrangements in individuals with SHFM1 mapped to human genome assembly 18 (hg18) and compared to the location of *DYNC1I1* eExon 15 and 17. An 880-kb microdeletion in an individual with a split foot phenotype was found to be at 95.39–96.27 Mb (Kouwenhoven et al. 2010). In the GK family, the 46,XY,t(7;20)(q22;p13) translocation breakpoints mapped to chr 7: 96.2–96.47 Mb. In the K6200 family, the chromosomal inversion breakpoints mapped to chr 7: 96,219,611 and 109,486,136. The breakpoint coordinates of a 7:46, XY, inv(7) (p22q21.3) with SHFM1 and pervasive developmental disorder-not otherwise specified (PDD-NOS) was found to be at chr 7: 95.53–95.72 Mb (van Silfhout et al. 2009). All of these chromosomal abnormalities overlap with *DYNC1I1* eExon 15 and 17 (orange ovals). Two of these chromosomal aberrations do not overlap with the BS1 AER enhancer (white oval). (Lightning bolts) Translocation and inversion breakpoints; (diamonds) deletion.

mations such as lissencephaly, subcortical laminar heterotopias, and pervasive developmental disorder-not otherwise specified (PDD-NOS) (Kato and Dobyns 2003). Interestingly, an individual with PDD-NOS and SHFM1 has an inversion in the *DYNC1I1* region (chr 7: 95.53–95.72 Mb) (Fig. 5), whose breakpoint has not been finely characterized (van Silfhout et al. 2009). Further analysis would be required in order to establish whether the PDD-NOS and SHFM phenotypes in this individual could be due to the disruption of both the *DYNC1I1* gene and our characterized eExons.

Two other characterized limb enhancers, *HDAC9* eExon 18 and 19, reside in the coding exons of *HDAC9*, a member of the histone acetyltransferase class II family. *HDAC9* eExon 19 has also been shown to be an exonic remnant in zebrafish and speculated to have a *cis*-regulatory function (Dong et al. 2010). Both *HDAC9* eExons 18 and eExon 19 appear in 3/9 *HDAC9* spice isoforms (RefSeq: NM_178423.1, NM_058176.2, NM_178425.2). *HDAC9* expression was shown to be more selective compared with that of other HDAC family members (de Ruijter et al. 2003). Our RNA analysis and whole-mount in situ hybridization results show that *Hdac9* is not expressed in the mouse limb at E11.5 (Supplemental Fig. 3D,G). *Hdac9*-null mice generated by deletion of exons 4 and 5 are fertile and survive a normal life span but develop cardiac hypertrophy with age and in response to pressure overload (Zhang et al. 2002). Interestingly, despite *Hdac9* not being expressed in the limb at E11.5, *Hdac9* homozygous knockout mice develop polydactyly in their hindlimbs with partial penetrance (Morrison and D'Mello 2008), similar to the polydactyly phenotype of *Twist1* heterozygous knockout mice (Bourgeois et al. 1998). Although *Hdac9* eExons 18 or 19 were not removed in these *Hdac9*-null mice, the regulation of *Twist1* by these and other potential *Twist1* enhancers could be disrupted leading to the polydactyly phenotype.

The ability of eExons to enhance the expression of their nearby genes, but not the gene they reside in, could suggest that mechanisms such as those involved in epigenetic regulation and high-order chromatin organization might control their function in each tissue. To our knowledge, the functional demonstration that DNA sequences can act as a protein coding sequence in one tissue but regulate the expression of a nearby gene/s in another tissue is novel. It raises the possibility that mutations in a certain gene, even synonymous ones, could potentially affect the regulation of a nearby gene. Therefore, careful analysis of the tissue-specific expression and function of a gene would be required in order to determine whether a phenotype is truly caused by a mutation within its coding sequence.

