Developmental Biology 387 (2014) 214–228

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Genomes and Developmental Control

A survey of ancient conserved non-coding elements in the *PAX6* locus reveals a landscape of interdigitated *cis*-regulatory archipelagos

Shipra Bhatia [a], Jack Monahan [a], Vydianathan Ravi [b], Philippe Gautier [a], Emma Murdoch [a], Sydney Brenner [b], Veronica van Heyningen [a], Byrappa Venkatesh [b], Dirk A. Kleinjan [a,*]

[a] MRC Human Genetics Unit at the MRC IGMM at the University of Edinburgh, Edinburgh EH4 2XU, United Kingdom
[b] Institute of Molecular and Cell Biology, Agency for Science Technology and Research (A*STAR), Biopolis, Singapore 138673, Singapore

ARTICLE INFO

Article history:
Received 2 September 2013
Received in revised form
29 December 2013
Accepted 11 January 2014
Available online 17 January 2014

Keywords:
Ancient conserved gnathostome enhancers
Reporter transgenics
PAX6
Elephant shark
Long-range enhancers
cis-Regulatory landscape
Regulatory archipelagos

ABSTRACT

Biological differences between cell types and developmental processes are characterised by differences in gene expression profiles. Gene-distal enhancers are key components of the regulatory networks that specify the tissue-specific expression patterns driving embryonic development and cell fate decisions, and variations in their sequences are a major contributor to genetic disease and disease susceptibility. Despite advances in the methods for discovery of putative *cis*-regulatory sequences, characterisation of their spatio-temporal enhancer activities in a mammalian model system remains a major bottle-neck. We employed a strategy that combines gnathostome sequence conservation with transgenic mouse and zebrafish reporter assays to survey the genomic locus of the developmental control gene *PAX6* for the presence of novel *cis*-regulatory elements. Sequence comparison between human and the cartilaginous elephant shark (*Callorhinchus milii*) revealed several ancient gnathostome conserved non-coding elements (agCNEs) dispersed widely throughout the *PAX6* locus, extending the range of the known *PAX6 cis*-regulatory landscape to contain the full upstream *PAX6-RCN1* intergenic region. Our data indicates that ancient conserved regulatory sequences can be tested effectively in transgenic zebrafish even when not conserved in zebrafish themselves. The strategy also allows efficient dissection of compound regulatory regions previously assessed in transgenic mice. Remarkable overlap in expression patterns driven by sets of agCNEs indicates that *PAX6* resides in a landscape of multiple tissue-specific regulatory archipelagos.

Introduction

Advances in modern genomics have clearly established that a comprehensive understanding of the causes of human genetic diseases requires greater knowledge about the functions not just of the coding, but also the non-coding parts of the genome. While mutations in the coding regions of a gene in most cases directly affect the functional integrity of the encoded protein, the effects of mutation or variation in the non-coding parts of a gene locus are often more subtle; affecting protein levels in a tissue-specific subset of the expression domain or disrupting the ability to respond to external signals. Developmental regulators are a class of genes, comprising transcription factors and signalling molecules with important roles during embryonic development and maintenance of adult homeostasis, that are particularly sensitive to non-coding variation. They orchestrate their often pleiotropic functions through participation in multiple gene regulatory networks whose connections are strictly regulated through the activity of long-range *cis*-regulatory elements (Davidson, 2006; Levine and Davidson, 2005).

The multi-functional developmental regulator PAX6, a paired and homeo-box containing transcription factor, has crucial roles in the development and maintenance of the central nervous system (CNS), the olfactory system and endocrine pancreas, although the gene is best known for its critical role in eye development (Ashery-Padan and Gruss, 2001; Osumi et al., 2008; Simpson and Price, 2002; van Heyningen and Williamson, 2002). In nearly every species known to use vision, development of the eyes is critically dependent on the presence or dosage of PAX6 (Pax6) (Gehring, 2005). In humans, heterozygous mutations in *PAX6* are the cause of the congenital eye malformation aniridia (MIM 106210), a pan-ocular disease characterised by a variable degree of iris hypoplasia, nystagmus, foveal hypoplasia, and ciliary body abnormalities. In the mouse, Pax6 haploinsufficiency is the cause of the *Smalleye* (*Sey*) phenotype (Hogan et al., 1986), of which several heterozygous loss-of-function alleles exist (e.g. $Pax6^{SeyEd}/+$ and $Pax6^{Sey1Neu}/+$) (Favor et al., 2001; Hill et al., 1991). Increased Pax6 expression also leads to reduced eye size, as shown in mice carrying a large YAC transgene

* Corresponding author. Fax: +44 131 467 8456.
E-mail address: Dirk-Jan.Kleinjan@igmm.ed.ac.uk (D.A. Kleinjan).

0012-1606/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ydbio.2014.01.007

for the human *PAX6* locus (Chanas et al., 2009; Manuel et al., 2008; Schedl et al., 1996). In addition to its expression in the retina, lens and cornea of the developing vertebrate eye, Pax6 is also expressed, at different developmental stages, in regions of the forebrain, hindbrain, cerebellum, the ventral neural tube, the olfactory system and pancreatic islet cells (Grindley et al., 1995; Kioussi et al., 1999; Nomura et al., 2007; St-Onge et al., 1997; Stoykova and Gruss, 1994; Walther and Gruss, 1991; Warren and Price, 1997). The homozygous *Smalleye* phenotype reflects this expression pattern: homozygous *Smalleye* animals die immediately after birth with no eyes, no nasal structures and severe brain and pancreatic abnormalities (Ashery-Padan et al., 2004; Hill et al., 1991; Hogan et al., 1986). A similar phenotype was described for a reported human case with functional loss of both copies of *PAX6* (Glaser et al., 1994).

The conspicuous eye phenotype resulting from PAX6 haploinsufficiency combined with the neonatal lethality of systemic homozygous loss of the gene have firmly focussed attention on PAX6 as a master control gene for the eye. However, several additional phenotypes, such as epilepsy (Strug et al., 2009), anosmia (Sisodiya et al., 2001), absence of pineal gland (Mitchell et al., 2003), myopia (Hammond et al., 2004), and diabetes (Ashery-Padan et al., 2004; Gosmain et al., 2012; Hart et al., 2013; Nishi et al., 2005; St-Onge et al., 1997) are also linked with disruption of Pax6 activity. Some of these additional phenotypes are found in aniridia patients, such as absence of anterior commissure and anosmia (Bamiou et al., 2004; Sisodiya et al., 2001), and an increased frequency of diabetes (Nishi et al., 2005; Yasuda et al., 2002). Other phenotypes may occur without presence of aniridia, if mutation or variation in regulatory control elements has led to tissue-specific disruption of PAX6 expression. Detailed dissection of the tissue- and stage- specific roles of Pax6 in development and tissue maintenance therefore remains of great interest, yet is hampered by its pleiotropy and the intricate connectivity of developmental processes and tissues.

The wide functional activity spectrum of PAX6 is achieved through strict control of spatio-temporal expression, requiring the activities of a large number of long-range *cis*-regulatory elements residing in a large and complex regulatory domain around the gene. The essential role of long range regulatory elements in control of PAX6 expression was brought to light by the discovery of a subset of aniridia patients carrying chromosomal abnormalities near, but not in *PAX6* that remove a large part of the gene regulatory domain (Fantes et al., 1995; Kleinjan et al., 2001; Lauderdale et al., 2000). All aniridia-associated breakpoints identified to date are located downstream of PAX6, within the large final intron of the adjacent *ELP4* gene, indicating the critical role of this region for expression of PAX6 in the eye. Several studies have investigated the transcriptional control of Pax6 expression (Griffin et al., 2002; Kammandel et al., 1999; Kim and Lauderdale, 2006; Kleinjan et al., 2004, 2006; McBride et al., 2011; Plaza et al., 1995; Williams et al., 1998; Xu et al., 1999), resulting in the identification of several *cis*-regulatory elements located immediately upstream of the promoters, within the introns of the gene, and in the extended domain downstream of the gene, within the *ELP4* introns (Griffin et al., 2002; Kleinjan et al., 2006, 2001; McBride et al., 2011). However, a large part of the *PAX6* gene locus, the 280 kb intergenic region upstream of the gene, remained unexplored for the presence of *cis*-regulatory sequences. Identification of these control elements has importance for disease diagnostics and to provide new insight into the regulatory networks in which the gene participates.

Among the variety of methods that can be used for the identification of *cis*-regulatory elements, multispecies sequence conservation is one of the simplest and most effective, but it requires subsequent functional testing in an appropriate model system to determine the tissue-specific activity of the identified element. The zebrafish is a cost-effective and relatively low-maintenance model system that is suitable for the generation of reporter transgenic animals. However, due to the additional whole genome duplication (WGD) in the teleost lineage and its consequently increased rate of evolutionary divergence, the zebrafish genome lacks many of the CNEs identifiable in other vertebrate species (Lee et al., 2011), raising some concern about the general usefulness of the zebrafish as a model of choice for testing CNEs.

