Developmental Biology 327 (2009) 526–540

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Genomes & Developmental Control

Systematic human/zebrafish comparative identification of *cis*-regulatory activity around vertebrate developmental transcription factor genes ☆

Pavla Navratilova [a], David Fredman [b], Thomas A. Hawkins [c], Katherine Turner [c], Boris Lenhard [a,b], Thomas S. Becker [a,d,*]

[a] *Sars Centre for Marine Molecular Biology, University of Bergen, 5008 Bergen, Norway*
[b] *Computational Biology Unit, University of Bergen, 5008 Bergen, Norway*
[c] *Department of Cell and Developmental Biology, UCL, Gower St. London WC1E 6BT, UK*
[d] *Brain and Mind Research Institute, University of Sydney, Camperdown, NSW 2050, Australia*

ARTICLE INFO

Article history:
Received for publication 31 July 2008
Revised 2 October 2008
Accepted 28 October 2008
Available online 12 November 2008

Keywords:
cis-regulation
Transcription
ELP4
IMMP1L
RCN1
WT1
CTCF insulator
Tol2 transposon

ABSTRACT

Pan-vertebrate developmental *cis*-regulatory elements are discernible as highly conserved noncoding elements (HCNEs) and are often dispersed over large areas around the pleiotropic genes whose expression they control. On the loci of two developmental transcription factor genes, *SOX3* and *PAX6*, we demonstrate that HCNEs conserved between human and zebrafish can be systematically and reliably tested for their regulatory function in multiple stable transgenes in zebrafish, and their genomic reach estimated with confidence using synteny conservation and HCNE density along these loci. HCNEs of both human and zebrafish function as specific developmental enhancers in zebrafish. We show that human HCNEs result in expression patterns in zebrafish equivalent to those in mouse, establishing zebrafish as a suitable model for large-scale testing of human developmental enhancers. Orthologous human and zebrafish enhancers underwent functional evolution within their sequence and often directed related but non-identical expression patterns. Despite an evolutionary distance of 450 million years, one pax6 HCNE drove expression in identical areas when comparing zebrafish vs. human HCNEs. HCNEs from the same area often drive overlapping patterns, suggesting that multiple regulatory inputs are required to achieve robust and precise complex expression patterns exhibited by developmental genes.

© 2008 Elsevier Inc. All rights reserved.

Introduction

About 20% of the sequence of the human genome consists of gene deserts, regions with few or no annotated genes (Venter et al., 2001). A subclass of gene deserts but also more gene-rich blocks of human: teleost conserved synteny are flanking key developmental regulatory genes (Ovcharenko et al., 2005) and highly conserved non-coding elements (HCNEs) cluster within these genomic regions (Bejerano et al., 2004; Sandelin et al., 2004; Woolfe et al., 2005). A substantial number of HCNEs have recognizable conservation across all vertebrate genomes (Sandelin et al., 2004; Woolfe et al., 2005), and many have been found to have enhancer activity in mouse, zebrafish, and frog (de la Calle-Mustienes et al., 2005; Nobrega et al., 2003; Pennacchio et al., 2006; Visel et al., 2008; Woolfe and Elgar, 2007).

Through enhancer detection in zebrafish (Ellingsen et al., 2005; Laplante et al., 2006) and computational analyses, we have described long-range gene regulation as a general basis of ancient synteny conservation in vertebrate and insect genomes (Becker and Lenhard, 2007; Engstrom et al., 2007; Gomez-Skarmeta et al., 2006; Kikuta et al., 2007a,b). We termed the chromosomal segments postulated to be due to the presence of long-range enhancers *genomic regulatory blocks* (GRBs). A GRB consists of a pleiotropic target gene and its enhancers in the form of HCNEs that occupy its introns and flanking regions, spanning a chromosomal region on the order of hundreds of kb to megabases (Kikuta et al., 2007b; Nobrega et al., 2003). Besides the target gene, GRBs also contain other neighboring (usually non-pleiotropic) protein-coding genes with HCNEs in their introns and in the intergenic sequence beyond these genes. In this work, we aim to show that a study that starts form the concept of GRB, including pan-vertebrate synteny and a region defined by high density of highly conserved non-coding elements leads to insights that are often missed in the study of long-range gene regulation. We chose GRBs targeting the genes *SOX3* and *PAX6* and their orthologous loci in zebrafish, because i) they have well-established developmental expression patterns, ii) many HCNEs defining their GRBs are strongly conserved across all vertebrates, and iii) they differ in the gene content of their

Abbreviations: HCNE, highly conserved noncoding element; GFP, green fluorescent protein; GRB, genomic regulatory block; MY, million years.

☆ Authors' contributions: PN and TB conceived of the study. PN did the experiments. PN, DF and BL did the bioinformatic analyses and generated the online database. PN, BL and TB wrote the paper. TAH and KT did the confocal study of PAX6_8.

* Corresponding author. Brain and Mind Research Institute, University of Sydney, Camperdown, NSW 2050, Australia. Fax: +61 2 93510997.
E-mail address: tsbecker@med.usyd.edu.au (T.S. Becker).

0012-1606/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2008.10.044

inferred GRB — while the *SOX3* region is largely devoid of other protein-coding genes, the *PAX6* region has a number of bystander genes containing inferred *PAX6* regulatory elements in their introns. We show that the GRBs of the two genes, defined using conserved synteny and HCNE density information alone represent equivalent genomic features. To demonstrate this, we i) computationally estimate the extent of GRBs targeting human *SOX3* and *PAX6* and their zebrafish orthologs, ii) systematically catalogue regulatory inputs and their genomic extent around the two genes, and iii) tested whether an HCNE interacts with different promoters. We amplified HCNEs from distal locations near, in, or beyond bystander genes from the GRBs of human and zebrafish transcription factor genes *SOX3*/*sox3* and *PAX6*/*pax6a*/*b* and tested them in transgenic zebrafish using the Tol2 transposon system (Fisher et al., 2006; Kawakami, 2004). We show that a subset of human and zebrafish HCNEs reproducibly drives reporter expression in transgenic zebrafish in a fashion that indicates that these elements regulate their target transcription factor genes, and that specific patterns are observed when either human or zebrafish HCNEs are used in these assays. Enhancers driving a single target gene are dispersed over large areas around these genes, and in the case of *PAX6* regulatory elements they are found inside the introns of at least two non-related neighboring genes. Despite the distant evolutionary origin of HCNEs, in one case an element showed identical target-gene specific patterns for human vs. zebrafish HCNEs, indicating that intense evolutionary pressure acts on these specific elements.

To fully understand GRB organization, we attempted to resolve enhancer–promoter specificity for the *SOX3-ATP11C* genomic region. We show that one specific enhancer always drives an identical *SOX3*-like pattern regardless whether the promoter used is the native *SOX3* promoter or the promoter of the neighboring *ATP11C* gene. We could not find any functional insulator sequences in the region between the *SOX3* enhancer and neighboring *ATP11C* bystander gene.

Our findings suggest that the concept of GRBs is both valid and can be efficiently investigated in the zebrafish system. The system is particularly suited for studying the open questions of enhancer interactions, enhancer:promoter responsiveness, and the effect of evolutionary HCNE differences on expression patterns. The understanding of human gene regulation will require systematic computational and transgenesis-based identification of regulatory inputs across large chromosomal segments and eventually the entire human genome. The testing of gene regulatory sequences in zebrafish should also permit to correctly link non-coding mutations to the altered expression of specific protein-coding genes in human disease.

Materials and methods

Detection and extraction of HCNE sequences

Human-zebrafish non-exonic (i.e. non-coding and non-UTR) evolutionarily conserved sequences were identified using the UCSC Genome Browser (Karolchik et al., 2003) using Vertebrate Multiz Alignment & PhastCons Conservation track and Zebrafish (Mar. 2006/danRer4) Chained Alignments track in Human Mar. 2006 Assembly (hg18) from where the human sequences were extracted. The zebrafish sequences were exported from the UCSC Genome Browser Zebrafish Mar. 2006 Assembly (zv6). Alignment plots of whole genomic regions were created using Ensembl for sequence and data extraction (human: NCBI 36, zebrafish: zv6 and for *pax6b* data zv7) and mVISTA (Dubchak and Ryaboy, 2006) under a sequence conservation threshold of at least 70% over 100 bp. The sequence coordinates together with the UCSC link are available online at http://zenbase.genereg.net/spp.

Primer design and cloning of HCNEs

PCR primers to amplify candidate human and zebrafish gene enhancers selected by evolutionary conservation analysis as described above were designed using NetPrimer (http://www.premierbiosoft.com/netprimer/). Primer sequences are a part of the online Supplementary data at http://zenbase.genereg.net/spp.

In order to increase the throughput for reporter construct generation, we exploited the Gateway Recombination System (Invitrogen). PCRs were performed on human or zebrafish genomic DNA using Advantage2 proof reading PCR system (Clontech), sequenced and successful products cloned into the standard Gateway entry vector (pCR8GW/TOPO vector, Invitrogen) using the manufacturer's recommended protocol. Using the LR recombination, the insert is inserted into the destination enhancer test vector (see below).

Structure of the Tol2-based Gateway™ destination enhancer test vector

The basic enhancer test vector contains the Gateway® C1 cassette, the zebrafish *gata2* promoter (Meng et al., 1997) coupled to the GFP gene and the polyA signal, flanked by Tol2 transposition sequences (Supplementary Fig. 1). The *atp11c*/*ATP11C* and other promoter sequences were amplified from zebrafish or human genomic DNA using BamHI-linker containing primers. The *gata2* promoter was replaced in the *gata2*-GFP Tol2 based vector and the SOX3_hs8A enhancer was cloned upstream of the promoter for testing the interaction of the specific promoter–enhancer combination. A putative CTCF site was inserted instead of the enhancer into the Gateway construct and SOX3_hs8A was cloned upstream subsequently. Genomic coordinates of the promoters and CTCF site are listed in Supplementary Table 1.

DNA and mRNA preparation

Transposase mRNA was synthesized as described (Kotani et al., 2006) except the mRNA purification step, which was performed using a MegaClear kit (Ambion). Capped mRNA was eluted with nuclease-free water. Enhancer test constructs were mixed with transposase mRNA to a final concentration of 25 ng/μl each.

Microinjections and zebrafish handling

For all experiments our wild-type zebrafish stocks were used. Details about fish strains, handling the embryos, maintenance and raising fish are available at http://www.sars.no/facilities/fishRaisingprotocol.doc.

The DNA/mRNA mixture was injected into one-cell stage embryos using glass capillaries made with a Sutter Instruments needle puller. Injected fish were observed at 1 dpf and 2 dpf for GFP expression and then raised to sexual maturity. Detailed description of all procedures can be found in (Fisher et al., 2006). Injected adult fish were outcrossed with wild type individuals and at least 50 F1 embryos were screened with a fluorescent microscope (TE2000-S inverted microscope; Nikon) equipped with 10× and 20× lenses, and a 500/20 nm excitation filter and a 515 nm BP emission filter (Chroma) for GFP. Photographs of live positive embryos were taken using a Spot monochrome digital camera and associated software (Diagnostic Systems). HCNEs were tested in at least 2 independent transgenic lines. Images were processed in Adobe Photoshop by adjusting levels. High-resolution images of the lines in this paper are available at http://zenbase.genereg.net/spp.

Immunohistochemistry, confocal microscopy and 3D imaging

Immunohistochemistry was performed using standard methods with polyclonal rabbit anti-GFP antibody (AMSbiotech (Europe), TP401; used at 1:1000) and alexa-fluor 488 secondary combined with TOPRO-3-iodide (Invitrogen) to label nuclei. Specimens were mounted and imaged using a Leica confocal microscope. Image stacks were processed using Volocity software (Improvision) to produce three-dimensional reconstructions.

Whole mount in situ hybridization and anti-YFP stains

The RNA *in situ* probes for zebrafish bystander genes were amplified from 24 h stage cDNA with the following primers:

immp1lF:GTGGTTTCTTTGTGAAGACGATCAGC,
immp1lR:AGGACTCTCAGCCAATACACCAAAACTC,
rcn1F:TGGAGGTTTCGACCTATTTGTGGC,
rcn1R:TTCCTGGTAAGGTCTTCACCGTAGTTGG,
atp11c_E17f:GAAGAGGCTGCAGAGACCATG,
atp11c_E17r:CTGGTGACGACTCCTGTTTTGAC.