Methods

Computational ChIP-seq data set analyses

We identified exonic sequences in the human hg18 and mouse mm9 genome assemblies using the UCSC knownGene track (http://genome.ucsc.edu). We downloaded all exonic sequences, including 5′ UTR and 3′ UTR, using the txStrat and txEnd filter field. All exon sequence sizes were divided by the number of exons to calculate the average exon size. We downloaded coding exon sequences using the cdsStart and cdsEnd filter field. The 22 ChIP-seq data sets of human cell lines were obtained from Ernst et al. (2011), Myers et al. (2011), and Rosenbloom et al. (2012) and were downloaded from the UCSC Genome Browser, and the three EP300 ChIP-seq data sets of mouse E11.5 tissues were obtained from Visel et al. (2009a) and downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) (Supplemental Table 1 includes links for all downloaded data). In order to unify our results, human sequences with hg19 coordinates were converted to hg18 using the UCSC Genome Browser LiftOver tool. A ChIP-seq peak was considered to overlap an exon if at least 1 bp of exonic sequence overlapped. BED files of all the ChIP-seq peaks that overlap exons in the various data sets can be obtained at http://bts.ucsf.edu/ahituv/resources.html. First exons for all splice isoforms of a gene were determined by the exonStarts exonEnds field in the UCSC knownGene track. To identify limb expressed genes, we used available mouse RNA in situ data from the Mouse Genome Informatics (MGI) gene expression data query form (http://www.informatics.jax.org/javawi2/servlet/WIFetch?page=expressionQF) and defined a limb expressing gene as one having RNA in situ expression data at either TS19 (E11.0–12.25) or TS20 (E11.5–13.0).

Transgenic enhancer assays

By use of primers designed to amplify the EP300 ChIP-seq peaks that overlap exons (Supplemental Table 5), we carried out polymerase chain reaction (PCR) on human genomic DNA (Qiagen). Primers were designed to have up to 500 bp additional sequence flanking the EP300 peak. Previous experiments have shown this to be a reliable method for obtaining positive enhancer activity when using evolutionary conserved regions (Pennacchio et al. 2006) and EP300 ChIP-seq peaks (Visel et al. 2009a). For the mouse enhancer assays, PCR products of the human genomic regions were cloned into a vector containing the *Hsp68* minimal promoter followed by the *LacZ* reporter gene (Pennacchio et al. 2006) and sequence verified. Transgenic mice were generated by the UCSF transgenic facility and by Cyagen Biosciences using standard procedures (Nagy et al. 2002). Embryos were harvested at E11.5 and stained for LacZ expression as previously described (Pennacchio et al. 2006). For the zebrafish enhancer assays,

the same human PCR products were cloned into the E1b-GFP-Tol2 enhancer assay vector containing an E1b minimal promoter followed by GFP (Li et al. 2009). They were injected following standard procedures (Nusslein-Volhard and Dahm 2002; Westerfield 2007) into at least 100 embryos per construct along with *Tol2* mRNA (Kawakami 2005) to facilitate genomic integration. GFP expression was observed and annotated at 48 and 72 hpf. An enhancer was considered positive if 60% of the GFP expressing fish showed a consistent expression pattern. All animal work was approved by the UCSF Institutional Animal Care and Use Committee.

Whole-mount in situ hybridization

Mouse E11.5 embryos were fixed in 4% paraformaldehyde. Clones containing mouse *Dync1i1* (MMM1013-9202215, Open Biosystems), *Dlx5* (Depew et al. 1999), *Dlx6* (OMM5895-99863403 Open Biosystems), *Hdac9* (EMM1032-601163 and EMM1002-6974502, Open Biosystems), and *Twist1* (Chen and Behringer 1995) were used as templates for digoxygenin-labeled probes. Mouse whole-mount in situ hybridizations were performed according to standard procedures (Hargrave et al. 2006).

RNA expression analysis

Mouse E11.5 limb and AER enriched tissues (limb buds where the AER region was carefully dissected), and adult mouse heart and cortex tissues were dissected. Total RNA was isolated using RNeasy (Qiagen) according to the manufacturer's protocol. qPCR was performed using SsoFast EvaGreen Supermix (Biorad) and run on the Eppendorf Mastercycler ep realplex 2 thermal cycler. Samples were tested in duplicates. Specificity and absence of primer dimers was controlled by denaturation curves. β-Actin (*Actb*) mRNA was used for normalization. Primer sequences used for amplification are listed in Supplemental Table 5.