We recently produced the complete genomic sequence of the *Pax6* (*Pax6.1*) locus from the elephant shark, *Callorhinchus milii* (Ravi et al., 2013). Sequence comparison with the human *Pax6* locus revealed the presence of several ancient conserved non-coding elements in the *PAX6* upstream region. Due to the putative connection of PAX6 mis-expression with susceptibility to a number of diseases we decided to investigate the regulatory potential of these sequences. We performed detailed sequence comparisons of the 280 kb genomic region between *PAX6* and the upstream gene, *RCN1*, in a spectrum of vertebrate species to allow categorisation of the elements into ancient CNEs or lineage specific CNEs. Additionally we selected two uncharacterised elements from the downstream region using the same specification. We tested the elements for enhancer activity using a combination of zebrafish and mouse transgenic studies, to identify and characterise novel distally located regulatory elements in the Pax6 locus. Our results extend the repertoire of known *PAX6* control elements and provide further insight into the detail and etiology of the complex regulatory landscape around this critical developmental control gene.

Materials and methods

Sequence analysis

Sequences for the human, mouse, opossum, chicken, coelacanth, *Xenopus* and zebrafish *Pax6* loci were extracted from the UCSC Browser (http://genome.ucsc.edu/) and the sequence of the elephant shark Pax6.1 was downloaded from GenBank (JX135563). Conserved sequences were identified with PIPmaker (Schwartz et al., 2003) and VISTA (Frazer et al., 2004). Alignments of ancient gnathostome CNEs (agCNEs) were made using ClustalOmega (Sievers et al., 2011). Comparative analysis and TFBS predictions were made using the Uniprobe (Newburger and Bulyk, 2009) and MEME suite (Bailey et al., 2009).

Cloning of ancient gnathostome conserved non-coding elements

agCNE elements were cloned for analysis in zebrafish transgenic reporter assays in a two-step process. A fragment containing the conserved element plus flanking sequence was PCR amplified from elephant shark BAC 23H6 or mouse genomic DNA using Phusion high-fidelity polymerase (NEB). attB4 and attB1r sequences were included in the PCR primers for use with the Gateway recombination cloning system (Invitrogen). For zebrafish transgenic studies the amplified fragment was first cloned into the Gateway pP4P1r entry vector and sequenced using M13 forward and reverse primers for verification. Next the agCNE containing pP4P1r construct was combined with a pDONR221 construct containing either a gata2 promoter-eGFP-polyA or a gata2 promoter-mCherry-polyA cassette, and recombined into a destination vector with a Gateway R4-R2 cassette flanked by Tol2 recombination sites. For mouse LacZ reporter studies the PCR amplified fragments with attB4 and attB1r sites were cloned directly into an hsp68-LacZ vector containing a P4-P1r entry cassette. An *Asc*I site was included in the forward primer to allow subsequent micro-injection fragment isolation. Primers used are shown in Supplementary Table 2.

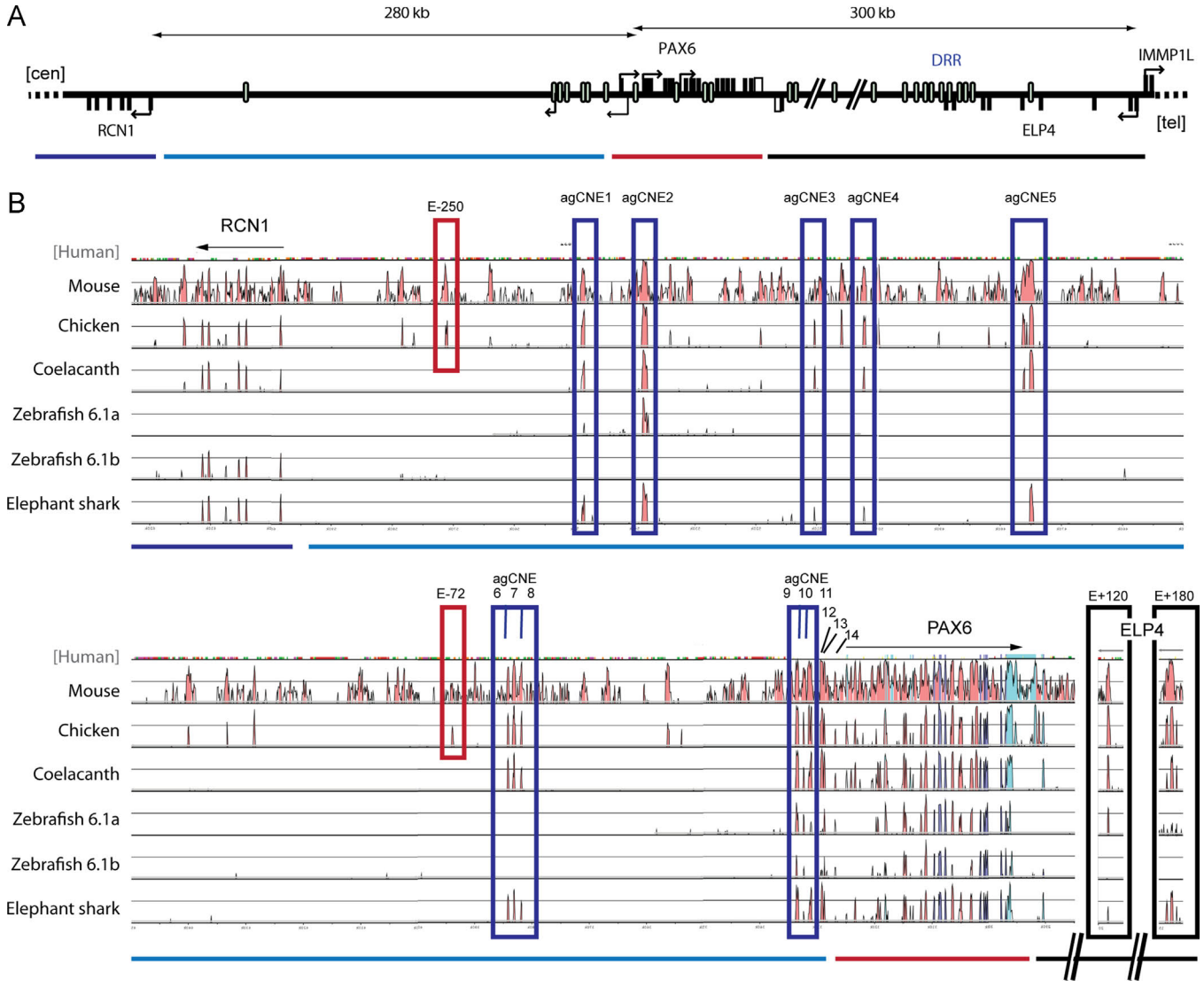

Fig. 1. Schematic overview of the *PAX6* locus. (A) The compact *PAX6* gene is flanked by adjacent *RCN1* and *ELP4* genes, with exons shown as black rectangles (*PAX6*, top strand; *ELP4* and *RCN1*, bottom strand). Arrows indicate transcription start sites (TSS). Many of the known long-range *cis*-elements (grey ellipses) for PAX6, including the Distal Regulatory Region (DRR) are located in the introns of *ELP4*. In contrast, the *cis*-regulatory potential of the gene desert between *RCN1* and *PAX6* was largely unexplored. The locus is sorted into regions indicated by coloured underlining: *WT1-RCN1* (dark blue), *RCN1-PAX6* (light blue), *PAX6* intragenic (red), *PAX6-ELP4* (black). (B) Visualisation of evolutionary sequence conservation by VISTA Plot of the *PAX6-RCN1* genomic region plus two additional previously uncharacterised conserved fragments from the *PAX6-ELP4* region. Genomic sequences from human, mouse, chicken, coelacanth, zebrafish and elephant shark were aligned. Peaks on the VISTA plot indicate fragments of significant sequence conservation. Fourteen distinct elements with conservation between human and elephant shark were identified and are scored as ancient gnathostome conserved non-coding elements (agCNEs).

Generation of mouse transgenic lines

All mouse experiments were approved by the University of Edinburgh ethical committee (TR-20-10) and performed under UK Home Office license number PPL 60/3785. Production, staining and imaging of transgenic mice was performed according to standard procedures as described previously (Kleinjan et al., 2001; McBride et al., 2011).

Generation of zebrafish transgenic lines

Zebrafish were maintained in a recirculating water system according to standard protocol (Westerfield, 2007). Embryos were obtained by breeding adult fish of standard strains (AB and WIK) and raised at 28.5 °C as described (Westerfield, 2007). Embryos were staged by hours post fertilization (hpf) as described (Kimmel et al., 1995).