The pax6a probe was amplified using primers available on the ZFIN web page.

The *sox3* and *pax6b* probes were obtained from Drs Laure Bally-Cuif and Terje Johansen, respectively. *In situ* hybridizations were carried out as described in Oxtoby and Jowett (1993) with minor modifications. Embedding of stained embryos and vibratome sections were prepared as in Hadrys et al. (2006). Whole mounts were photographed using a Spot digital camera attached to a Leica MZ FLIIII stereomicroscope, while sections were photographed with a Spot 7.4 Slider camera on a Nikon Eclipse E800 microscope equipped with DIC optics.

Results

SOX3 resides in a gene desert with high HCNE density

The *SOX3* gene is a key transcriptional activator in vertebrate neural and craniofacial development (Bylund et al., 2003; Dee et al., 2008). Human *SOX3* is located in a gene desert extending about 1 Mb at Xq27.1. Mutations in the coding region, and non-coding deletions 67 kb downstream of the gene, are the cause of X-linked mental retardation (Laumonnier et al., 2002) or hypoparathyroidism and hypopituitarism (Bowl et al., 2005), respectively. Our Ancora genome browser (Engstrom et al., 2008) was used to visualize HCNE density and conserved synteny with the zebrafish genome (Fig. 1A). The human *SOX3* region, when compared with the corresponding mouse×chromosomal segment, stands out as a region of high HCNE density (Fig. 1A). There is a single contiguous human: zebrafish conserved synteny block mapping to zebrafish chromosome 14: the block includes *sox3* and the far downstream neighboring genes *atp11c* and *mcf2*. In all vertebrate genomes the *SOX3* region contains numerous HCNEs. In the chicken and frog genomes, the distal part containing *Atp11c* and *Mcf2* has broken off and is detected elsewhere in respective genomes (data not shown), and the minimal conserved synteny extends just across the segment with high HCNE density. We do not know whether this is a result of a true evolutionary event or whether the area might be misassembled in these genomes. There are

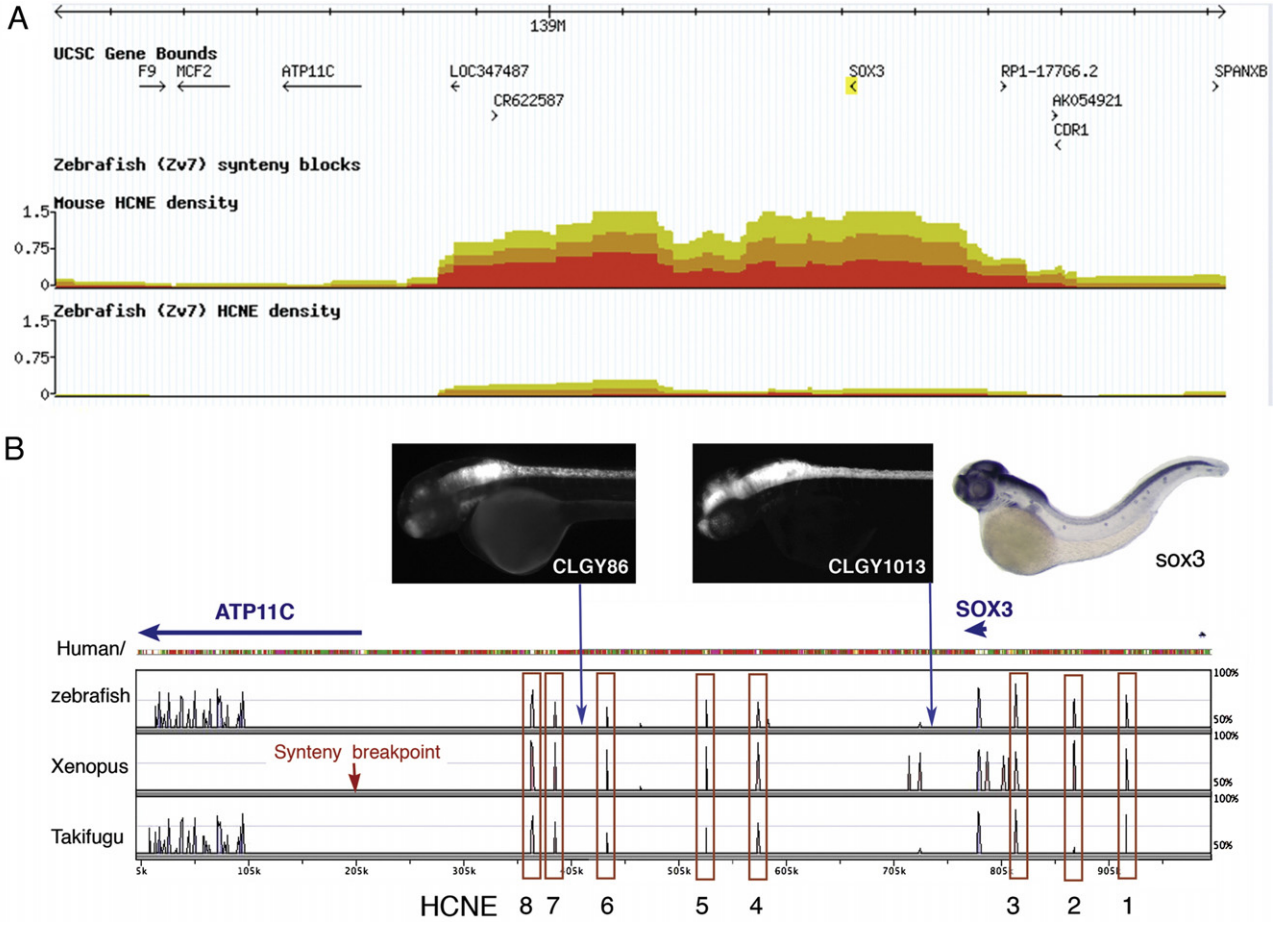


Fig. 1. The overview of genomic locus of *SOX3*. (A) Ancora view of a 2 Mb window across the human locus (upper track), density of HCNEs compared with the mouse genome (lower track) and conserved synteny in the zebrafish genome. (B) 1 Mb window mVISTA plot of human *SOX3* compared with zebrafish (dr, upper track), *Xenopus tropicalis*, (xt, middle track) and *Fugu rubripes* (fr, lower track). HCNEs conserved in all four species were tested in zebrafish and are boxed in red. The insets above tracks show two enhancer detection insertions in the *sox3* gene desert in zebrafish and the RNA *in situ* hybridization pattern of sox3 in a 2-day zebrafish embryo. Expression is mainly confined to the nervous system (for detailed expression domains see also Supplementary online database).

a few annotated genes in this area, but since these are not conserved in other vertebrates nor contain known protein domains, it is likely that they are not protein-coding (Clamp et al., 2007).

Zebrafish sox3 and atp11c are expressed in distinct patterns

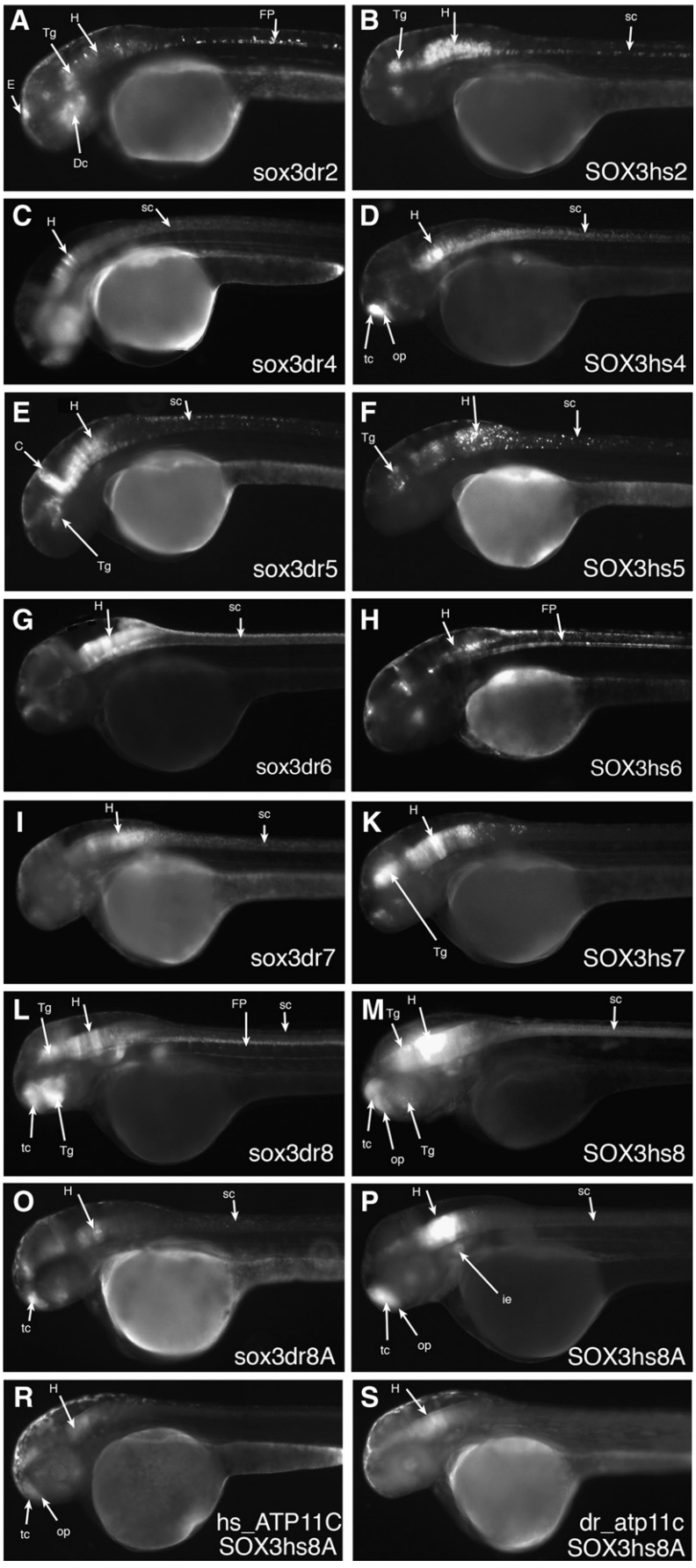
The zebrafish *sox3* gene is initially expressed in the neural plate and ventral mesoderm, and later in the presumptive forebrain, hindbrain, and spinal cord; there is also low-level expression at the forebrain–midbrain and midbrain–hindbrain boundaries (Fig. 1B; for detailed expression analysis on sections, see Supplementary Fig. 2). At 2 days post fertilization (2 dpf), *sox3* is expressed in the central nervous system, where high levels of the transcript are detected mainly in the ventricular zones of the forebrain and hindbrain, as well as in the spinal cord and the lateral line neuromasts. Zebrafish *sox3* is also expressed in the retina, gut, olfactory pits, pharyngeal arches and in epibranchial placodes (Supplementary Fig. 2). Two enhancer detection insertions in the area, one near *sox3* and the other 96 kb downstream (Fig. 1B), have almost identical expression patterns, although the more distal insertion lacks the high level of expression in the midbrain seen in the insertion near the target gene. An unrelated gene, *ATP11C*, is found 672 kb downstream from *SOX3* in the human genome. In contrast to *sox3*, the expression of *atp11c* in zebrafish is ubiquitous at low level, although somewhat higher expression can be seen in the lateral line neuromasts (Supplementary Figs. 3M–Q). Thus, while there is some overlap between the two expression patterns, they differ significantly in terms of restriction and signal strength.