ChIP followed by qPCR

ChIP following standard techniques (Nelson et al. 2006) was performed on mouse E11.5 AER-enriched tissue. For each ChIP, 100–500 mg of chromatin was used. For immunoprecipitation, we used 2 μg of H3K4me1 (ab8895, Abcam), H3K4me3 (ab8580, Abcam), H3K27ac (ab4729, Abcam), and H3K36me3 (ab9050; Abcam) antibodies. qPCR was carried out using SsoFast EvaGreen Supermix (Biorad) and run on the Eppendorf Mastercycler ep realplex 2 thermal cycler. ChIP-qPCR signals were standardized to input chromatin (percentage of input). Primer sequences used for amplification are listed in Supplemental Table 5.

3C assay

3C was performed following standard procedures (Dostie and Dekker 2007). Mouse E11.5 heart and AER enriched tissues were dissected from 30 embryos, cross-linked with 1% formaldehyde, and processed to get single cell preparations. Cells were lysed to purify nuclei and digested with HindIII (1200 units) restriction enzyme (New England Biolabs). Cross-linked fragments were ligated with 2000 units of T4 DNA ligase (New England Biolabs) for 3 d at 4°C. The samples were reverse cross-linked, and purified DNA was amplified by whole-genome amplification (WGA2, Sigma-Aldrich). Product detection was done in triplicate by qPCR, as described above for ChIP, and averaged for each primer pair (Supplemental Table 5). Each data point was first corrected for PCR bias by dividing the average of three PCR signals by the average signal of an internal control template. Data from AER and heart were normalized to a BAC library containing seven BACs obtained from the CHORI BACPAC resource center covering the SHFM1 minimal region (RP23-430G21, RP24-73K21, RP23-336P10, RP23-389M11, RP24-343G1, RP24-270A16).

DNA florescent in situ hybridization

DNA florescent in situ hybridization (FISH) was carried out as previously described (Lomvardas et al. 2006). BAC clones RP23-77O3 for *Dlx5/6* and RP23-430G21 for *Dync1i1* were obtained from the CHORI BACPAC resource center. Probes were labeled with Digoxigenin-11-dUTP or Biotin-16-dUTP by Nick Translation (Roche). Limb or heart tissues (E11.5) were embedded without fixation, and 10 μM cryosections were collected on Superfrost Plus slides (Fisher). After drying, sections were fixed in 4% PFA for 5 min at 4°C. DNA was fragmented by incubation with 0.1 M HCl for 5 min at room temperature, and slides were treated with RNase A for 1 h at 37°C. Slides were dried by an ethanol series, denatured in a solution of 75% formamide in 2xSSC for 5 min at 85°C, rinsed immediately in ice-cold 2xSSC, and dried again by 4°C ethanol series. Pre-denatured, CotI-annealed probes were applied overnight. The probe was washed three times for 15 min in 55% formamide, 0.1% NP-40 in 2xSSC at 42°C. Probes were detected using Dylight 488 anti-digoxigenin and Dylight 549 anti-biotin (Jackson Immunoresearch). Antibody washes were carried out in a solution of PBS containing 0.1% Triton-x-100 and 8% formamide at room temperature. All images were obtained using confocal fluorescence microscopy (Nikon C1 Spectral). FISH signals were recorded in three separate RGB channels. The image stacks were reconstructed using the Volocity program (PerkinElmer), and the shortest distance between the gravity centers of the *Dlx5/6* and *Dync1i1* signals was calculated.