Reporter plasmids were isolated using Qiagen miniprep columns and were given extra purification via a Qiagen PCR purification column (Qiagen), and diluted to 50 ng/μl with DNAse/RNAse free water. *tol2* transposase RNA was synthesized from a *NotI*-linearized *pCS2-TP* plasmid (Fisher et al., 2006b) using the SP6 mMessage mMachine kit (Ambion), and similarly diluted to 50 ng/μl. Equal volumes of the reporter construct (s) and the transposase RNA were mixed immediately prior to injections to give an injection solution containing 25 ng/μl of DNA and 25 ng/μl of transposase RNA. 1-2 nl of the solution was micro-injected per embryo into the cytoplasm of 200 embryos at the 1- to 2-cell stage. Embryos were screened for mosaic fluorescence at 1–5 days post-fertilization (dpf) and raised to adulthood. Germline transmission was identified by mating of sexually mature adults to wild-type fish and examining their progeny for fluorescence. Positive embryos were raised to adulthood and lines were maintained by outcrossing.

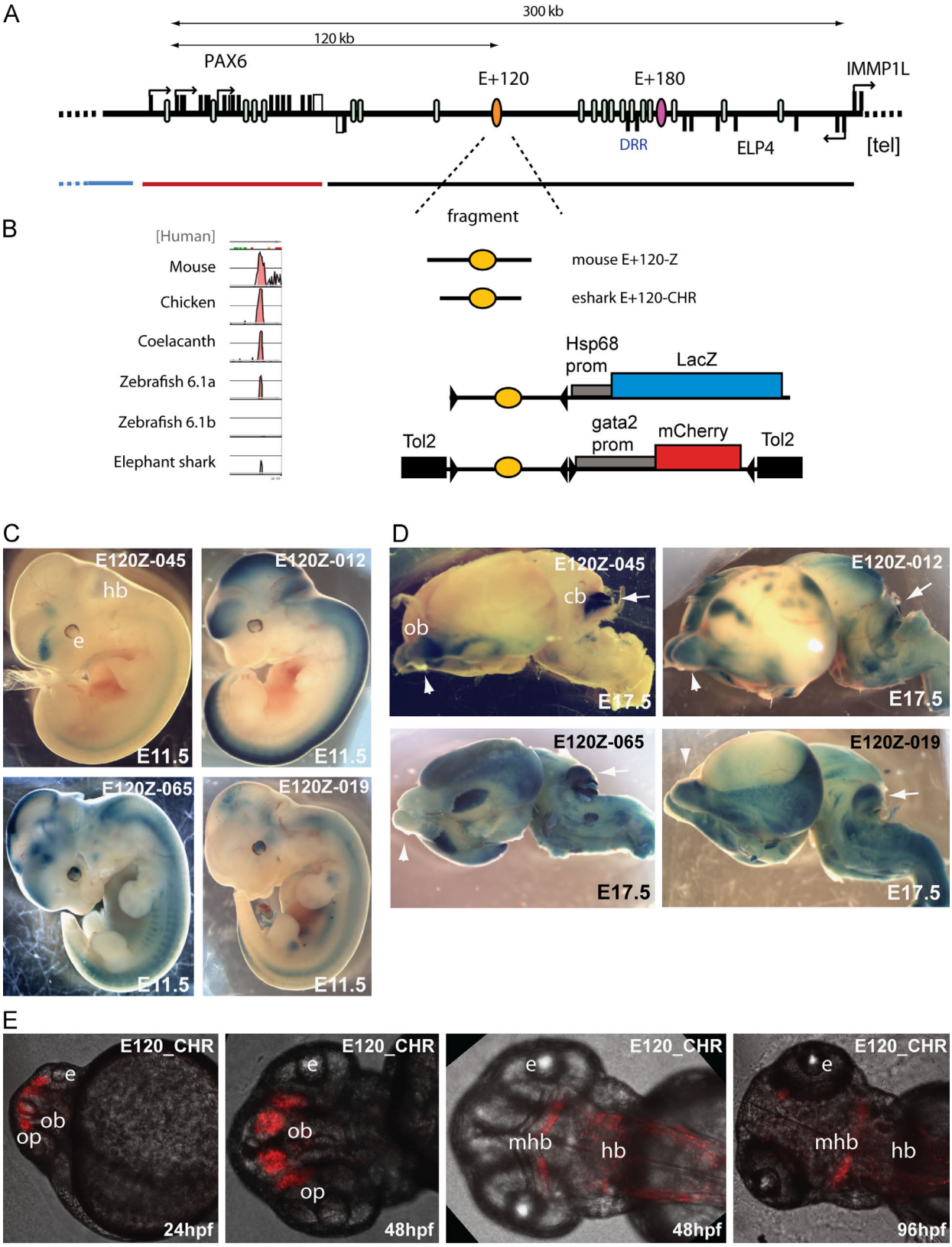

Fig. 2. Characterisation of a novel conserved element E+120 in mouse and zebrafish reporter transgenics. (A) The E+120 element (orange ellipse) is located 120 kb 3′ of the PAX6 P1 promoter in the human locus, in the large final intron of ELP4. (B) VISTA plot using the human sequence as a base, showing the deep conservation of E+120 in mouse, chicken, coelacanth, one of the zebrafish pax6 loci and the elephant shark. The constructs used for the reporter transgenic experiments are shown. (C) Four independent transgenic mouse lines were analysed for the E120Z construct, showing variable patterns of expression at E11.5. (D) At E17.5 strong expression is seen in the cerebella (black arrow) and olfactory bulbs (white arrowhead) in 3 out of 4 lines. (E) Expression pattern of the elephant shark E+120 element in transgenic zebrafish, shown by mCherry (CHR) fluorescence. Signal is seen in the olfactory placodes (op) and olfactory bulbs (ob) from 24 hpf. At 48 hpf (shown at two different confocal plains) and 96 hpf expression is also found at the midbrain-hindbrain boundary (mhb) and in the hindbrain (hb) region.

Imaging of zebrafish reporter transgenic embryos

Embryos for imaging were treated with 0.003% PTU (1-phenyl 2-thio-urea) from 24 hpf to prevent pigmentation. Embryos selected for imaging were anaesthetised with tricaine and mounted in 1% low-melting agarose. Images were taken on a Nikon A1R confocal microscope and processed using Nikon A1R analysis software.

RNA isolation and rtPCR

RNA was isolated from cell lines and mouse embryo tissues using TRI reagent (Sigma) according to manufacturer's protocol. RT-PCR was performed using Superscript II reverse transcriptase and random hexamers for first strand synthesis. PCR primers used were:

Pax6: 5′-TAGATGGGCGCAGACGGCATG-3′ and 5′-AGATCTATTTT GGCTGCTAGTC-3′; AK032637: 5′-CTCTCCGGTTCGAGTTATGC-3′ and 5′-CGCAGTTCTGGTCGTGTAGA-3′; Gapdh: 5′-CATGGCCTTCCG TGTTCCTA-3′ and 5′-CCTGCTTCACCACCTTCTTGAT-3′.

mRNA in situ hybridization

RNA *in situ* hybridization on fish embryos was performed as previously described (Thisse and Thisse, 2008).

Results

Ancient conserved non-coding elements (aCNEs) are evolutionarily conserved, non-repetitive intergenic or intronic sequence fragments present in a variety of contemporary species that are separated by a large time-span of divergence from a common ancestor. In the jawed vertebrate lineage, non-coding fragments that are conserved between mammals and the cartilaginous fish species the elephant shark, *Callorrhinchus millii* (Venkatesh et al., 2006), are a category of aCNEs that can be defined as ancient gnathostome CNEs (agCNEs). Cartilaginous fishes are among the most anciently diverged vertebrate lineages to split from the lineage leading to mammals. Conserved sequences in common between these species have survived at least 450 million years of vertebrate evolution (Venkatesh et al., 2005). We had previously obtained complete genomic sequence for the elephant shark *Pax6* locus (Ravi et al., 2013). Here we carry out comparative sequence analyses across the *PAX6* locus to look for agCNEs with a putative role in *PAX6* gene regulation, using genomic sequences from human, elephant shark and a selection of vertebrate species whose lineages have diverged at various more recent timepoints in evolutionary history.

We obtained genomic sequence for the *Pax6* locus from a range of organisms representing the major groups of vertebrate species by downloading their latest available genomic sequences. We selected the region ranging from the nearest upstream gene, *RCN1*, located centromeric to *PAX6*, to the promoter of the adjacent downstream gene, *ELP4*, located telomeric to *PAX6* in tail-to-tail orientation (Fig. 1A). To identify regions of ancient evolutionary sequence conservation we performed multispecies sequence alignments using PIPmaker (Schwartz et al., 2003) and VISTA (Frazer et al., 2004) using a window size of 50 bp and minimum conservation of 70%. Both programs were in close agreement and the VISTA result is shown in Fig. 1B.