Enhancer activity of SOX3 and sox3 HCNEs

We located eight HCNEs conserved from human to zebrafish genomes across the *SOX3* region at the thresholds we used (red boxes in Fig. 1B), of which one, HCNE 8, corresponds to ultraconserved element 482 (Bejerano et al., 2004). We tested all eight elements (both human and corresponding zebrafish sequences) in our transgenic assay and found six of them to drive reproducible expression patterns and thus likely to be functional enhancers of *SOX3*/*sox3*. HCNE 1 (both human and zebrafish sequences) when tested in several transgenic lines resulted in no observable GFP expression and we therefore assume this to be a negative regulatory element. The function of HCNE 3 remained unresolved as a number of different expression patterns were isolated, but none reproducibly. The patterns of reproducibly positive enhancers (in at least two, but usually three or more independent transgenic lines) were generally related for fish and human orthologous elements tested in transgenic zebrafish, but often with different relative expression strengths and/or extent of expression domains: for example the human sequence of HCNE 2 (SOX3_hs2), drove a broader domain in the hindbrain than its zebrafish counterpart (sox3_dr2) (Figs. 2A, B). Likewise, HCNE 4, denoted as SOX3_hs4, drove much wider expression in the hindbrain than the corresponding zebrafish sequence, sox3_dr4, which labeled specific stripes within the rhombomeres (Figs. 2C, D). HCNE 5, tested also by Pennacchio et al. (2006) in the mouse (as element 667) resulted in a highly similar pattern using the human sequence (SOX3_hs5) to that in the mouse, but sox3_dr5, in contrast, also directed expression to the cerebellum in transgenic zebrafish (Figs. 2E, F). HCNE 6 showed specific activity in roof plate and floor plate in both orthologous sequences (Figs. 2G, H), but there were also expression differences in signal strength in zebrafish vs., human elements (see Figs. 2A, B, G, H, O, P). Notably, many of the tested HCNEs drove expression in the hindbrain, in the spinal cord, in the tegmentum, the hypothalamus and in the telencephalon. An ultraconserved element overlapping with SOX3_hs8 was recently deleted from the mouse genome, with no observable phenotype (Ahituv et al., 2007). Although we do not know the reasons for this surprising finding, it may be possible that the overlapping expression patterns of other HCNEs in the area shown here are able to compensate for the loss of expression resulting from the deletion of this regulatory element. Nevertheless, the highly similar expression patterns driven by the two orthologous sequences suggest, in addition to the ultraconservation observed between human and rodent genomes, that this element is under considerable evolutionary constraint across vertebrates (Figs. 2L, M). Furthermore, SOX3_hs8 is divided into two parts, which show split conservation peaks in the human-fish comparison — SOX3_hs8A and 8B (Fig. 7A). Although not overlapping with the core UCR (Bejerano et al., 2004), part 8A directed highly specific *SOX3* expression patterns (Fig. 2P) whereas part B showed no activity (data not shown). This HCNE is closer to the neighboring *atp11c*/*ATP11C* than it is to *SOX3* (255 kb vs. 416 kb in the human genome). Since *atp11c* is, like *sox3*, expressed in the lateral line we asked whether the promoters of *atp11c*/*ATP11C* would be able to interact with human element SOX3_hs8A. In *Drosophila*, the differential responsiveness of target- and bystander genes within GRBs is likely regulated through enhancer/promoter specificity mediated through different promoter subtypes (Engstrom et al., 2007; Hendrix et al., 2008). Both human and zebrafish promoters were used and tested with the human element. As shown in Fig. 2 (R, S), both *ATP11C*/*atp11c* promoters did interact with SOX3_hs8A and gave transcriptional activity in the correct patterns, but both with strikingly lower activity than when the *gata2* promoter was used. In the human locus there is a CTCF binding site mapped between SOX3_hs8 and *ATP11C* (Kim et al., 2007), but this site is not conserved in zebrafish. Placing the human sequence between SOX3_hs8A and the basal promoter:GFP did not result in any difference in reporter expression strength or pattern (Supplementary Fig. 4H). Thus, at least in this case, the promoter of a neighboring gene can interact with a specific enhancer of the target gene, albeit at low level. However, this was observed by placing the HCNE directly in front of the *ATP11C* promoter, and the effect at native distances may be considerably attenuated. We confirmed independence of the enhancer SOX3_hs8A of minimal promoter sequence by combining it with various types of human and zebrafish minimal promoters including the endogenous *SOX3* promoter. The SOX3_hs8A enhancer was tested with six further promoters to test GFP expression. The promoters chosen were from the human β-globin gene, of the zebrafish *hsp70* gene and from the zebrafish *ngn1* gene. In the injected embryos, all of the combinations resulted in reproducible expression patterns virtually identical to the combination with the *gata2* promoter (Supplementary Figs. 4A–G). The enhancer was proven not to activate GFP by itself (not shown). In summary, the *SOX3*/*sox3* orthologs are embedded in a large area of synteny conserved from human to fish genomes, and conserved *cis*-regulatory elements driving the gene(s) are dispersed over more than 600 kb in the human genome. Gene deserts such as the one described here are large regions with few genes, and multiple regulatory elements, with *SOX3*/*sox3* the only candidate regulatory gene around in this case. Extended regulatory domains exist around other transcription factor genes, but often these regions contain additional protein coding genes (Kikuta et al., 2007a, b). To be able to compare our results to those obtained previously in mouse and zebrafish, we chose as an example the human *PAX6* and zebrafish *pax6a*/*b* regulatory regions.

Human PAX6 resides in a large block with conserved vertebrate gene order

The human *PAX6* gene maps to chromosome band 11p13, close to the border to p14.1; the region contains numerous genes including the Wilms tumor oncogene *WT1* Fig. 3A). The entire area has been conserved from human to fugu genomes (Miles et al., 1998). Unlike in fugu, in the zebrafish genome the *pax6* gene has been retained in two copies after teleost whole genome duplication (Kikuta et al., 2007b; Nornes et al., 1998), which may indicate that the zebrafish and fugu lineages diverged soon after the whole genome duplication in an


A
H
Tg
FP
E
Dc
sox3dr2
B
Tg
H
sc
SOX3hs2
C
H
sc
sox3dr4
D
H
sc
tc
op
SOX3hs4
E
H
sc
C
Tg
sox3dr5
F
H
sc
Tg
SOX3hs5
G
H
sc
sox3dr6
H
H
FP
SOX3hs6
I
H
sc
sox3dr7
K
H
Tg
SOX3hs7
L
Tg
H
FP
sc
tc
Tg
sox3dr8
M
Tg
H
sc
tc
op
Tg
SOX3hs8
O
H
sc
tc
sox3dr8A
P
H
sc
ie
tc
op
SOX3hs8A
R
H
tc
op
hs_ATP11C
SOX3hs8A
S
H
dr_atp11c
SOX3hs8A

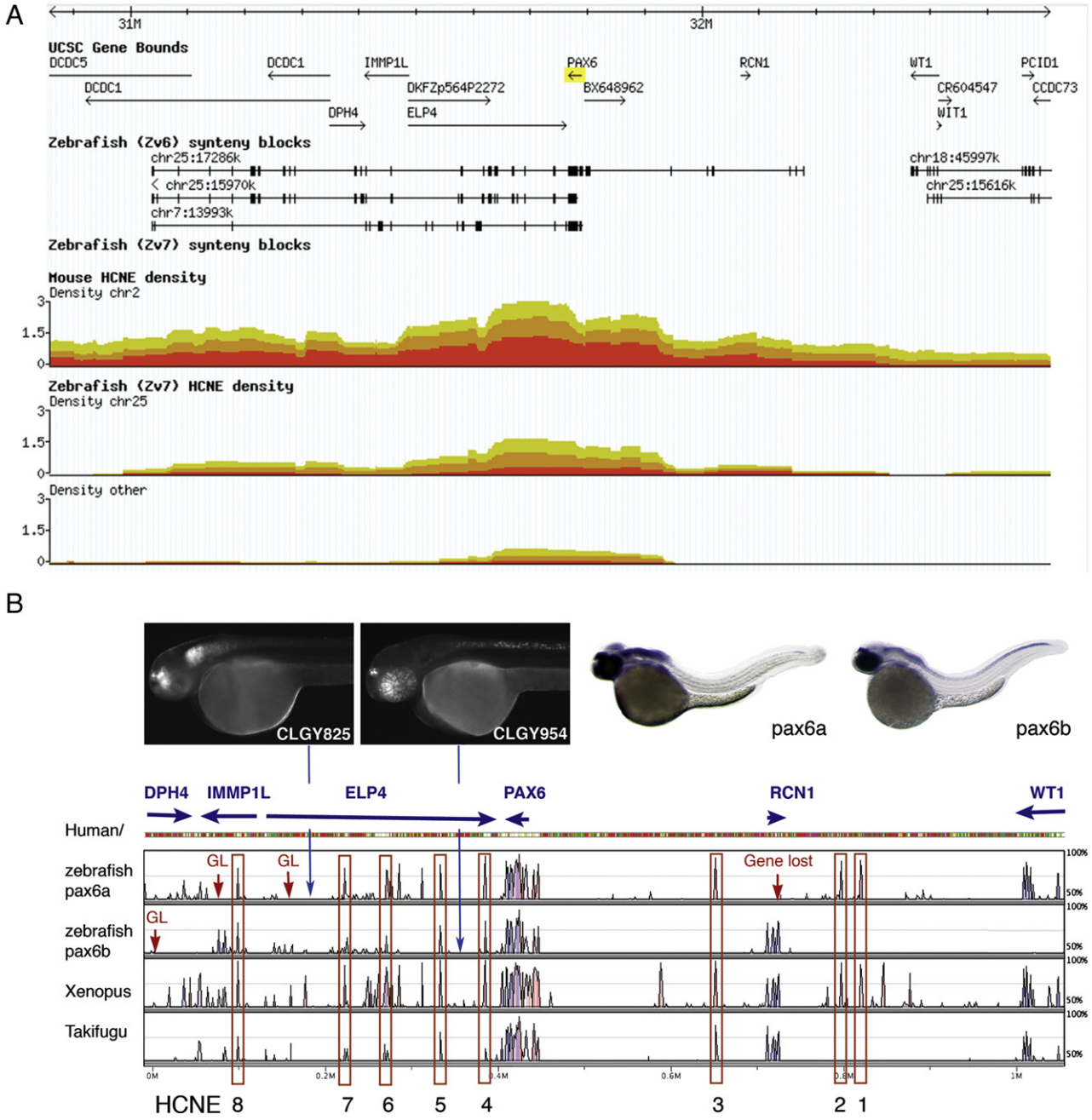


Fig. 3. The genomic locus of *PAX6*. (A) Ancora view of a 5 Mb window across the human locus (upper track), density of HCNEs compared with the mouse genome (lower track) and conserved synteny in the zebrafish genome. (B) 1 Mb mVISTA plot of human *PAX6* compared with zebrafish *Xenopus tropicalis*, (xt, middle track) and *Fugu rubripes* (fr, lower track). HCNEs conserved in all four species were tested in zebrafish and are boxed in red. Insets above tracks show two enhancer detection insertions in the *pax6a* and *pax6b* genomic regulatory block in zebrafish and the RNA *in situ* hybridization pattern of these genes in a 2 day zebrafish embryo. Expression is mainly confined to the nervous system (for detailed expression see also Supplementary data).

ancestral teleost, when the genome was still largely tetraploid. Fig. 3A shows a 5 Mb window across the human *PAX6* region, with mouse HCNE densities annotated on the lower track and zebrafish synteny blocks in the middle track. There are two homologous chromosomal segments in zebrafish, one containing *pax6b* on chromosome 7, and one containing *pax6a* on chromosome 25. Of note, we use synteny information to show that the two synteny blocks overlap in a core region encompassing *PAX6* and the downstream genes *ELP4*, *IMMP1L* and *DPH4* in the human genome, while the distal breakpoints of these two blocks led to the retention of *mpped2* region in only one copy (at the *pax6b* locus), and *wt1* region in only one copy (adjacent to *pax6a*). Thus, there is a duplicated core region containing the *pax6* target gene (s), and two flanking regions that have been kept in only one copy in the zebrafish, one with each copy of *pax6*. The entire human region covered by HCNEs, shown as peaks in the lower track, appears to contain the same set of conserved regulatory inputs for *PAX6* as the

Fig. 2. Lateral views of HCNE-test transgenic zebrafish larvae expressing GFP at 2 days post fertilization. Left column shows transgenic made with zebrafish HCNEs, right column shows transgenes made with the orthologous human sequence. (A–M) Elements depicted in Fig. 1, direct expression of GFP to telencephalon (C, D, L–P), diencephalon (A, B, E, F, L–P), hindbrain (A–P) and spinal cord (A–M). (O, P) Partial sequence of element dr8 and hs8; (R, S) Transgenes using hs8A element but the zebrafish/human *atp11c*/*ATP11C* promoters instead of the zebrafish *gata2* promoter. Note similarity of pattern, but much weaker activation. Abbreviations: h: hindbrain; Tg: tegmentum; E: epiphysis; FP: floor plate; R: rhombomeres; Dc: diencephalon; tc: telencephalon; ie: inner ear; sc: spinal cord; C: cerebellum.

two zebrafish *pax6* genes together, where some of the enhancers were kept in duplicates (HCNE #8), but several have been lost from one copy (#1, 2, 3, 6, 7) (Fig. 3B).