Subjects and chromosomal breakpoint mapping

The GK family consisted of a male who had ectrodactyly, micrognathia, an elongated neck, and bilateral microtia with neurosensory deafness and his female offspring who died before birth and had ectrodactyly, micrognathia, and bilateral microtia. Karyotypes of the father and his offspring demonstrated a reciprocal balanced chromosomal translocation 46,XY,t(7;20)(q22;p13) that was not found in GK's healthy mother. By use of FISH, following standard techniques (Trask 1991), with two BACs (RP11-94N7, RP11-78B12), the breakpoint coordinates at chromosome 7 were mapped to be between 96.2 and 96.47 Mb. The K6200 family had autosomal dominant SHFM and variable sensorineural hearing loss as previously reported (Tackels-Horne et al. 2001). Subsequent studies of this family by pulse field gel electrophoresis and FISH identified a chromosome inversion with breakpoints in the SHFM1 critical region (Everman et al. 2005; Everman et al. 2006). Southern blot analysis and inverse PCR as previously described (Vervoort et al. 2002) were then used to identify the inversion breakpoints (D.B. Everman, C.T. Morgan, M.E. Laughridge, T. Moss, S. Ladd, B. DuPont, D. Toms, A. Dobson, K.D. Clarkson, F. Gurrieri, et al., unpubl.). The inversion in this family was balanced, with minimal changes in the normal sequence at each breakpoint and segregated with the SHFM/hearing loss phenotype.

Acknowledgments

We thank members of the Ahituv laboratory for helpful comments on the manuscript. We also thank Juhee Jeong and John L.R. Rubenstein for reagents. This research was supported by NICHD grant no. R01HD059862. N.A. and G.B. are also supported by NHGRI grant number R01HG005058, and N.A. is also supported by NIGMS award number GM61390. M.J.K. was supported in part by NIH Training Grant T32 GM007175 and the Amgen Research Excellence in Bioengineering and Therapeutic Sciences Fellow-

ship. O.A. and O.S.B. were supported by the Morris Kahn family foundation. D.B.E and C.E.S. were supported in part by a grant from the South Carolina Department of Disabilities and Special Needs, the Genetic Endowment of South Carolina, and a previous grant (no. 8510) from Shriners Hospitals for Children. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NICHD, NHGRI, NIDCR, or the NIGMS.

References

Agarwal P, Verzi MP, Nguyen T, Hu J, Ehlers ML, McCulley DJ, Xu SM, Dodou E, Anderson JP, Wei ML, et al. 2011. The MADS box transcription factor MEF2C regulates melanocyte development and is a direct transcriptional target and partner of SOX10. *Development* **138:** 2555–2565.

Ahituv N, Prabhakar S, Poulin F, Rubin EM, Couronne O. 2005. Mapping *cis*-regulatory domains in the human genome using multi-species conservation of synteny. *Hum Mol Genet* **14:** 3057–3063.

Bourgeois P, Bolcato-Bellemin AL, Danse JM, Bloch-Zupan A, Yoshiba K, Stoetzel C, Perrin-Schmitt F. 1998. The variable expressivity and incomplete penetrance of the *twist*-null heterozygous mouse phenotype resemble those of human Saethre-Chotzen syndrome. *Hum Mol Genet* **7:** 945–957.

Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* **7:** 98–108.

Chen ZF, Behringer RR. 1995. *twist* is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev* **9:** 686–699.

Crackower MA, Sinasac DS, Xia J, Motoyama J, Prochazka M, Rommens JM, Scherer SW, Tsui LC. 1999. Cloning and characterization of two cytoplasmic dynein intermediate chain genes in mouse and human. *Genomics* **55:** 257–267.

Depew MJ, Liu JK, Long JE, Presley R, Meneses JJ, Pedersen RA, Rubenstein JL. 1999. *Dlx5* regulates regional development of the branchial arches and sensory capsules. *Development* **126:** 3831–3846.

de Ruijter AJ, van Gennip AH, Caron HN, Kemp S, van Kuilenburg AB. 2003. Histone deacetylases (HDACs): Characterization of the classical HDAC family. *Biochem J* **370:** 737–749.

Dong X, Navratilova P, Fredman D, Drivenes O, Becker TS, Lenhard B. 2010. Exonic remnants of whole-genome duplication reveal *cis*-regulatory function of coding exons. *Nucleic Acids Res* **38:** 1071–1085.

Dostie J, Dekker J. 2007. Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc* **2:** 988–1002.

Eichenlaub MP, Ettwiller L. 2011. De novo genesis of enhancers in vertebrates. *PLoS Biol* **9:** e1001188. doi: 10.1371/journal.pbio.1001188.