In total we identified 41 human-elephant shark non-exonic sequence fragments that qualify as agCNEs in the *RCN1-ELP4* genomic locus. In the human genome this region spans 580 kb. Comparison between human and chicken for the same region revealed 74 CNEs, indicating that in addition to the large number of agCNEs, regulatory potential in the locus is actively evolving. To categorise the elements we divided the locus into regions: from *RCN1* to the 5′-end of the *PAX6* gene ending at the P0 promoter, the extended *PAX6* promoter and intragenic region up to its 3′–UTR, and the downstream region most of which covers the *ELP4* gene body from the 3′–UTR to the *ELP4* promoter. CNEs in the latter region thus reside in the *ELP4* introns, and the majority of these have been assayed previously (Griffin et al., 2002; Kleinjan et al., 2006, 2001; McBride et al., 2011). Nevertheless, two hitherto uncharacterised agCNEs from this downstream region are included in the current study: E+120 and E+180 (see Fig. 1).

The upstream region (Fig. 1, blue underlined section) harbours 14 agCNEs. The three elements located closest to *PAX6* (agCNEs12-14) correspond to previously characterised lens and pancreas enhancers (Carbe et al., 2012; Delporte et al., 2008; Kammandel et al., 1999; Williams et al., 1998; Zhang et al., 2003). The eleven more distal elements (agCNEs 1-11) were selected for characterisation in zebrafish reporter transgenics. In addition we selected two CNEs that are not found in elephant shark, coelacanth or zebrafish, but are conserved between chicken, mouse and human. At present it is unclear whether these elements have arisen after divergence of these species or have been lost in the former species following lineage divergence. These elements are named based on their approximate location in the human locus with respect to the *PAX6* P1 promoter as E-250 and E-72.

We first assessed the validity of the approach of using zebrafish reporter transgenics for analysis of human-elephant shark conserved sequences by creating both zebrafish and mouse transgenic lines for the E+120 element, which is also conserved in zebrafish (Fig. 1b). We obtained four expressing transgenic mouse lines out of eight lines in total. We could not recognise a consistent expression pattern at E11.5, but at E17.5strong expression was seen in the cerebellum in three of the lines, while some weaker expression was present in the fourth line (Fig. 2). Strong expression was also seen in the olfactory bulbs at E17.5 in three of the lines (Fig. 2). In zebrafish transgenics we examined expression at 1, 2 and 4 dpf. Reporter signal was seen in the olfactory bulbs and placodes in the forebrain and in the cerebellum and hindbrain.

Next we compared the expression driven by the E+180 region in both mouse and zebrafish transgenic lines. While this region, containing three closely spaced peaks of sequence conservation, is conspicuously present in elephant shark, coelacanth, chicken

Fig. 3. Characterisation of the E+180 conserved element. (A) Overview of the PAX6 downstream region indicating the position of the E+180 element at the telomeric end of the DRR. (B) The E+180 element contains three peaks of sequence conservation between human, mouse, chicken, coelacanth and elephant shark, but no significant conservation is seen in the zebrafish pax6 loci. A schematic overview of putative *cis*-elements in the telomeric half of the DRR shows the location of the E+180 elements and the fragments used in the transgenic reporter experiments. (C)–(H) Reporter transgenic mice with the E180 fragment from elephant shark (E180A/B/C). Strong expression is present at E13.5 in the trigeminal ganglion (TG) and in the dorsal root ganglia (DRG) along the neural tube (NT). At E13.5 expression also appears in the ganglion cell layer (GCL) of the eye (E) and in the optic nerves (ON). At E17.5 staining is present in the olfactory bulbs (OB) and olfactory epithelium (OE), in nerve tracts along the neural tube and hindbrain, and remains strong in the trigeminal ganglion and nerves. Expression is also present in the retinal ganglion cells, and staining extends along the optic nerve, chiasm (OC) and optic tracts (OT) to the lateral geniculate nuclei (LGN). (I)–(Q) Transgenic embryos with the mouse E180B only fragment. Reporter expression first becomes apparent from E11.5 in the trigeminal ganglion and increases in strength towards E13.5. Expression in the ganglion cell layer of the retina and in the optic nerve starts around E13.5 and continues to manifest at E17.5 in the retinal ganglion cells, optic nerve, chiasm and optic tracts and lateral geniculate nuclei. Staining is also seen in the olfactory bulbs and epithelium. (R)–(U) The mouse and elephant shark E180 fragments were also tested in reporter transgenic zebrafish. *In situ* hybridisation with a GFP antisense probe on mE180B-GFP transgenic fish shows the expression in the trigeminal nerve and tracts along the neural tube. Dual colour fluorescence in mouse E180B-GFP and elephant shark E180-mCherry double transgenic fish show the near-complete overlap in expression.

and mammals, it is not well conserved in either the zebrafish *pax6.1a* or *pax6.1b* loci, suggesting the element has been lost in the zebrafish lineage (Fig. 3). Alternatively the sequence of the element may have diverged, while functional activity has been retained, as shown for some other elements (Fisher et al., 2006a; McGaughey et al., 2008). We first used the elephant shark sequence containing all three subregions of conservation in transgenic mice. The construct containing the E180 region

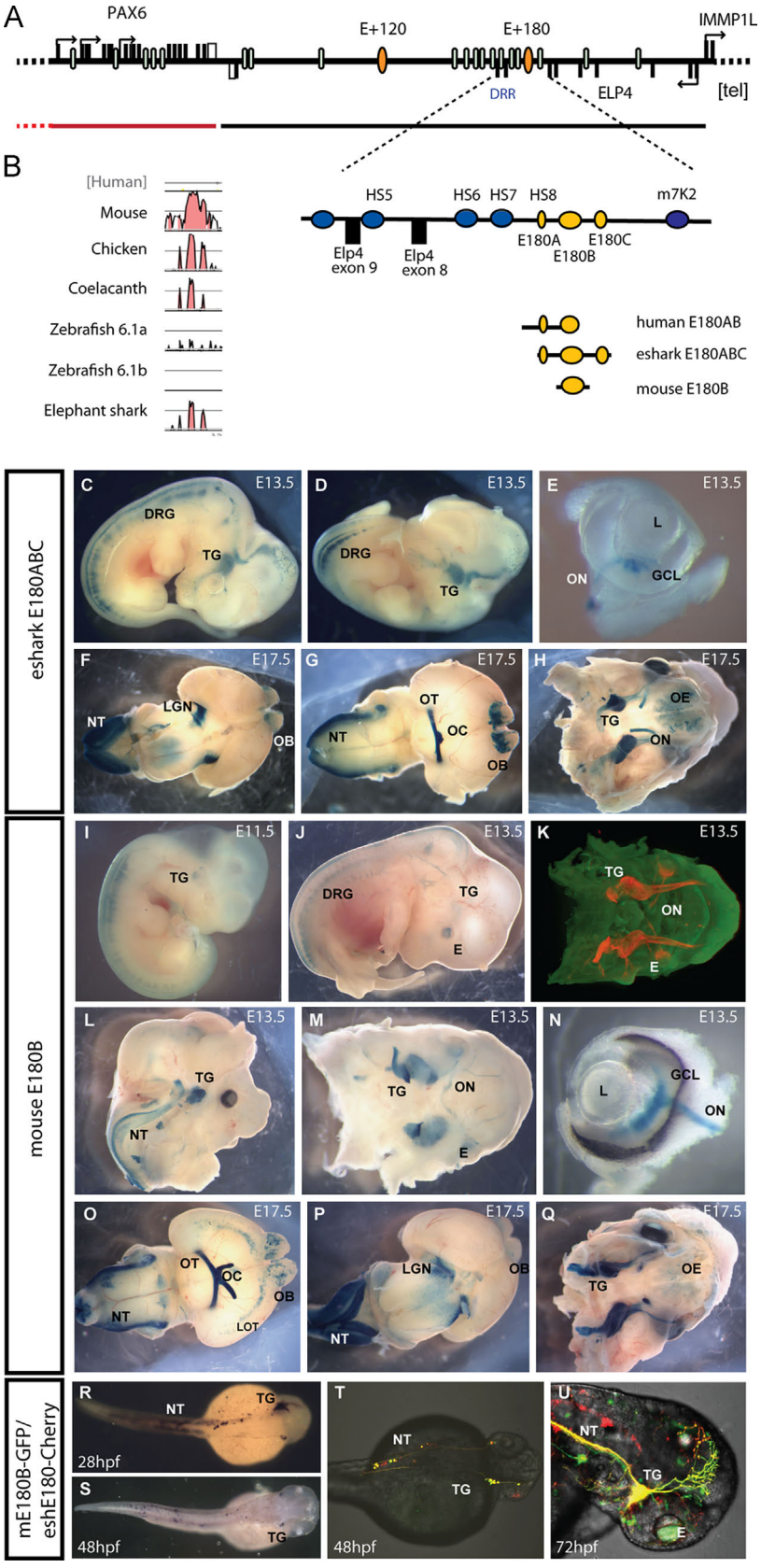

Table 1
Ancient gnathostome non-coding elements conserved between human and elephant shark, tested for expression patterns in stable zebrafish reporter transgenic fish at 2 and 4 dpf.