Large-scale heterozygous deletions in this region in the human genome give rise to WAGR syndrome (Wilms tumor, aniridia, genitourinary abnormalities and mental retardation) (Francke et al., 1979), while heterozygous missense mutations in *PAX6* cause aniridia in humans and the small eye phenotype in mouse, respectively (Glaser et al., 1992; Hill et al., 1992; Jordan et al., 1992). Correct expression of *PAX6* is dependent on regulatory elements inside the last intron of the neighboring gene *ELP4*: breakpoints within *ELP4*, which leave the coding sequence of *PAX6* intact, have also been shown to cause aniridia, (Fantes et al., 1995; Kleinjan and van Heyningen, 2005; Lauderdale et al., 2000). It was shown that the phenotype is not caused by loss of *ELP4* function, but rather by loss of *PAX6* expression, thus suggesting that essential regulatory elements driving *PAX6* reside inside *ELP4* (Kleinjan et al., 2001) (see Fig. 3B). In tetrapod genomes, the *PAX6* chromosomal domain encompasses the genes *IMMP1L* and *DPH4* further downstream and the *RCN1* gene upstream. Both paralogous *pax6* gene regions in the zebrafish genome contain HCNEs, many of which are inside introns of the neighboring bystander genes *elp4*, *immp1l* and *dph4*, as well as several elements near and beyond the upstream neighboring gene *rcn1*. As noted by (Kikuta et al., 2007b), these unrelated genes within the *pax6a/b* neighborhood are present in single copy in the zebrafish genome at only one of the two paralogous loci/synteny blocks, while both *pax6* and *wt1* have been retained as duplicates (a and b). However, only *wt1a* remained adjacent to the *pax6a* locus, while *wt1b* is located on chromosome 18 in a small block of conserved synteny. Human (and ancestral) *elp4* and *immp1l* contain HCNEs in their introns (Fig. 3); however, after whole-genome duplication and rediploidization some of them were not kept in introns of the remaining copies of *elp4* and *immp1l* next to *pax6b*, but are rather found in an equivalent location in *cis* to *pax6a*, which has become a gene desert. This suggests that the HCNEs inside *elp4* and *immp1l* are not essential for regulation of these genes, but must be involved in regulation of *pax6a*, a conclusion further corroborated by the finding that enhancer detection insertions in zebrafish in this region assume a *pax6* expression pattern (Fig. 3B); (Kikuta et al., 2007b). Of the two genes further beyond *IMMP1L* in the human genome, *DPH4* and *DCDC1*, the former is found to contain non-coding elements in a human-chicken alignment (data not shown), whereas the latter is present only in mammals: the fact that the HCNEs in *DCDC1* introns are conserved between human and chicken indicates that the gene has been lost from other vertebrate genomes, rather than acquired in the mammalian lineage. Our synteny and HCNE density approach shows that the corresponding GRB extends to additional HCNEs conserved between human and zebrafish beyond the location of *DCDC1* in the human genome, extending the *PAX6*/*WT1* conserved region to more than 3 Mb. While the *ELP4* has been shown to harbor *PAX6* enhancers previously (Kleinjan et al., 2006, 2008; Woolfe et al., 2005), we show that *PAX6* enhancers are distributed over a larger than expected region and reside in several neighboring

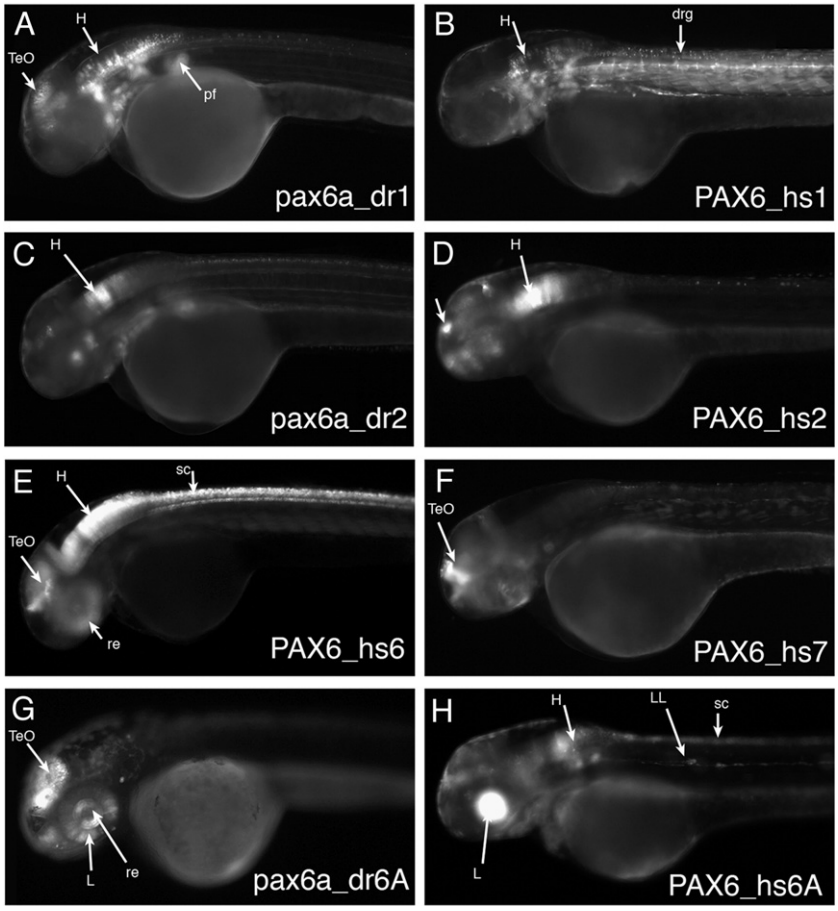


Fig. 4. Lateral views of HCNE-test transgenic zebrafish embryos at 2 days post fertilization. Transgenes are labeled by locus name (*pax6a/b/PAX6*), organism (zebrafish=dr/human=hs) and HCNE number. (A, B) Show transgenes made with zebrafish HCNE 1 expressing GFP in optic tectum, pectoral fins and hindbrain vs. human element driving expression in dorsal root ganglia and hindbrain, which indicate its functional connection to WT1 expressed in these domains. HCNE 2 in (C, D) directs expression mainly to hindbrain with both zebrafish and human elements. Embryos in (E) made with human sequence PAX6_hs6 express in optic tectum, tegmentum, hindbrain, spinal chord and retina, embryo (F) expressing in dorsal diencephalon (PAX6_hs7). Sub-sequence pax6a_dr6A (as seen in Fig. 6B) drives expression only in optic tectum and retina but the human homologous PAX_hs6A element directs expression to optic tectum, hindbrain, spinal chord, lateral line and lens.

genes, and that the location of *PAX6* enhancers inside *ELP4* is not an exceptional occurrence.

Zebrafish pax6a/b and bystander genes have distinct expression patterns

We asked whether the bystander genes of *pax6* could themselves be regulated by the HCNEs they contain. We compared the expression patterns of the two *pax6* genes (Fig. 3B, Supplementary Fig. 2) and the bystander genes *immp1l* and *rcn1* (Supplementary Fig. 3), as well as the previously investigated *elp4* (Kikuta et al., 2007b). None of these bystander genes showed a pattern resembling either of the two *pax6* paralogs. Moreover, compared to *pax6a/b* these bystander genes were expressed at relatively low level, as estimated through the much longer time required to develop the RNA *in situ* hybridization signals. The expression of the *immp1l* and *elp4* genes is widespread at the assayed stages, while *rcn1* is also expressed in cartilage. In contrast, both *pax6a* and *pax6b* expression is highly restricted mainly to the CNS; both of the genes are expressed in the retina, midbrain, hindbrain, epiphysis and diencephalon. Notable differences in expression between the two paralogous genes, seen in both enhancer detection insertions in either locus and by RNA *in situ* hybridization (Fig. 3B; Supplementary Fig. 2) are that *pax6a* is expressed at higher levels in the optic tectum and telencephalon, and only at low levels in the anterior spinal cord, whereas *pax6b* is expressed more strongly in the entire spinal cord, and also in the pancreas. The observed differences are presumably the result of subfunctionalization after selective loss or mutation of *cis*-regulatory elements in either locus (Canestro et al.,

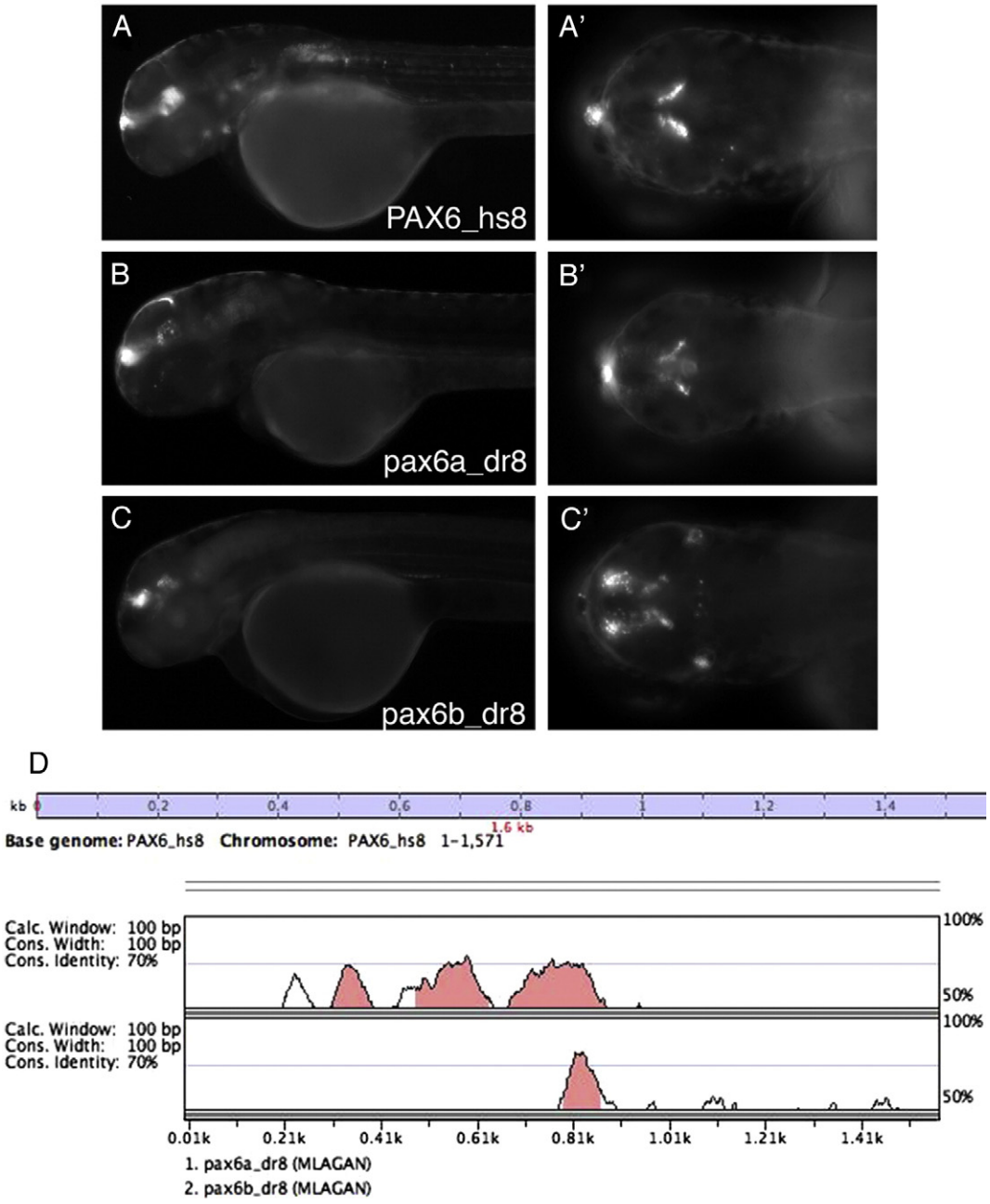


Fig. 5. Expression driven by PAX6 HCNE8, human and both zebrafish orthologous sequences. (A–C′) All sequences drive expression to pineal and diencephalon. Left column shows lateral views of 3 days old embryos, right column dorsal views of heads of the same larvae. The signals differ in relative intensity in these structures between human *PAX6/pax6a/pax6b*. (D) VISTA conservation plot of the alignment of zebrafish pax6a_dr8 and pax6b_dr8 with PAX6_hs8 (baseline) indicating asymmetrical divergence of the zebrafish duplicates. Common peak indicates core HCNE.