Elliott AM, Evans JA. 2006. Genotype-phenotype correlations in mapped split hand foot malformation (SHFM) patients. *Am J Med Genet A* **140:** 1419–1427.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473:** 43–49.

Everman D, Morgan C, Clarkson K, Gurrieri F, McAuliffe F, Chitayat D, Stevenson R, Schwartz C. 2005. Submicroscopic rearrangements involving the SHFM1 locus on chromosome 7q21-22 are associated with split-hand/foot malformation and sensorineural hearing loss. *Proc Greenwood Genet Cent* **24:** 137.

Everman D, Morgan C, Stevenson R, Schwartz C. 2006. Chromosome rearrangements: An emerging theme in the causation of split-hand/foot malformation. *Proc Greenwood Genet Cent* **25:** 138–139.

Firulli BA, Krawchuk D, Centonze VE, Vargesson N, Virshup DM, Conway SJ, Cserjesi P, Laufer E, Firulli AB. 2005. Altered Twist1 and Hand2 dimerization is associated with Saethre-Chotzen syndrome and limb abnormalities. *Nat Genet* **37:** 373–381.

Gilbert SF. 2000. *Developmental Biology*, 6th ed. Sinauer Associates, Sunderland, MA.

Hall B.K. 2007. *Fins into limbs*. The University of Chicago Press, Chicago.

Hargrave M, Bowles J, Koopman P. 2006. In situ hybridization of whole-mount embryos. *Methods Mol Biol* **326:** 103–113.

Heintzman ND, Ren B. 2009. Finding distal regulatory elements in the human genome. *Curr Opin Genet Dev* **19:** 541–549.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459:** 108–112.

Hon GC, Hawkins RD, Ren B. 2009. Predictive chromatin signatures in the mammalian genome. *Hum Mol Genet* **18:** R195–R201.

Iovine MK. 2007. Conserved mechanisms regulate outgrowth in zebrafish fins. *Nat Chem Biol* **3:** 613–618.

Itzkovitz S, Alon U. 2007. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res* **17:** 405–412.

Kardon JR, Vale RD. 2009. Regulators of the cytoplasmic dynein motor. *Nat Rev Mol Cell Biol* **10:** 854–865.

Kato M, Dobyns WB. 2003. Lissencephaly and the molecular basis of neuronal migration. *Hum Mol Genet* **12:** R89–R96.

Kawakami K. 2005. Transposon tools and methods in zebrafish. *Dev Dyn* **234:** 244–254.

Kothary R, Clapoff S, Brown A, Campbell R, Peterson A, Rossant J. 1988. A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature* **335:** 435–437.

Kouwenhoven EN, van Heeringen SJ, Tena JJ, Oti M, Dutilh BE, Alonso ME, de la Calle-Mustienes E, Smeenk L, Rinne T, Parsaulian L, et al. 2010. Genome-wide profiling of p63 DNA-binding sites identifies an element that regulates gene expression during limb development in the 7q21 SHFM1 locus. *PLoS Genet* **6:** e1001065. doi: 10.1371/journal.pgen.1001065.

Lampe X, Samad OA, Guiguen A, Matis C, Remacle S, Picard JJ, Rijli FM, Rezsohazy R. 2008. An ultraconserved Hox-Pbx responsive element resides in the coding sequence of *Hoxa2* and is active in rhombomere 4. *Nucleic Acids Res* **36:** 3214–3225.

Li Q, Ritter D, Yang N, Dong Z, Li H, Chuang JH, Guo S. 2009. A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Dev Biol* **337:** 484–495.

Lin MF, Kheradpour P, Washietl S, Parker BJ, Pedersen JS, Kellis M. 2011. Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res* **21:** 1916–1928.

Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. 2006. Interchromosomal interactions and olfactory receptor choice. *Cell* **126:** 403–413.

Mercader N. 2007. Early steps of paired fin development in zebrafish compared with tetrapod limb development. *Dev Growth Differ* **49:** 421–437.

Morrison BE, D'Mello SR. 2008. Polydactyly in mice lacking HDAC9/HDRP. *Exp Biol Med (Maywood)* **233:** 980–988.

Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, Birney E, Wold B, et al. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9:** e1001046. doi: 10.1371/journal.pbio.1001046.

Nagy A, Gertsenstein M, Vintersten K, Behringer R. 2002. *Manipulating the mouse embryo: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nelson JD, Denisenko O, Bomsztyk K. 2006. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc* **1:** 179–185.

Neznanov N, Umezawa A, Oshima RG. 1997. A regulatory element within a coding exon modulates keratin 18 gene expression in transgenic mice. *J Biol Chem* **272:** 27549–27557.

Nissim S, Tabin C. 2004. Development of the limbs. In *Inborn errors of development* (ed. C.E.R. Erickson and T. Wynshaw-Boris), pp. 148–167. Oxford University Press, New York.

Nusslein-Volhard C, Dahm R. 2002. *Zebrafish*. Oxford University Press, Oxford.

O'Rourke MP, Soo K, Behringer RR, Hui CC, Tam PP. 2002. *Twist* plays an essential role in FGF and SHH signal transduction during mouse limb development. *Dev Biol* **248:** 143–156.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. 2006. *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444:** 499–502.

Ritter DI, Dong Z, Guo S, Chuang JH. 2012. Transcriptional enhancers in protein-coding exons of vertebrate developmental genes. *PLoS One*. doi: 10.1371/journal.pone.0035202.

Robledo RF, Rajan L, Li X, Lufkin T. 2002. The *Dlx5* and *Dlx6* homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. *Genes Dev* **16:** 1089–1101.

Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, Cline MS, Karolchik D, Barber GP, Clawson H, et al. 2012. ENCODE whole-genome data in the UCSC Genome Browser: Update 2012. *Nucleic Acids Res* **40:** D912–D917.

Shamseldin HE, Faden MA, Alashram W, Alkuraya FS. 2012. Identification of a novel *DLX5* mutation in a family with autosomal recessive split hand and foot malformation. *J Med Genet* **49:** 16–20.

Tackels-Horne D, Toburen A, Sangiorgi E, Gurrieri F, de Mollerat X, Fischetto R, Causio F, Clarkson K, Stevenson RE, Schwartz CE. 2001. Split hand/split foot malformation with hearing loss: First report of families linked to the SHFM1 locus in 7q21. *Clin Genet* **59:** 28–36.

Trask BJ. 1991. Fluorescence *in situ* hybridization: Applications in cytogenetics and gene mapping. *Trends Genet* **7:** 149–154.

Tumpel S, Cambronero F, Sims C, Krumlauf R, Wiedemann LM. 2008. A regulatory module embedded in the coding region of *Hoxa2* controls expression in rhombomere 2. *Proc Natl Acad Sci* **105:** 20077–20082.

van Silfhout AT, van den Akker PC, Dijkhuizen T, Verheij JB, Olderode-Berends MJ, Kok K, Sikkema-Raddatz B, van Ravenswaaij-Arts CM. 2009. Split hand/foot malformation due to chromosome 7q aberrations(SHFM1): Additional support for functional haploinsufficiency as the causative mechanism. *Eur J Hum Genet* **17:** 1432–1438.

Vervoort VS, Viljoen D, Smart R, Suthers G, DuPont BR, Abbott A, Schwartz CE. 2002. Sorting nexin 3 (*SNX3*) is disrupted in a patient with a translocation t(6;13)(q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype. *J Med Genet* **39:** 893–899.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009a. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457:** 854–858.

Visel A, Rubin EM, Pennacchio LA. 2009b. Genomic views of distant-acting enhancers. *Nature* **461:** 199–205.

Westerfield M. 2007. *The zebrafish book*. University of Oregon, Eugene.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* **3:** e7. doi: 10.1371/journal.pbio.0030007.

Zeller R, Lopez-Rios J, Zuniga A. 2009. Vertebrate limb bud development: Moving towards integrative analysis of organogenesis. *Nat Rev Genet* **10:** 845–858.

Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA, Olson EN. 2002. Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* **110:** 479–488.

Received October 19, 2011; accepted in revised form March 19, 2012.