agCNE name	Element name	Human genomic position (hg19)	Expression at 2 dpf	Expression at 4 dpf	Reference
	E-250	chr11:32085525-32086304	No specific pattern	No specific pattern	This study
agCNE1		chr11:32062882-32063230	Trigeminal ganglia Dorsal spinal cord neurons	Trigeminal ganglia Dorsal spinal cord neurons	This study
agCNE2	E-200	chr11:32052624-32053243	Olfactory bulbs, Cerebellum, Olfactory bulbs, Lateral olfactory tracts	Olfactory bulbs, Cerebellum, Olfactory bulbs, Lateral olfactory tracts	Ravi et al. (2013)
agCNE3		chr11:32024810-32024977	Olfactory bulbs, olf placodes	Olfactory bulbs, olf placodes	This study
agCNE4		chr11:32016455-32017112	Hindbrain	Hindbrain	This study
agCNE5	Id855	chr11:31989455-31989799	Forebrain	Forebrain	Visel et al. (2007) This study
	E-72	chr11:31908464-31908883	No specific pattern	No specific pattern	This study
agCNE6	E-55/A	chr11:31898728-31898965	Hindbrain, Neural tube	Hindbrain, Neural tube	This study
agCNE7	E-55/B	chr11:31897563-31898031	No specific pattern	No specific pattern	This study
agCNE8	E-55/C	chr11:31896241-31896769	No specific pattern	No specific pattern	This study
agCNE9	Up-10	chr11:31,847,850-31,848,450	Pineal gland	Pineal gland	This study
agCNE10	Up-9	chr11:31,846,835-31,847,159	No specific pattern	No specific pattern	This study
agCNE11	Up-8	chr11:31,845,331-31,846,404	Pineal gland	Pineal gland	This study
agCNE12	P	chr11:31,843,629-31,844,067	Pancreas	Pancreas	Delporte et al., 2008; Kammandel et al. (1999), Williams et al. (1998), Zhang et al. (2003)
agCNE13	EE	chr11:31,843,215-31,843,530	Lens	Lens	
agCNE14	P2	chr11:31,841,310-31,841,722	Pancreas	Pancreas	
	P0 promoter Exon 0	chr11:31,839,356-31,839,510			
E+120	E+120	chr11:31712679-31713314	Forebrain, olfactory bulb, Cerebellum, hindbrain	Forebrain, olfactory bulb, Cerebellum, hindbrain	This study
E+180B	E+180B	chr11:31661947-31663424	Trigeminal ganglia Dorsal spinal cord neurons	Trigeminal ganglia Dorsal spinal cord neurons	This study

produced a consistent pattern (three expressing/five total lines), with expression observed in the trigeminal ganglion from E11.5 onwards, as well as in the retinal ganglion cells of the eye from E13.5, and in the optic nerves projecting through the optic chiasm and optic tracts to the lateral geniculate nuclei (LGN) at E17.5 (Fig. 3). Expression was also seen in the olfactory system, dorsal root ganglia and some nerve tracts running along the spinal cord. To distinguish between the three peaks of sequence conservation in the E180 region we named them E180A, E180B and E180C. The E180B element is the most highly conserved, while the moderately conserved E180A peak coincides with a previously identified constitutive HS site (HS8) (Kleinjan et al., 2001; McBride et al., 2011). We made transgenic mice with a human fragment centred on the HS8/E180A fragment but also containing the E180B sequence (three expressing/seven total transgenic lines), and observed staining in the same pattern as seen in the elephant shark E180 reporter mice (not shown). Finally we made transgenic mice with the mouse E180B fragment alone (two expressing/four total). The expression pattern driven by the mouse E180B fragment was indistinguishable from the pattern generated by the elephant shark E180A/B/C or human E180A/B elements, suggesting the E180B element alone is sufficient for the tissue-specific expression pattern driven by this region. Although we have not tested the E180A and E180C peaks separately in reporter mice, they appear unnecessary for the tissue-specific expression from the E180 region and may instead carry some other function. These results indicate that both sequence and function of the E180B element have been strictly conserved over at least 450 million years.

As the E180B element is conserved between elephant shark and human we concluded that it must have been present in the ancestral *Pax6.1* locus, and subsequently lost in the teleost loci. This view is supported by the presence of a rudimentary fragment of conservation in the medaka *pax6.1* locus (not shown). To test whether, despite losing the conserved E180 enhancer sequence itself, zebrafish had retained the molecular machinery to drive expression from the element, we made reporter transgenic zebrafish with both the elephant shark E180 and mouse E180B fragments. A highly specific expression pattern was seen in the trigeminal nerves and spinal nerve tracts of transgenic fish with both the elephant shark and mouse E180B fragments (Fig. 3). However, expression was limited to the trigeminal and spinal nerves and not seen in the olfactory system, the retinal ganglion cells, optic nerves, tracts or LGN of transgenic fish, suggesting the element's capacity to drive expression in those tissues has been lost in the fish. Thus some species-specific difference between mouse and fish exist in the interpretation of regulatory function for this enhancer.

Having established the validity of zebrafish reporter transgenics for the characterisation of ancient CNEs, we next focussed on the *PAX6* centromeric region, between the upstream *RCN1* gene and the *PAX6* promoter region (Fig. 1, blue underline). We surveyed the regulatory potential of the eleven agCNEs in this region by generating zebrafish reporter lines with the elephant