2007; Force et al., 1999; Kikuta et al., 2007b; Kleinjan et al., 2008; Woolfe and Elgar, 2007). Furthermore, additional HCNEs driving specific aspects missing in either paralog may hide in the regions outside of the core region of overlapping synteny blocks. In conclusion, there is a marked difference in the pattern and strength of expression between *pax6a/b* and the neighboring bystander genes, and even though one might argue that the expression patterns of target genes and bystander genes overlap due to the widespread expression of the latter, a more likely explanation is that the bystanders are expressed constitutively at low levels throughout development and are therefore not regulated specifically by HCNEs in the region.

HCNEs in the PAX6/pax6a/b GRBs drive pax6-like patterns

We tested HCNEs from the broader *PAX6/pax6a/b* region for their ability to direct GFP expression in transgenic assays. The tested elements are conserved from human to zebrafish and are indicated by red boxes in Fig. 3B. HCNEs 6 and 7 are found within the bystander gene *ELP4*. HCNE 8, which is in an intron of *IMMP1L* in the human genome, has been kept in duplicate in the zebrafish, even though *immp1l* itself has disappeared downstream from *pax6a*. HCNEs 1, 2 and 3 are located upstream of *PAX6*, and out of which elements 1 and 2 are located beyond *RCN1* (Fig. 3).

The pattern driven by HCNE 6 was shown by Kleinjan et al. (2001) in the mouse (PAX6_hs6 overlapping element EI/Z), while PAX6_hs7 (Fig. 4F) corresponds to element RB tested in mouse by (Kleinjan et al., 2006). PAX6_hs6, PAX6_hs7, PAX6_hs2 and PAX6_hs8 (Figs. 4E, F, D and Fig. 5) correspond to elements tested by (Pennacchio et al., 2006) — these are elements 234, 565, 113 and 863 respectively. Transgenes in zebrafish using these HCNEs resulted in GFP patterns in the mouse equivalent to the patterns in the above publications. The human HCNEs PAX6_hs6A, PAX6_hs1 and PAX6_hs3 and their

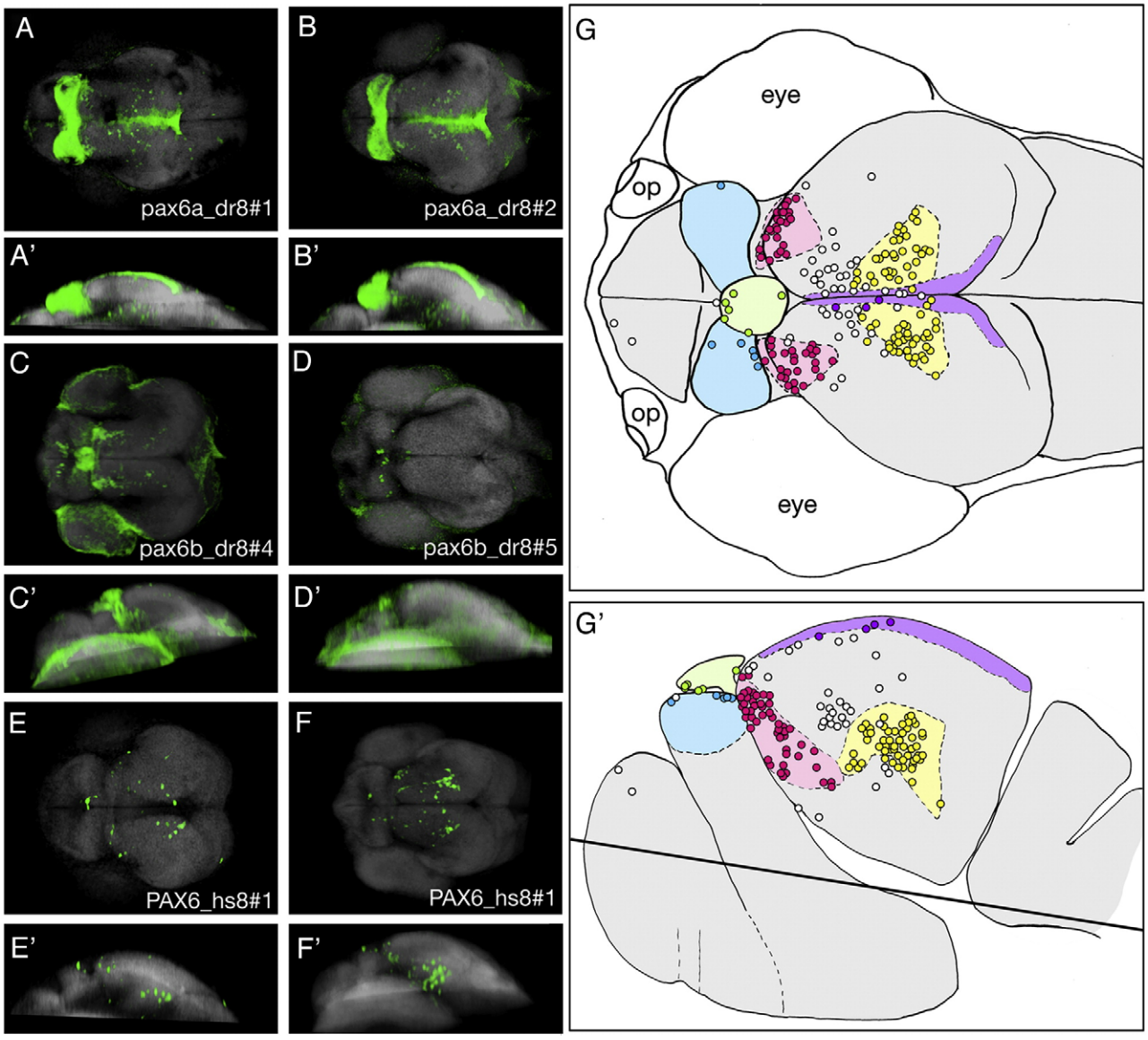


Fig. 6. Confocal study of PAX6 HCNE8. (A–F′) 3D reconstructions of dorsally-imaged confocal stacks of 3 dpf embryos, transgenic genotype indicated top right, stained using immunohistochemistry for GFP (green) combined with TOPRO nuclear staining (grey). Viewed from dorsal (A, B, C, D, E, F) and lateral (A′, B′, C′, D′, E′, F′) aspects. Areas of GFP expression in embryos bearing GFP transgenes driven by zebrafish *pax6a* (A, A′, B, B′) and *pax6b* (C, C′, D, D′) HCNE 8 were mapped onto schematic diagrams of the brain of a 3 dpf embryo drawn from dorsal (G) and lateral (with eye removed; G′) viewpoints as coloured areas: pineal=green, habenulae=blue, pretectal area=red, dorsal (roofplate) tectum=purple, dorsal tegmentum=yellow. 3D reconstructions were used to plot the locations of GFP-positive cells for multiple PAX6_hs8 embryos (n=6) (examples E, E′, F, F′) using coloured circles. Where GFP-positive cells fell within the zones of either zebrafish HCNE 8-driven transgene expression the circle was coloured the equivalent colour. Where they fell outside these areas they were left white. The lateral schematic (G′) represents summated expression from both left and right sides of the brain, the black line across the image represents the depth of the 3D reconstructions employed (op=olfactory pits).

corresponding zebrafish sequences have not been tested before. PAX6_hs6A is a sub-sequence of PAX6_hs6 (Fig. 7B), and its orthologous element from zebrafish drove expression in distinct patterns (lens expression in human vs. retina, telencephalon and optic tectum in zebrafish pax6_dr6A) that were consistent with *pax6* regulation (Figs. 4G, H).

A duplicated zebrafish enhancer and a human enhancer driving the same pattern in zebrafish

The human and both paralogous zebrafish sequences of HCNE 8 drove expression of GFP in the diencephalon and pineal complex in which *pax6* is expressed, and showed signs of both conserved function and subfunctionalization (compare expression by the human element, PAX6_hs8, to the element pax6a_dr8, and to the duplicate pax6b_dr8); (Fig. 5).

To compare expression areas in more detail between embryos expressing GFP driven by zebrafish and human *pax6* HCNEs we examined 3D reconstructions of the brains of confocally-imaged specimens immunostained for GFP and counterstained with a nuclear dye (TOPRO) (Fig. 6) to determine precisely the exact areas of GFP expression. The zebrafish *pax6a* HCNE 8-driven GFP transgene (pax6a_dr8; Figs. 6A–B′) gave strong expression in the left and right habenulae, some pretectal areas and the extreme dorsal tectum (roofplate); there was also a lower level of expression in the dorsal tegmentum of the midbrain. GFP expression in *pax6b* HCNE 8-driven transgenics (pax6b_dr8; Figs. 6C–D′) was strong in the pineal and the pretectum, there was also weak expression in the medial habenulae and in some cells of the dorsal tegmentum (in some transgene integrations; e.g. Figs. 6C, C′). These distinct expression patterns demonstrate some subfunctionalization between the two zebrafish *pax6* HCNE8 elements although there is some overlap, notably in the pretectum and in the dorsal tegmentum.

GFP expression in the human HCNE8 transgenics (PAX6_hs8; Figs. 6E–F′) was more variable between individuals than the zebrafish HCNE8-lines with GFP-positive cells being more widely spread apart; however, individual GFP-positive cells seemed to be present within the areas of GFP expression found in the zebrafish *pax6a* and *pax6b* HCNE8-driven lines. To determine if this general impression was valid we plotted the position of GFP-positive cells in several PAX6_hs8 embryos onto schematic diagrams where the areas of GFP expression in the zebrafish HCNE8 lines had been demarcated (Figs. 6G and G′). This generated a summated GFP expression pattern for the human HCNE8-driven line. Summation of the expression in this way showed robustly that the majority of GFP-positive cells in the PAX6_hs8 line fell within the areas of GFP expression in the pax6a_dr8 and pax6b_dr8 lines. This being so, not all the areas were densely populated by GFP-expressing cells: the dorsal tegmentum and pretectum had many GFP-positive cells in the PAX6_hs8 line but the pineal and dorsal tectum bore only a few and the habenulae had even fewer. Since the dorsal tegmentum and pretectum are areas that were found to be common between the two zebrafish *pax6* lines this suggests that the sequences driving subfunctionalization should be distinct between the two zebrafish HCNEs but there should be sequence that is common between all three forms.

On the sequence level, there is a short core segment shared by all three elements 8 (Figs. 5D, 7D — note the zebrafish chained alignment track). Thus, while differential loss of HCNEs, either through mutational events or through differential extent of the two *pax6* GRBs contributes to subfunctionalization, these results also demonstrate subfunctionalization through sequence divergence, as hypothesized by Woolfe and Elgar (2007). Notably, although this HCNE is kept in duplicate in the zebrafish genome, and therefore might be expected to be under lower evolutionary constraint than single copy elements, GFP expression driven by PAX6_hs8 compared to the two zebrafish pax6a_dr8 and pax6b_dr8 showed areas of common expression.

Despite high levels of conservation, transgenes testing HCNEs 3, 4 and 5 resulted in variable expression patterns, often unrelated to any of the genes in the region and therefore were deemed inconclusive. HCNE 4, overlapping with ultraconserved element uc198, has been tested in mouse with the same unspecific result by Visel et al. (2007). However, a larger mouse sequence, including a neighboring conserved element, was tested by Kleinjan et al. (2008), who also demonstrated loss of one zebrafish copy of the enhancer after teleost whole genome duplication. Unlike the selectively lost HCNE, tested in their publication, sequences orthologous to HCNE 4 were kept in both pax6a and pax6b loci, but showed no specific activity in our assay. HCNEs 3 and 5 correspond to element PAX6_1 and PAX6_19 tested in transient assays by Woolfe et al. (2005). We generated 4 and 2 independent transgenic lines with human element 3 and zebrafish element 3, respectively, and 4 transgenic lines each with human element 5 as well as with both zebrafish duplicates of this element and found neither consistent patterns nor in fact any overlap with the patterns for PAX6_1 and PAX6_19 in Woolfe et al. (2005). Since these authors used the orthologous elements from the Fugu genome, it is possible that these elements direct different patterns, but regardless of this possibility it is clear that transient testing is unable to distinguish between elements that result in reproducible patterns in multiple transgenes and those that do not and yield ectopic expression patterns.

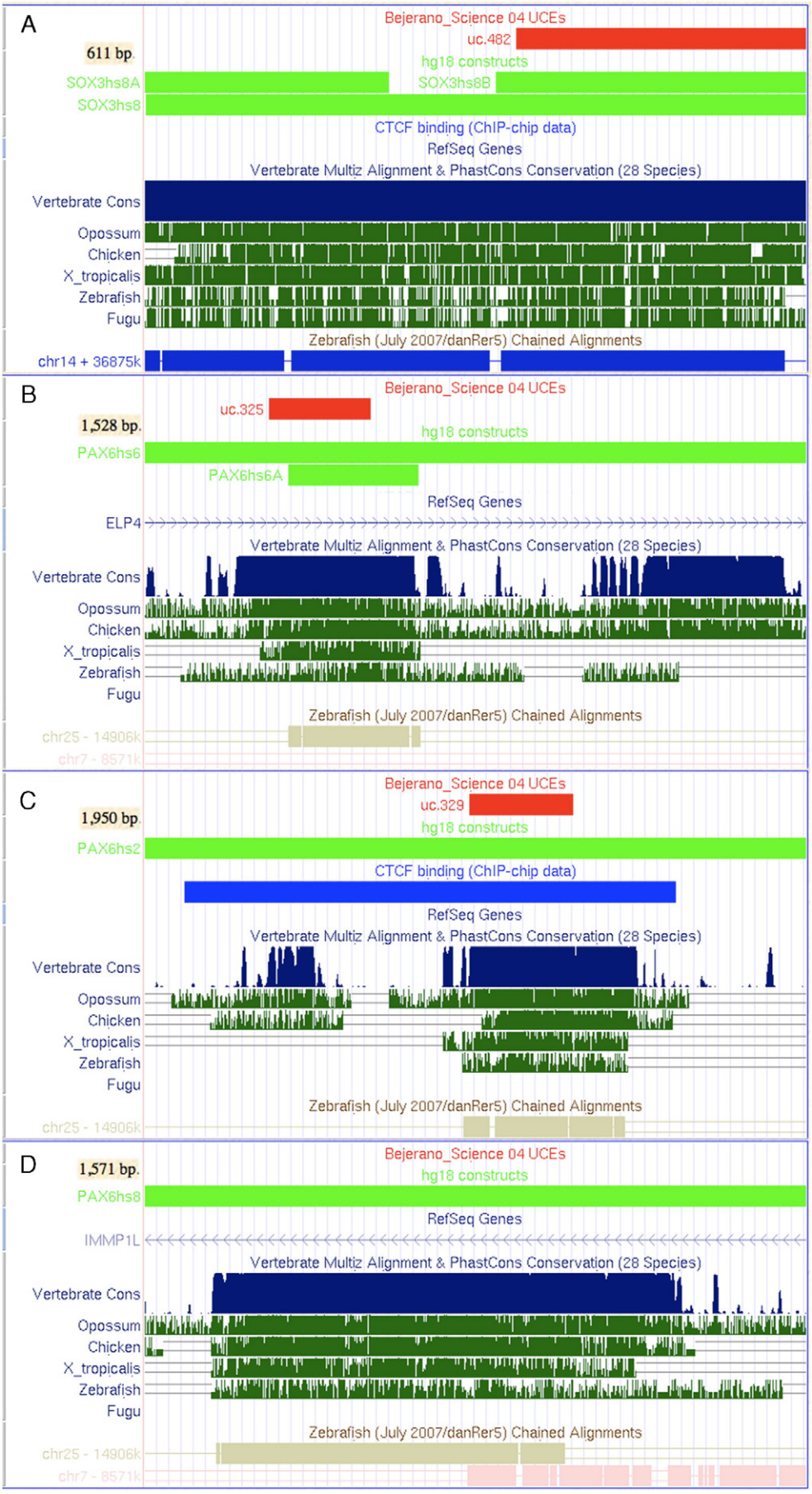
WT1 is controlled separately from PAX6

Human element PAX6_hs2 corresponds to a mammalian conserved CTCF binding site, mapped by ChIP-chip in human fibroblasts by Kim et al. (2007). Surprisingly, although thought to be functioning as an insulator, this element had enhancer activity consistent with regulation of *pax6*, directing reporter expression to the hindbrain (Figs. 4C, D). This element also overlaps with ultraconserved element uc329, recently deleted from the mouse genome with no discernible resulting phenotype (Ahituv et al., 2007) (Fig. 7C). PAX6_hs1, on the other side of the boundary element towards the *WT1* oncogene, drove expression in the dorsal root ganglia (Fig. 4B), which is part of the *Wt1* pattern in mouse (Dame et al., 2006) — thus this element appears to regulate *WT1*, not *PAX6*. However, there are also other mapped CTCF sites distributed throughout the *PAX6* GRB, but from our data there is no clear indication that these function as boundary elements, since HCNEs beyond them still harbor activity that is consistent with being part of the *pax6a/b* expression pattern.

In conclusion, synteny and the distribution of noncoding conservation mark the presence of *cis*-regulatory elements of the *PAX6* genomic regulatory block located within bystander genes well beyond the previously described *ELP4*. The association of these elements with *PAX6* is seen both through their evolutionary segregation patterns (i.e. they are present despite the absence of the bystander gene) as well as through their specific patterns of expression in transgenic assays. This analysis shows that the region inhabited by long-range regulatory elements is matched by conserved synteny at this locus and extends the regulatory domain of *PAX6* to *IMMP1L*, several hundred kb away from the *PAX6* coding region, and further. The entire region in the human genome, which includes the *PAX6* and *WT1* GRBs, spans over three megabases and demonstrates the utility of using human: zebrafish conserved synteny and HCNE densities across the human *PAX6* region for GRB delineation.

Discussion

Developmental regulatory transcription factors are usually involved in multiple regulatory processes (Carroll, 2008) and for this reason are surrounded by a large number of regulatory inputs, many conserved across large evolutionary distances and evident through comparative genomics.


A
611 bp.
Bejerano_Science 04 UCEs
uc.482
hg18 constructs
SOX3hs8A
SOX3hs8B
SOX3hs8
CTCF binding (ChIP-chip data)
RefSeq Genes
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Vertebrate Cons
Opossum
Chicken
X_tropicalis
Zebrafish
Fugu
Zebrafish (July 2007/danRer5) Chained Alignments
chr14 + 36875k
B
1,528 bp.
Bejerano_Science 04 UCEs
uc.325
hg18 constructs
PAX6hs6
PAX6hs6A
RefSeq Genes
ELP4
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Vertebrate Cons
Opossum
Chicken
X_tropicalis
Zebrafish
Fugu
Zebrafish (July 2007/danRer5) Chained Alignments
chr25 - 14906k
chr7 - 8571k
C
1,950 bp.
Bejerano_Science 04 UCEs
uc.329
hg18 constructs
PAX6hs2
CTCF binding (ChIP-chip data)
RefSeq Genes
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Vertebrate Cons
Opossum
Chicken
X_tropicalis
Zebrafish
Fugu
Zebrafish (July 2007/danRer5) Chained Alignments
chr25 - 14906k
D
1,571 bp.
Bejerano_Science 04 UCEs
hg18 constructs
PAX6hs8
RefSeq Genes
IMMP1L
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)
Vertebrate Cons
Opossum
Chicken
X_tropicalis
Zebrafish
Fugu
Zebrafish (July 2007/danRer5) Chained Alignments
chr25 - 14906k
chr7 - 8571k

Using two prominent vertebrate genes of this category, we have demonstrated several important principles regarding the organization of GRBs and their regulation. These genes and, by extension, other genes of this functional category are embedded in large chromosomal regions containing highly conserved non-coding regulatory elements and represent target genes of those elements. Some of these regions additionally encompass neighboring bystander genes in conserved synteny blocks in sequenced vertebrate genomes, in which they are intertwined with HCNEs that regulate the target gene. The bystander genes have sequences, expression patterns, and biological functions unrelated to those of the corresponding target genes and are typically non-pleiotropic. In contrast to regulatory landscapes, where locus control regions maintain conserved synteny through co-regulation of two or more genes by one or more enhancers (Spitz et al., 2003; Zuniga et al., 2004) it is important to note that the GRBs contain regulatory elements whose primary role is to regulate the target gene, while the bystander genes are either not regulated by them, or any regulatory effect by these elements is without consequence: after the whole genome duplication in teleosts, the duplicate copies of the bystander genes in the *pax6a*/*b* regions disappeared in the rediploidization process in zebrafish, leaving behind the HCNEs that formerly resided in their introns (Kikuta et al., 2007b). Since these HCNEs are putative *cis*-regulatory elements, we asked whether in the case of *pax6* and *sox3* they drive expression in patterns that are characteristic of the GRB target gene. We found that there are elements that have reproducible expression patterns that are consistent with being sub-patterns of the target gene irrespective of whether the zebrafish or the human elements are used in transgenic zebrafish, and that these patterns are characteristic for the tested element. In this context it is also noteworthy that the different characteristics of enhancers (i.e. reproducible and target gene-specific vs. variable and ectopic) might indicate that the latter are not standalone inputs but modify or are modified by neighboring elements to achieve a precise expression pattern in specific spatiotemporal contexts. A way to address this in the future will be through motif search and mutational and combinatorial assays to understand how these sequences act to regulate their target gene (Rastegar et al., 2008).

Comparison of expression patterns of human elements tested in both zebrafish and mouse

A subset of the elements tested in our study has been previously tested in other experimental systems. For the *SOX3* locus, human elements SOX3_hs5 and SOX3_hs8 were tested in mouse (Pennacchio et al., 2006; Visel et al., 2007). Whereas SOX3_hs5 showed a broader expression in the mouse CNS than the pattern we reproducibly recovered in zebrafish, the SOX3_hs8 – also corresponding to an UCR (uc482) – showed essentially the same pattern (telencephalon, rhombencephalon and spinal cord) in zebrafish as in mouse. Of the *PAX6* elements tested in this paper, PAX6_hs2, PAX6_hs6, PAX6_hs7 and PAX6_hs8 were tested in the mouse (Kleinjan et al., 2001, 2006; Pennacchio et al., 2006; Visel et al., 2007). Notably, PAX6_hs6 corresponds to EI/Z tested by Kleinjan et al. (2001) with essentially the same expression pattern in mouse that we found in zebrafish. Element PAX6_hs7, which corresponds to element RB in Kleinjan et al. (2006) also directed the same pattern as in the mouse. Finally, PAX6_hs8, which we showed to be duplicated in zebrafish and which is located in *IMMP1L*, also had a corresponding pattern in mouse (Pennacchio et al., 2006; Visel et al., 2007). Thus, human *cis*-regulatory elements that gave consistent patterns in transgenic mice resulted in essentially the same patterns in transgenic zebrafish, suggesting that zebrafish is a suitable system to test the function of human regulatory sequences in this category. Since gene regulatory networks are bound to have evolved since the divergence of man, mouse, and fish, we can only infer that the patterns driven by these elements would be similar in human embryos, but the finding that patterns are so similar when the human elements are tested in the mouse indicates that the information gleaned from our experiments can serve as a first approximation for the activity of these elements.