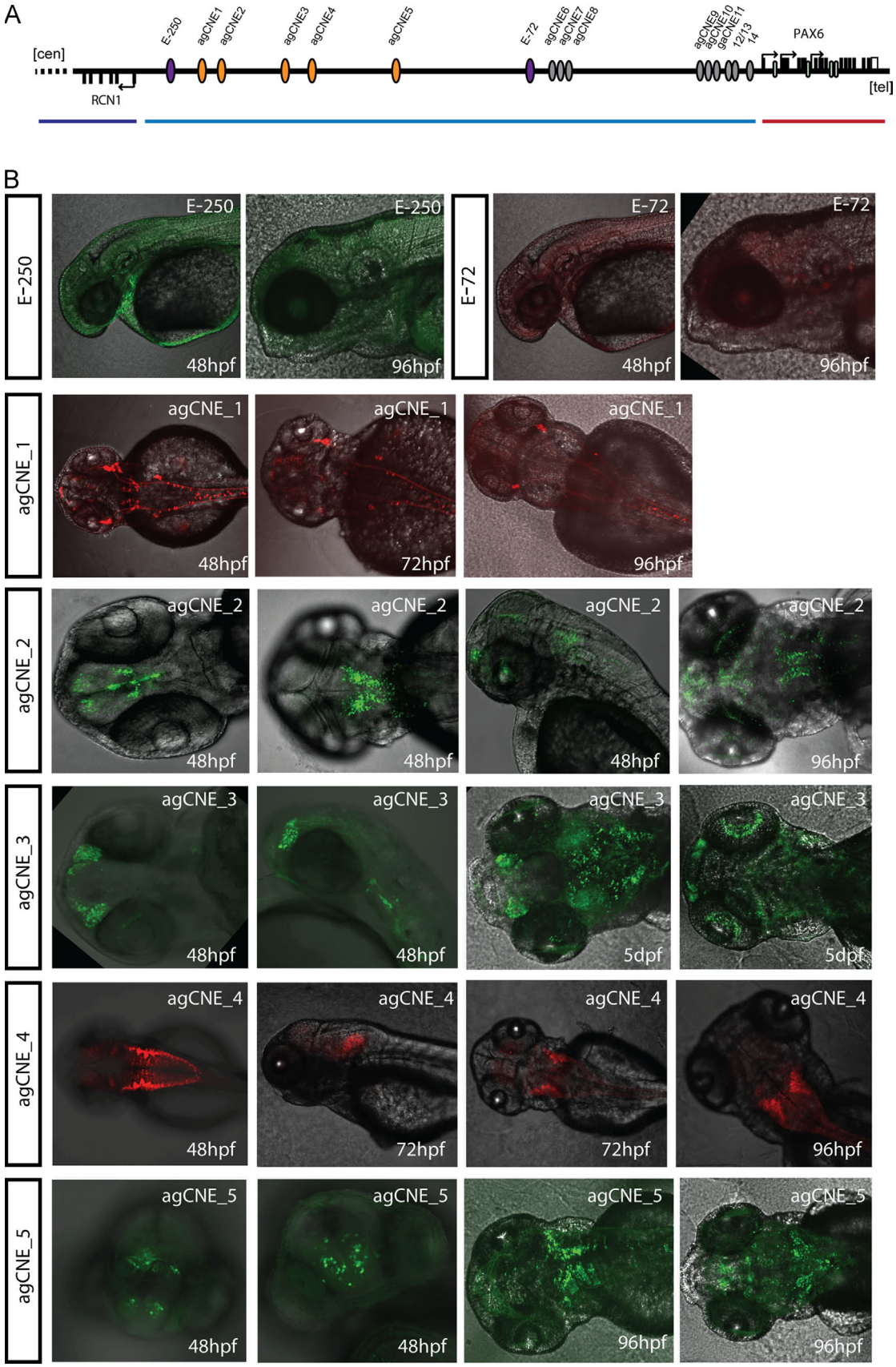

Fig. 4. Characterisation of tissue-specific expression patterns of CNEs from the *RCN1* to *PAX6* intergenic region. (A) Schematic overview of the region with the positions of the ancient gnathostome CNEs indicated by ellipses. The agCNEs (orange) and the chick-mammalian specific CNEs (purple) characterised in this figure are highlighted. (B) Representative reporter transgenic embryos are shown for agCNE1-5 and for the less deeply conserved E-250 and E-72 elements. Only background signal is seen with the E-250 and E-72 fragments. agCNE1 drives expression in the trigeminal ganglia and dorsal spinal cord neurons. agCNE2 transgenic embryos show signal in forebrain, olfactory bulbs and placodes and hindbrain. Expression from agCNE3 is seen in the olfactory placodes and olfactory bulbs with weak expression in hindbrain in some of the lines. In contrast agCNE4 shows strong hindbrain specific expression. agCNE5 shows consistent expression in the forebrain, with hindbrain expression in some lines.

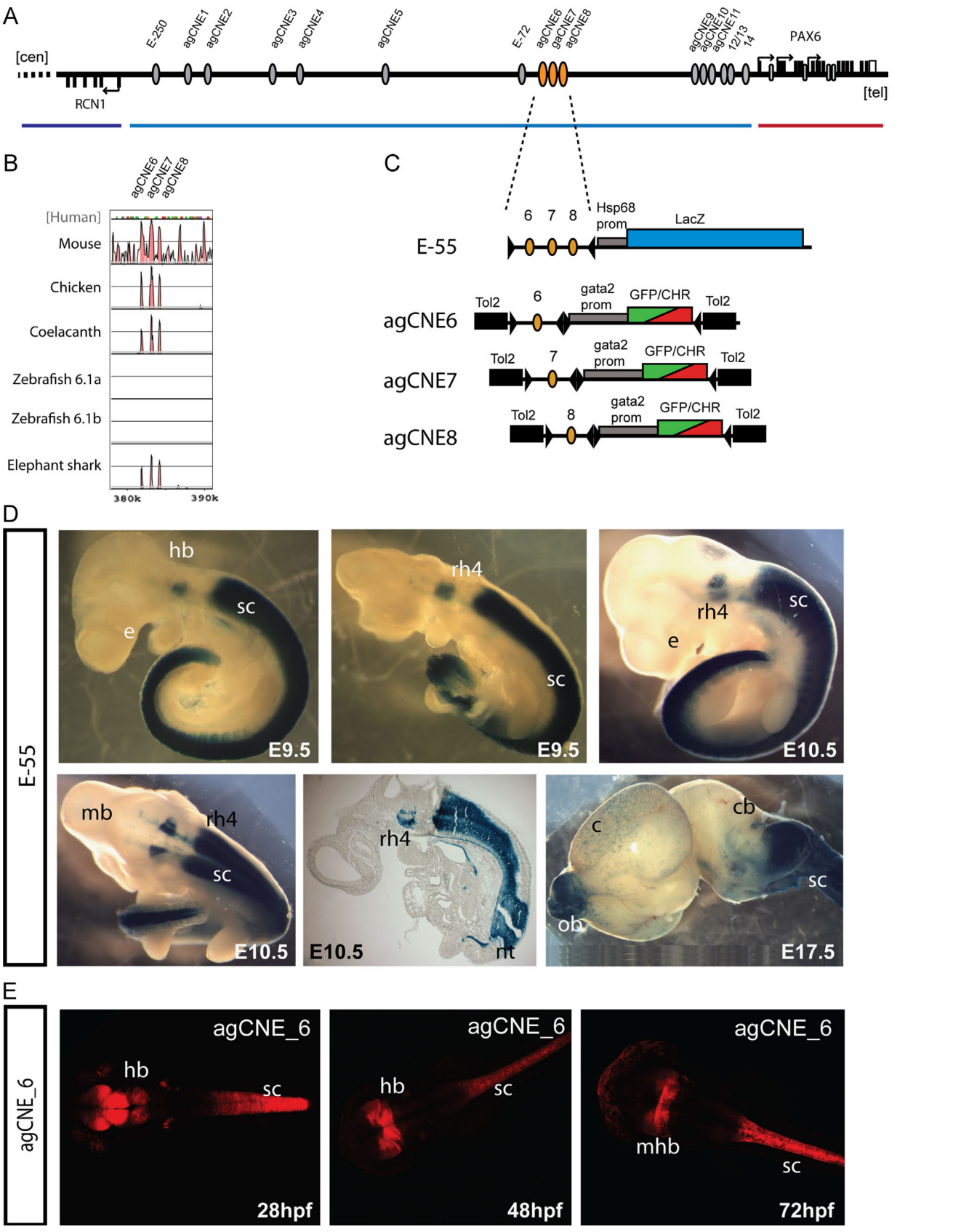

Fig. 5. Functional dissection of the E-55 region. (A) Schematic overview of the *RCN1* to *PAX6* intergenic region with the positions of agCNEs indicated by grey ellipses and agCNE6-8 in orange. (B) VISTA plot of the multispecies sequence alignment of the fragment showing the three peaks of conservation in elephant shark but not in either of the zebrafish *pax6* loci. (C) Schematic depiction of the reporter constructs used to generate transgenic mice and fish. (D) Transgenic embryos for the E-55 fragment, containing all three agCNEs, show X-gal staining in rhombomere 4 (rh4) of the hindbrain (hb) and from rhombomere 7 towards caudal along the length of the spinal cord, covering its full area on cross-section (sc). Expression in the neural tube is maintained at E17.5 and also seen in the olfactory bulbs (ob) in some embryos. (E) Transgenic reporter embryos carrying agCNE6 only show restricted expression in part of the hindbrain (hb), particularly at the midbrain-hindbrain boundary (mhb) and along the length of the spinal cord (sc).