SOX3 is embedded in a gene desert containing numerous HCNEs with overlapping enhancer activities

SOX genes encode HMG domain-containing transcriptional regulators, with 20 members of this gene family in the human genome (Kiefer, 2007). SOX genes are often located in gene deserts, e.g. (Gomez-Skarmeta et al., 2006). Uchikawa et al. (2003) isolated and tested regulatory elements of *Sox2* in chicken and found regulatory elements over a 50 kb interval around the gene. Our analysis of the *SOX2* gene desert in the human genome suggests that the genomic regulatory block of this gene extends over 1 Mb (Supplementary Fig. 5). In this paper we have shown that out of eight regulatory elements of *SOX3* conserved between human and zebrafish and dispersed over more than 500 kb, six drive expression in overlapping domains of the zebrafish CNS — an observation in agreement with the study of Werner et al. (2007) for regulatory elements of *Sox10* in the mouse. This seeming "redundancy" of enhancer function, hypothesized to ensure robustness of gene expression (Werner et al., 2007), appears to be a common feature of Sox gene regulation; however, it can also serve to enable complex regulatory interactions of enhancers (see Ertzer et al., 2007; Rastegar et al., 2008; and below) that refine spatiotemporal expression of the target genes beyond the level achievable by single elements.

Human enhancer SOX3hs8A is able to activate different core promoters

Considering the unrelated expression patterns and functions of the SOX3_hs8A target gene and its bystander *ATP11C*, as well as the relative proximity of this element to the bystander vs. target gene, we expected a mechanism preventing inappropriate activation of the *ATP11C*. However, this enhancer activated a range of heterologous minimal promoters in a *SOX3*-specific pattern, which was confirmed by combination with the endogenous human *SOX3* promoter. The promoters we tested fall into both categories of TATA-box containing, single TSS, and CpG island-overlapping, multiple TSS types (Carninci et al., 2006). The SOX3_hs8A element was also able to activate expression from the *ATP11C* promoter, but with a much lower intensity. No insulating sequences between the *SOX3* and *ATP11C* were found, and the only CTCF binding site detected in this region by a genome-wide screen (Kim et al., 2007) did not block enhancer–promoter interaction. We cannot exclude promoter competition or extremely weak activation of the bystander gene by this enhancer. Additionally, standard reporter constructs test the activity of enhancers cloned immediately in front of the promoter: based on currently available data, we are not able to judge the ability of different kinds of promoters to respond to enhancers from the extreme distances encountered in GRBs.

The PAX6 regulatory domain spans several adjacent genes and is functionally separate from the adjacent WT1 regulatory domain

By examining minimal conserved synteny and HCNE density across human and teleost genome (Fig. 3A) we determined that the *PAX6* GRB encompasses the upstream *RCN1* gene, as well as the

Fig. 7. Screenshots from UCSC browser on human hg18 assembly. Individual tested elements shown as green bars, CTCF binding sites (Kim et al., 2007) in blue and ultraconserved sequences (Bejerano et al., 2004) behind red bars. Visualized tracks show RefSeq gene annotation, conserved transcription factor binding sites, and vertebrate conservation plot together with 7 individual species plots. Bottom bar indicates human–zebrafish conserved regions.

downstream genes *ELP4*, *IMMP1L*, and *DPH4*. Even though the more distally located gene *DCDC1* is not conserved in non-mammalian vertebrates, an alignment of the region in human and chicken shows that the introns of human *DCDC1* gene contain several elements that are conserved in the chicken genome. Since *DCDC1* belongs to a protein family with multiple members, most of which are present in all vertebrates, it could not have been created de novo nor could its exons have gotten intertwined with the HCNEs already present in the region. The only plausible explanation for the current configuration is that the gene has been lost from the analyzed non-mammalian genomes. Both conserved synteny and the occurrence of HCNEs suggest that the *PAX6*/*pax6a*/*b* area, including the genes listed above, is held together by HCNEs that regulate their target gene over several gene distances and in that way maintain conserved synteny across this locus. The observation that *wt1b* is found in a synteny block on chromosome 18 and has been removed from *pax6b* suggests that *WT1* is also a target gene and that the 3 Mb synteny block in 11p13 containing *PAX6* and *WT1* in the human genome consists of two conjoined, but separable, GRBs. A further point to be made is that all the bystander genes in the region rediploidized by neutral evolution. An obvious explanation is that these genes do not have developmental regulatory function, but rather have little regulation at the transcriptional level, and as such have small amounts of sequence dedicated to receiving regulatory inputs. This property would make them less prone to regulatory subfunctionalization. Several HCNEs (e.g. HCNEs 5 and 8) have been kept in duplicate (Fig. 3B). For one of them, pax6a/b_dr8, we have shown evidence for subfunctionalization: even though the two elements are still similar at the sequence level, the differences are sufficient to subtly change the domain in which they drive expression (Fig. 5). It is important to note that, in the case of the *PAX6*/*pax6a*/*b* GRB, HCNEs are distributed over a megabase area including the introns of all the bystander genes, a feature that can predict that these elements are likely involved in pax6 regulation, and this should also be true for similar cases. For instance, Kimura-Yoshida et al. (2004) tested HCNEs conserved around the human and fugu *OTX2*/*otx2* genes in transgenic mice and found seven elements in a 60 kb interval (in fugu). A quick inspection of the region using the Ancora browser (Engstrom et al., 2008) shows that the human *OTX2* region extends over more than 1 Mb, and human:zebrafish conserved synteny includes five neighboring genes, most with additional HCNEs in their introns (Supplementary Fig. 5). This synteny block corresponds to 150 kb in fugu, more than twice the span of sequence searched by Kimura-Yoshida et al. (2004). The pattern of HCNEs spilling out into the introns of genes that neighbor developmental control genes, specifically those encoding transcriptional regulators, is common across vertebrate genomes (Lindblad-Toh et al., 2005; Sandelin et al., 2004). We propose that this arrangement, with numerous GRBs in the human genome where bystander genes contain HCNEs regulating a distant target gene, has profound significance for the mapping of human disease mutations.

Sequence- and functional similarity of pan-vertebrate HCNEs

The HCNEs tested in this study are conserved from human to zebrafish genomes, as seen from their sequence similarity and equivalent genomic location. When tested in zebrafish, both human and zebrafish HCNEs showed similar activities in several respects: generally, if an HCNEs of one species directed reproducible expression pattern the orthologous HCNE from the other species tended to do the same; the same was true for elements which did not, despite recognizable conservation at the sequence level, direct any reporter expression and were deemed negative elements (as opposed to those that resulted in variable patterns and were deemed inconclusive, which, in the case of *SOX3*/*sox3*, was also assigned to the same element, HCNE 3). In addition, the expression patterns driven by orthologous elements were remarkably similar, often directed to the same brain regions, despite the large evolutionary distance between human and zebrafish readily apparent from Fig. 2. In *Drosophila* species, for example, enhancers can exhibit considerable functional differences across relatively short evolutionary distances (Ludwig et al., 2005). Moreover, the same authors have also observed that the even skipped stripe 2 enhancer precisely retains its regulatory activity between *D. melanogaster* and *D. pseudoobscura*, although these species diverged between 40 and 60 MY ago (Ludwig et al., 2000). Thus both the *Drosophila* and our study suggest that enhancer sequences, although under sequence constraint, can in some instances subtly change the expression pattern they drive while keeping high overall sequence similarity, or, in other cases, change their sequence to a significant degree and nevertheless drive the same expression pattern. Nevertheless, it is evident that HCNEs conserved across all vertebrates are recognizably similar with respect to function as assayed in this study.

CTCF binding sites in SOX3 and PAX6 GRBs

The *SOX3* gene desert is bounded by clusters of CTCF binding elements: one in between the element SOX3hs8 and the neighboring *ATP11C* gene, and the other far upstream of *SOX3* and close to the next gene, *CDR1* (Kim et al., 2007). However, the former is not conserved in zebrafish nor did the human sequence show insulator activity in transgenic assays. In the case of the adjacent GRBs of *PAX6* and *WT1*, there is a CTCF binding site that appears to mark a boundary between their respective GRBs, which is supported by the reporter expression patterns driven by HCNEs from the two sides of the boundary. However, apart from this case, the significance of these elements with respect to the *PAX6* GRB is far from obvious, as there are three additional CTCF binding sites between *RCN1* and *PAX6* and one inside *ELP4*, and expression patterns obtained with the HCNEs described here point that neither of these sites seems to insulate *PAX6* from long-range influences from the other side. However, there are also indications that *PAX6* is repressed by *CTCF* (Li et al., 2006), and it is therefore possible that this could be mediated through the mapped sites. We could not detect any conservation of these reported CTCF motifs between human and zebrafish. Thus, while it is tempting to believe that GRBs are bounded by such elements and thus are insulated from influences from HCNEs outside of them, there are CTCF sites interspersed between them that may act conditionally and their function is still unexplained.

The evolution of GRBs: from in silico patterns to experimental observations to general principles

We have proposed previously that the majority of gene deserts are functionally equivalent to GRBs that contain bystander genes, such as the *PAX6* GRB (Kikuta et al., 2007a). The finding that bystander genes are lost after duplication of a given GRB points to the possibility that this is the origin of gene deserts: repeated duplication can eventually lead to the complete loss of all bystanders, leaving behind areas populated by HCNEs and a single target gene. This would in turn lead to the question of how GRBs evolved in the first place and how HCNEs came to be situated inside neighboring genes. One possibility would be that HCNEs were inserted into the introns of the bystander genes through mobile elements, a possibility supported by the finding that subsets of regulatory elements are descended from retroposons or retroviruses (Bejerano et al., 2006; Lowe et al., 2007; Wang et al., 2007; Xie et al., 2006). If this were the major mode of GRB evolution, the insertion of elements that later developed into regulatory modules would have to be ancient, before the radiation of the vertebrates, since otherwise one would not expect the observed conservation, of conserved synteny around developmental regulatory genes from human to fish, and even to the very root of chordates (Putnam et al., 2008). Indeed, there is a

growing body of evidence that GRBs are present in most metazoan genomes (Engstrom et al., 2007; Vavouri et al., 2006). Regardless of the evolution of GRBs, it is tempting to ask what advantage such an arrangement would have, given that GRBs are such a pervasive and successful feature of vertebrate genomes. Susumu Ohno predicted that only a small number of loci can be under selective pressure in any given population, due to limits in population size (Ohno, 1973). It is tempting to speculate that GRBs represent these loci under selective pressure: since GRBs contain developmental control genes plus numerous regulatory sequences, new evolutionary inventions could arise in these neighborhoods, which would constitute large targets for insertion of new sequences and regulatory mutations. Perhaps the large size of GRBs also allows faster selection since recombination between the target gene and any fortuitous regulatory mutation or new insertion in the GRB periphery should be more probable if the region under selection is large. This would also suggest that one should expect mutations somewhere along the entire length of GRBs that are disadvantageous and that might represent in the human genome variants predisposing to genetic disease.

Links

The UCSC genome browser [http://genome.ucsc.edu].
ZFIN: The Zebrafish Model Organism Database [http://zfin.org].
Ancora Genome Browser [http://ancora.genereg.net].
ZenBase [http://zenbase.genereg.net/spp].

Acknowledgments

We thank Drs. José Luis Gómez Skarmeta, Koichi Kawakami, Laure Bally-Cuif and Terje Johansen for reagents, and Caterina Sunde, Rosaebel Elmer and Øyvind Reinshol for expert animal care. We acknowledge Pär G Engstrom's input on database development. The research reported in this paper was funded by a core grant from the Sars Centre, by grants from the University of Bergen, and by a grant from the Institut du Cerveau et de la Moelle Epinière, Paris, France to TSB; and by the Functional Genomics Programme (FUGE) of the Research Council of Norway and a core grant of the Sars Centre to BL. Our large-scale enhancer detection screen was additionally funded by the National Programme in Functional Genomics in Norway (FUGE) in the Research Council of Norway and by the European Commission as part of the ZF-Models Integrated Project in the 6th Framework Programme (Contract No. LSHG-CT-2003-503496). TAH and KT were also supported by ZF-Models, and their work was carried out in the laboratory of Stephen W. Wilson.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2008.10.044.