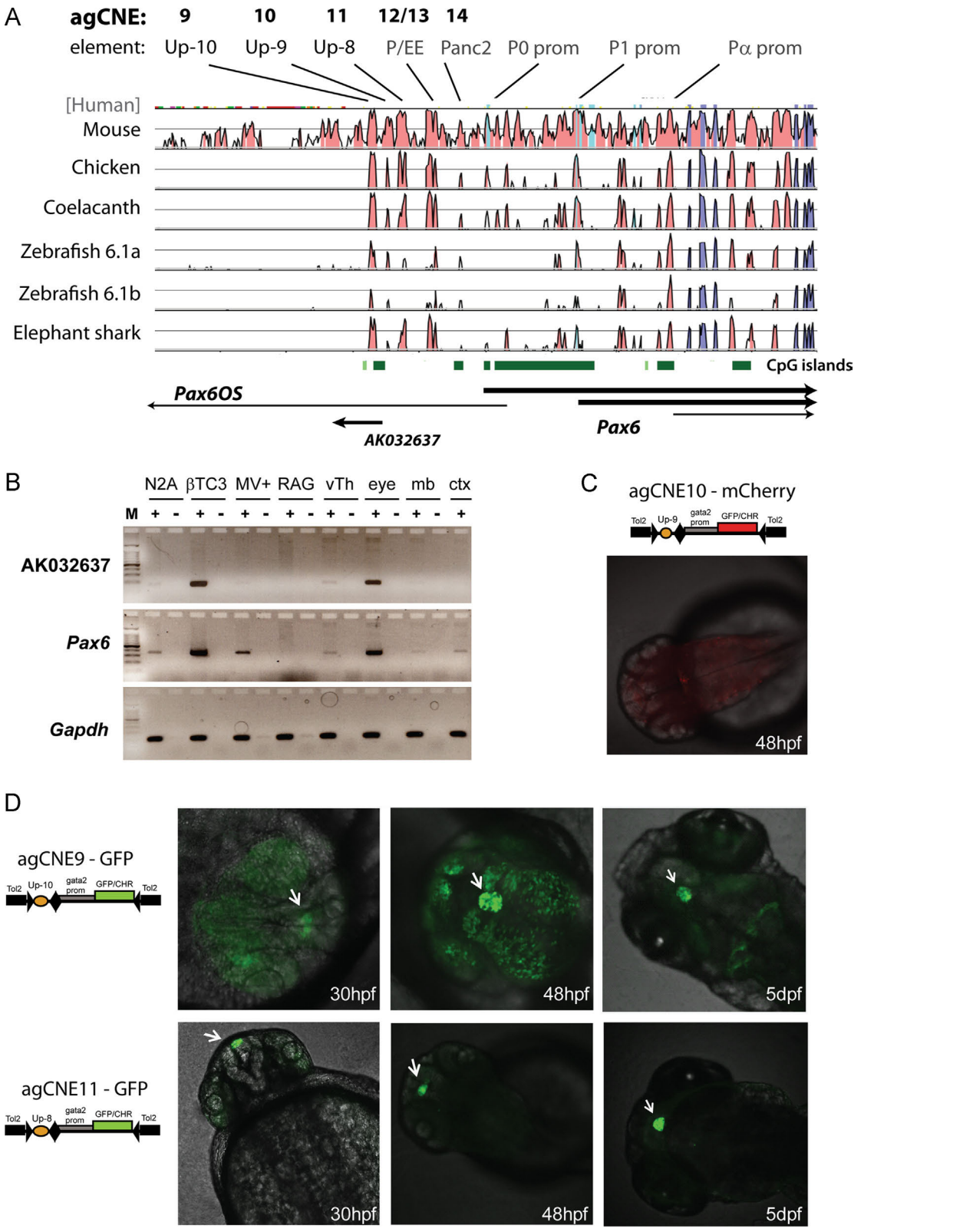

Fig. 6. Characterisation of anciently conserved elements in the *Pax6* promoter and proximal region. (A) VISTA plot of multispecies sequence conservation in the *PAX6* genomic region containing the promoters and immediate upstream CNEs. *PAX6* exons are shown as blue peaks and non-coding peaks of conservation are in pink. The positions of the validated Pax6 promoters P0, P1 and Pα are indicated, as are the known enhancers P/EE and Panc2 corresponding to agCNEs 12-14. agCNEs 9-11 are well conserved in all vertebrate *Pax6* loci analysed with the exception of agCNE11 which is absent in the zebrafish *pax6b* locus. Tracks underneath the plot show the presence of several CpG islands in the region, including one at the agCNE9-10 location, and a simplified overview of transcripts produced from the region. In addition to the various Pax6 transcripts two transcripts are found on the opposite strand, the multi-exon Pax6OS and the 2.8 kb single exon ncRNA AK032637. (B) rtPCR analysis of AK032637 in comparison with *Pax6* and *Gapdh* in a selection of cell lines and embryo tissues. N2A, neuroblastoma, bTC3, pancreatic β-cell line, MV+, lens epithelium, RAG, renal adenocarcinoma, vTh, ventral thalamus, eye, E17.5dpc whole eye, mb, midbrain, ctx, cortex. (C) Reporter transgenic zebrafish with construct agCNE10-mCherry show ubiquitous low level signal but lack specific expression. (D) Transgenic fish for agCNE9-GFP show expression in the pineal gland with additional signal in the fore- and hindbrain in some lines. Fluorescent signal in transgenic fish with the agCNE11-GFP construct is restricted to the pineal gland only.

shark or mouse sequence elements. The position of the agCNEs in the human genome (hg19) is indicated in Table 1. Expression patterns consistent between a minimum of four independent lines are shown in Fig. 4 and summarised in Table 1, with further details in Supplementary Table 2.

The most distal agCNE in the *RCN1-PAX6* region, agCNE1, is located 230 kb upstream of the *PAX6* P1 promoter in the human locus. The element is conserved in zebrafish and shows a specific expression pattern that is limited to the trigeminal ganglion, nerves and dorsal spinal cord neurons (Fig. 4).

agCNE2 is a highly conserved element with significant level of conservation over a 900 bp stretch between elephant shark and human. The element has previously been reported as E-200 (Ravi et al., 2013). Its sequence has also been maintained in the zebrafish *pax6.1a* locus. Expression is seen in the olfactory system and the hindbrain at 2 dpf and 4 dpf. In contrast, conservation of agCNE3 between elephant shark and human is found in a narrow fragment, and the element is absent in the zebrafish genome. Nevertheless, the agCNE3 element drives consistent expression in the olfactory bulbs and placodes (Fig. 4). Similarly, the agCNE4 element acts as a tissue-specific enhancer for expression in the hindbrain despite its lack of conservation in the zebrafish genome.

The next ancient conserved element, agCNE5, acts as a forebrain enhancer (Fig. 4). It is conserved in a wide variety of vertebrate species, including the zebrafish *pax6.1a* locus. The element coincides with the human element Id855, which has been shown to drive weak forebrain expression in transient transgenic mice at E11.5 (Visel et al., 2007).

The next three agCNEs in the *Rcn1-Pax6* upstream region, agCNEs6, 7, and 8, are located within close proximity of each other. This made it convenient to clone all three elements in a single fragment from the mouse genome, which was named E-55. We made transgenic reporter mice (four expressing/seven total lines) with the triple CNE-containing fragment. Strong expression was found along the spinal cord from rhombomere 8 towards the caudal end of the embryo. In addition a strictly contained site of expression was present in rhombomere 4 of the hindbrain (Fig. 5). Unlike the endogenous Pax6 pattern in the spinal cord (Fig. S5), the pattern driven by the E-55 element covered its full width and height. This wider expression domain is likely due to its assessment as an isolated element integrated at a random site in the genome, whereas in the endogenous locus a potential variety of restricting factors, such as repressors, epigenetic marking or locus conformation, would impose a more restricted functional output sensed by the Pax6 promoters. Since further dissection of the E-55 fragment by the mouse transgenic approach would be expensive and labour intensive, we made use of the relative ease of the zebrafish agCNE reporter assay to dissect the enhancer activity of the three individual elements contained with the E-55 fragment. We generated zebrafish reporters with the individual elephant shark elements agCNE6, 7 and 8. Surprisingly the full E-55 pattern was recapitulated in reporter fish with agCNE6, including both the