References

Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A., Rubin, E.M., 2007. Deletion of ultraconserved elements yields viable mice. PLoS Biol. 5, e234.
Becker, T.S., Lenhard, B., 2007. The random versus fragile breakage models of chromosome evolution: a matter of resolution. Mol. Genet. Genomics 278, 487–491.
Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. Science 304, 1321–1325.
Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., Haussler, D., 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441, 87–90.
Bowl, M.R., Nesbit, M.A., Harding, B., Levy, E., Jefferson, A., Volpi, E., Rizzoti, K., Lovell-Badge, R., Schlessinger, D., Whyte, M.P., Thakker, R.V., 2005. An interstitial deletion–insertion involving chromosomes 2p25.3 and Xq27.1, near SOX3, causes X-linked recessive hypoparathyroidism. J. Clin. Invest. 115, 2822–2831.
Bylund, M., Andersson, E., Novitch, B.G., Muhr, J., 2003. Vertebrate neurogenesis is counteracted by Sox1–3 activity. Nat. Neurosci. 6, 1162–1168.
Canestro, C., Yokoi, H., Postlethwait, J.H., 2007. Evolutionary developmental biology and genomics. Nat. Rev. Genet. 8, 932–942.
Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., Forrest, A.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A., Hayashizaki, Y., 2006. Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. 38, 626–635.
Carroll, S.B., 2008. Evo–devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134, 25–36.
Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K., Lander, E.S., 2007. Distinguishing protein-coding and noncoding genes in the human genome. Proc. Natl. Acad. Sci. U. S. A. 104, 19428–19433.
Dame, C., Kirschner, K.M., Bartz, K.V., Wallach, T., Hussels, C.S., Scholz, H., 2006. Wilms tumor suppressor, Wt1, is a transcriptional activator of the erythropoietin gene. Blood 107, 4282–4290.
Dee, C.T., Hirst, C.S., Shih, Y.H., Tripathi, V.B., Patient, R.K., Scotting, P.J., 2008. Sox3 regulates both neural fate and differentiation in the zebrafish ectoderm. Dev. Biol. 320 (1), 289–301.
de la Calle-Mustienes, E., Feijoo, C.G., Manzanares, M., Tena, J.J., Rodriguez-Seguel, E., Letizia, A., Allende, M.L., Gomez-Skarmeta, J.L., 2005. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. Genome Res. 15, 1061–1072.
Dubchak, I., Ryaboy, D.V., 2006. VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. Methods Mol. Biol. 338, 69–89.
Ellingsen, S., Laplante, M.A., Konig, M., Kikuta, H., Furmanek, T., Hoivik, E.A., Becker, T.S., 2005. Large-scale enhancer detection in the zebrafish genome. Development 132, 3799–3811.
Engstrom, P.G., Ho Sui, S.J., Drivenes, O., Becker, T.S., Lenhard, B., 2007. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Res. 17, 1898–1908.
Engstrom, P.G., Fredman, D., Lenhard, B., 2008. Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. Genome Biol. 9, R34.
Ertzer, R., Muller, F., Hadzhiev, Y., Rathnam, S., Fischer, N., Rastegar, S., Strahle, U., 2007. Cooperation of sonic hedgehog enhancers in midline expression. Dev. Biol. 301, 578–589.
Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M., et al., 1995. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum. Mol. Genet. 4, 415–422.
Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., Urasaki, A., Kawakami, K., McCallion, A.S., 2006. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat. Protoc. 1, 1297–1305.
Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.
Francke, U., Holmes, L.B., Atkins, L., Riccardi, V.M., 1979. Aniridia–Wilms' tumor association: evidence for specific deletion of 11p13. Cytogenet. Cell Genet. 24, 185–192.
Glaser, T., Walton, D.S., Maas, R.L., 1992. Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. Nat. Genet. 2, 232–239.
Gomez-Skarmeta, J.L., Lenhard, B., Becker, T.S., 2006. New technologies, new findings, and new concepts in the study of vertebrate *cis*-regulatory sequences. Dev. Dyn. 235, 870–885.
Hadrys, T., Punnamoottil, B., Pieper, M., Kikuta, H., Pezeron, G., Becker, T.S., Prince, V., Baker, R., Rinkwitz, S., 2006. Conserved co-regulation and promoter sharing of hoxb3a and hoxb4a in zebrafish. Dev. Biol. 297, 26–43.
Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., Levine, M.S., 2008. Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. Proc. Natl. Acad. Sci. U. S. A. 105, 7762–7767.
Hill, R.E., Favor, J., Hogan, B.L., Ton, C.C., Saunders, G.F., Hanson, I.M., Prosser, J., Jordan, T., Hastie, N.D., van Heyningen, V., 1992. Mouse Small eye results from mutations in a paired-like homeobox-containing gene. Nature 355, 750.
Jordan, T., Hanson, I., Zaletayev, D., Hodgson, S., Prosser, J., Seawright, A., Hastie, N., van Heyningen, V., 1992. The human PAX6 gene is mutated in two patients with aniridia. Nat. Genet. 1, 328–332.
Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., Weber, R.J., Haussler, D., Kent, W.J., 2003. The UCSC Genome Browser Database. Nucleic Acids Res. 31, 51–54.
Kawakami, K., 2004. Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. Methods Cell Biol. 77, 201–222.
Kiefer, J.C., 2007. Back to basics: Sox genes. Dev. Dyn. 236, 2356–2366.
Kikuta, H., Fredman, D., Rinkwitz, S., Lenhard, B., Becker, T.S., 2007a. Retroviral enhancer detection insertions in zebrafish combined with comparative genomics reveal genomic regulatory blocks — a fundamental feature of vertebrate genomes. Genome Biol. 8 (Suppl. 1), S4.
Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A.Z., Engstrom, P.G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., Ghislain, J., Pezeron, G., Mourrain, P., Ellingsen, S., Oates, A.C., Thisse, C., Thisse, B., Foucher, I., Adolf, B., Geling, A., Lenhard, B., Becker, T.S., 2007b. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res. 17, 545–555.
Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., Ren, B., 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128, 1231–1245.
Kimura-Yoshida, C., Kitajima, K., Oda-Ishii, I., Tian, E., Suzuki, M., Yamamoto, M., Suzuki, T., Kobayashi, M., Aizawa, S., Matsuo, I., 2004. Characterization of the pufferfish

Otx2 *cis*-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. Development 131, 57–71.
Kleinjan, D.A., van Heyningen, V., 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am. J. Hum. Genet. 76, 8–32.
Kleinjan, D.A., Seawright, A., Schedl, A., Quinlan, R.A., Danes, S., van Heyningen, V., 2001. Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. Hum. Mol. Genet. 10, 2049–2059.
Kleinjan, D.A., Seawright, A., Mella, S., Carr, C.B., Tyas, D.A., Simpson, T.I., Mason, J.O., Price, D.J., van Heyningen, V., 2006. Long-range downstream enhancers are essential for Pax6 expression. Dev. Biol. 299, 563–581.
Kleinjan, D.A., Bancewicz, R.M., Gautier, P., Dahm, R., Schonthaler, H.B., Damante, G., Seawright, A., Hever, A.M., Yeyati, P.L., van Heyningen, V., Coutinho, P., 2008. Subfunctionalization of duplicated zebrafish pax6 genes by *cis*-regulatory divergence. PLoS Genet. 4, e29.
Kotani, T., Nagayoshi, S., Urasaki, A., Kawakami, K., 2006. Transposon-mediated gene trapping in zebrafish. Methods 39, 199–206.
Laplante, M., Kikuta, H., Konig, M., Becker, T.S., 2006. Enhancer detection in the zebrafish using pseudotyped murine retroviruses. Methods 39, 189–198.
Lauderdale, J.D., Wilensky, J.S., Oliver, E.R., Walton, D.S., Glaser, T., 2000. 3′ deletions cause aniridia by preventing PAX6 gene expression. Proc. Natl. Acad. Sci. U. S. A. 97, 13755–13759.
Laumonnier, F., Ronce, N., Hamel, B.C., Thomas, P., Lespinasse, J., Raynaud, M., Paringaux, C., Van Bokhoven, H., Kalscheuer, V., Fryns, J.P., Chelly, J., Moraine, C., Briault, S., 2002. Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. Am. J. Hum. Genet. 71, 1450–1455.
Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas III, E.J., Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.R., DeJong, P.J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.P., Parker, H.G., Pollinger, J.P., Searle, S.M., Sutter, N.B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., et al., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819.
Li, T., Lu, Z., Lu, L., 2006. Pax6 regulation in retinal cells by CCCTC binding factor. Investig. Ophthalmol. Vis. Sci. 47, 5218–5226.
Lowe, C.B., Bejerano, G., Haussler, D., 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc. Natl. Acad. Sci. U. S. A. 104, 8005–8010.
Ludwig, M.Z., Bergman, C., Patel, N.H., Kreitman, M., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403, 564–567.
Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., Kreitman, M., 2005. Functional evolution of a *cis*-regulatory module. PLoS Biol. 3, e93.
Meng, A., Tang, H., Ong, B.A., Farrell, M.J., Lin, S., 1997. Promoter analysis in living zebrafish embryos identifies a *cis*-acting motif required for neuronal expression of GATA-2. Proc. Natl. Acad. Sci. U. S. A. 94, 6267–6272.
Miles, C., Elgar, G., Coles, E., Kleinjan, D.J., van Heyningen, V., Hastie, N., 1998. Complete sequencing of the Fugu WAGR region from WT1 to PAX6: dramatic compaction and conservation of synteny with human chromosome 11p13. Proc. Natl. Acad. Sci. U. S. A. 95, 13068–13072.
Nobrega, M.A., Ovcharenko, I., Afzal, V., Rubin, E.M., 2003. Scanning human gene deserts for long-range enhancers. Science 302, 413.
Nornes, S., Clarkson, M., Mikkola, I., Pedersen, M., Bardsley, A., Martinez, J.P., Krauss, S., Johansen, T., 1998. Zebrafish contains two pax6 genes involved in eye development. Mech. Dev. 77, 185–196.
Ohno, S., 1973. Ancient linkage groups and frozen accidents. Nature 244, 259–262.
Ovcharenko, I., Loots, G.G., Nobrega, M.A., Hardison, R.C., Miller, W., Stubbs, L., 2005. Evolution and functional classification of vertebrate gene deserts. Genome Res. 15, 137–145.
Oxtoby, E., Jowett, T., 1993. Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. Nucleic Acids Res. 21, 1087–1095.
Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B.L., Couronne, O., Eisen, M.B., Visel, A., Rubin, E.M., 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499–502.
Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin, I.T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071.
Rastegar, S., Hess, I., Dickmeis, T., Nicod, J.C., Ertzer, R., Hadzhiev, Y., Thies, W.G., Scherer, G., Strahle, U., 2008. The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. Dev. Biol. 318, 366–377.
Sandelin, A., Bailey, P., Bruce, S., Engstrom, P.G., Klos, J.M., Wasserman, W.W., Ericson, J., Lenhard, B., 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics 5, 99.
Spitz, F., Gonzalez, F., Duboule, D., 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell 113, 405–417.
Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y., Kondoh, H., 2003. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Dev. Cell 4, 509–519.
Vavouri, T., McEwen, G.K., Woolfe, A., Gilks, W.R., Elgar, G., 2006. Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. Trends Genet. 22, 5–10.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al., 2001. The sequence of the human genome. Science 291, 1304–1351.
Visel, A., Minovitsky, S., Dubchak, I., Pennacchio, L.A., 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res. 35, D88–92.
Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., Pennacchio, L.A., 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat. Genet. 40 (2), 158–160.
Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., Haussler, D., 2007. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc. Natl. Acad. Sci. U. S. A. 104, 18613–18618.
Werner, T., Hammer, A., Wahlbuhl, M., Bosl, M.R., Wegner, M., 2007. Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis. Nucleic Acids Res. 35, 6526–6538.
Woolfe, A., Elgar, G., 2007. Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. Genome Biol. 8, R53.
Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J., Cooke, J.E., Elgar, G., 2005. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3, e7.
Xie, X., Kamal, M., Lander, E.S., 2006. A family of conserved noncoding elements derived from an ancient transposable element. Proc. Natl. Acad. Sci. U. S. A. 103, 11659–11664.
Zuniga, A., Michos, O., Spitz, F., Haramis, A.P., Panman, L., Galli, A., Vintersten, K., Klasen, C., Mansfield, W., Kuc, S., Duboule, D., Dono, R., Zeller, R., 2004. Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. Genes Dev. 18, 1553–1564.