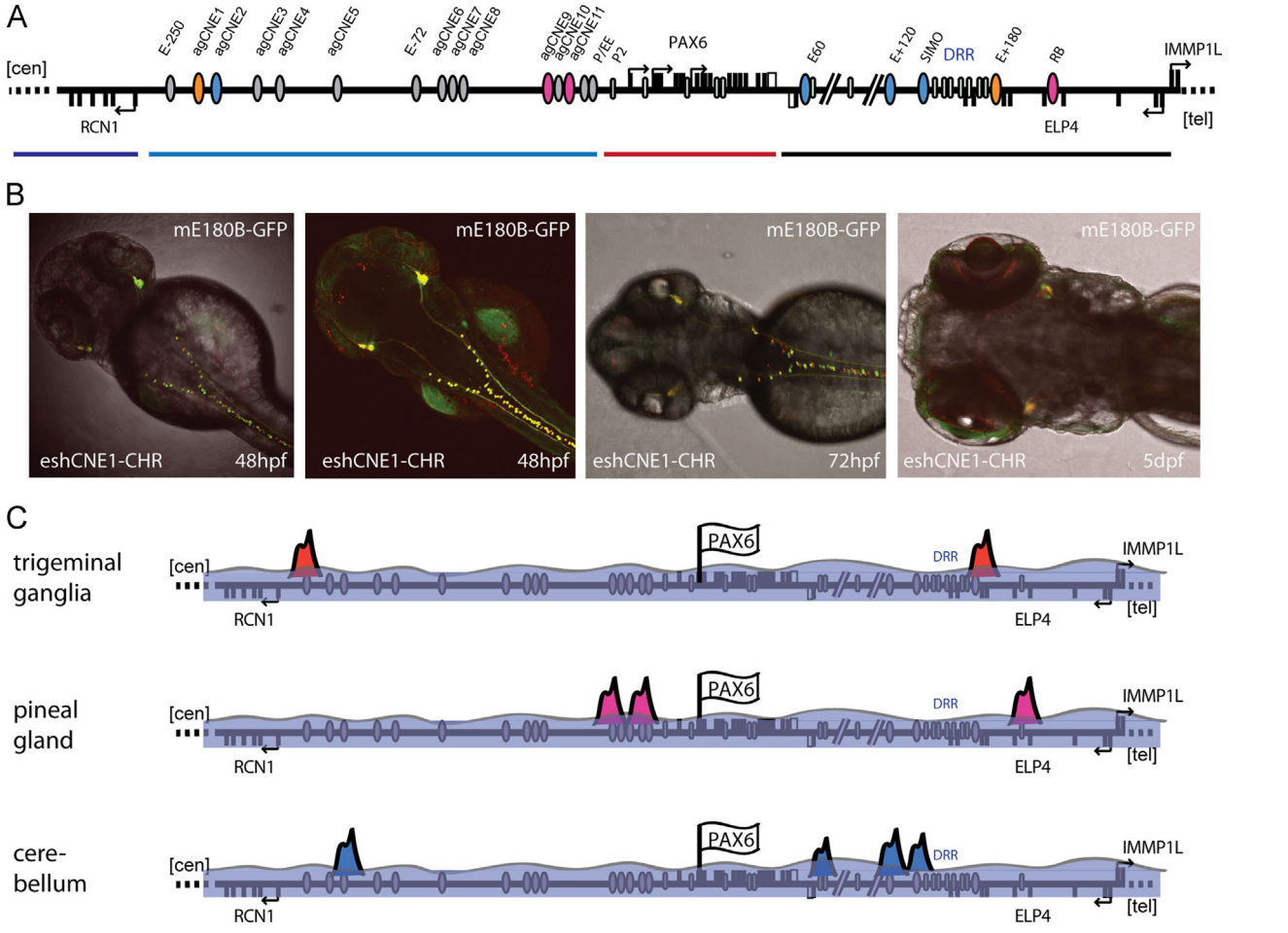

Fig. 7. Multiple CNEs with overlapping expression patterns form enhancer archipelagos at the *PAX6* locus. (A) Overview of the locus with newly characterised agCNEs depicted as ellipses. CNEs with overlapping regulatory activities found in this study are indicated with matching colours. (B) Overlap in the expression patterns of the agCNE1 and E180B enhancers is shown by dual-fluorescence in double transgenic zebrafish at 24 hpf, 48 hpf and 72 dpf. (C) Schematic depiction of the inter-digitated co-existence of multiple regulatory archipelagos at the *PAX6* genomic locus. Selected archipelagos are activated in a tissue-specific manner by the binding of TFs and cofactors at defined CNEs.

neural tube expression and a restricted expression in the hindbrain, albeit that the exact position was shifted anteriorly from rhombomere 4 in mouse to the mid-hindbrain boundary (MHB) in fish indicating that some species specific differences occur (Fig. 5). We did not observe any consistent pattern of expression in the reporter fish with agCNE7 or agCNE8. Together these results suggest that while aCNE6 acts as a tissue-specific enhancer, aCNE7 and 8 may fulfil other functions in the locus.

A further set of three closely positioned agCNEs is found approximately 10 kb upstream of the *PAX6* P0 promoter. All three elements are clearly recognisable in a multispecies conservation plot (Fig. 5A) and were originally named Up-10, Up-9 and Up-8 (corresponding to agCNE9, 10 and 11). agCNE10 (Up-9) is the least conserved between elephant shark and human, while agCNE11 shows variable conservation between the zebrafish *pax6.1a* and *pax6.1b* loci. Examination of this region of the murine *Pax6* locus in the UCSC genome browser suggested the presence of a non-coding RNA transcribed from the region encompassing agCNE9 in the orientation opposite to the *Pax6* transcripts. This transcript is represented by mouse EST AK032637, and starts between the agCNE9 and 10 elements (Fig. 6A) in a region marked by a high CpG content (CpG island). Limited expression profiling of AK032637 by rtPCR on a small panel of cell lines and embryonic brain tissues suggests it has an expression pattern that is similar to the Pax6 pattern (Fig. 6B). Expression is absent in RAG (renal adenocarcinoma) cells. A low level of expression is found in the N2A (neuroblastoma) and MV+ (lens epithelium) cell lines, E11.5 embryo head (not shown) and E17.5 embryonic hypothalamus and midbrain regions, while stronger expression is seen in βTC3 (pancreatic β) cells and E17.5 embryonic eyes (Fig. 6B). AK032637 fully encompasses the agCNE9 element, posing the question whether sequence conservation is due to selection on a potential enhancer function of the underlying DNA or on the functional activity contained of the ncRNA. We generated zebrafish reporter transgenics with each of the agCNE9, 10 and 11 elements. Remarkably both agCNE9 and agCNE11 produced strong and specific expression in the pineal gland, which is present from 30 hpf and persists until at least 5 dpf. While agCNE9 also showed some additional fluorescence in the fore- and hindbrain regions, expression from agCNE11 was highly restricted to the pineal gland. No consistent expression was detected for agCNE10, which would fit with this element having a potential minimal promoter function for the AK032637 ncRNA.

Next, we were interested to test whether our approach for the evaluation of putative enhancer elements would also work for tetrapod conserved sequences that are not anciently conserved, i.e. elements without conspicuous sequence conservation in alignments with either zebrafish, coelacanth or elephant shark. We selected two elements, E-250 and E-72 (see Fig. 1) and tested the mouse sequences in the zebrafish reporter transgenic assay. Though variable expression sites were seen in some transgenic founders, neither of the two elements showed any consistent pattern of expression (Fig. 4), indicating that neither of the E-250 or E-72 elements is recognised as an enhancer by the zebrafish transcriptional machinery. To test if the elements would function as enhancers in the mouse we generated transient transgenic embryos. We examined E-72 transgenic embryos at E13.5 but did not observe any obvious expression at this stage. In contrast, the E-250 element drove strong expression of the LacZ gene in the diencephalon, nasal region and midbrain of E11.5 and E13.5 embryos (Fig. S6). Staining was most obvious in the midbrain even though this is not a Pax6 expression site, but both the nasal and diencephalon are parts of the Pax6 expression domain (Fig. S5).

Finally, drawn by the similarity of expression patterns of several of the agCNEs we wanted to investigate whether this functional overlap would be reflected at the sequence level. First we confirmed the complete overlap in expression of the novel trigeminal ganglion elements agCNE1 and E180B in dual colour reporter transgenic embryos carrying both elements linked to GFP and mCherry respectively (Fig. 7B). Next we compiled two sets of conserved enhancer fragment sequences, one for the agCNE1 and E180B enhancers, and another for the pineal gland enhancers agCNE9, agCNE11 and the previously described RB (Kleinjan et al., 2006; Navratilova et al., 2009). In contrast to the strong conservation observed for each individual element across multiple species we found no significant sequence similarity between the separate co-expressing enhancers using BLAST or ClustalOmega (Altschul et al., 1990; Sievers et al., 2011), suggesting the elements are not ancient duplicates but have evolved independently. Using the MEME suite (Bailey et al., 2009) to look for shared sequence motifs yielded no convincing evidence for common motifs containing binding sites for shared transcription factors (data not shown), suggesting that the elements can achieve similar expression patterns through largely different sets of interacting binding factors.

Discussion

PAX6 is a transcription factor with pleiotropic functions during development and adult maintenance. Eye development is particularly sensitive to correct levels of PAX6, yet several other organs are also critically dependent on expression of PAX6, including the brain, neural tube, pancreas and olfactory system. It is well known that the spatio-temporal expression patterns of many developmental regulators are strictly regulated by a multitude of *cis*-regulatory elements. These *cis*-elements can be located upstream, downstream or within introns of the gene itself or of its neighbours (Kleinjan and van Heyningen, 2005). Identification of these elements has importance for disease diagnostics and as tools to gain insight into the regulatory networks in which the target genes participate and which underlie organismal development and the diversity in constituent cell types. Multispecies sequence conservation combined with functional testing for the spatio-temporal specific activity of the identified element in an appropriate model system remains one of the most effective strategies, even though it will miss lineage-specific elements and those whose sequences have strongly diverged between the species used in the comparison.

In this study we have employed a strategy based on ancient sequence conservation to survey the genomic region upstream of the medically important *PAX6* gene. Using this approach we have characterised the 280 kb intergenic interval between *PAX6* and its upstream neighbour *RCN1* for the presence of novel remote enhancers. The chosen model for mammalian regulatory elements is most often the mouse, yet reporter studies in this organism are expensive and time consuming. In accordance with other studies, e.g. (Bessa et al., 2009; Fisher et al., 2006b; Ishibashi et al., 2013), our results indicate that transgenic reporter zebrafish are a suitable alternative model system to test enhancer activity, even when the element is not obviously conserved in zebrafish itself. Due to the additional whole genome duplication (3R) that occurred at the base of the teleost lineage, leading to increased freedom from selective pressure on gene duplicates, teleost fish have experienced an enhanced rate of evolutionary divergence (Lee et al., 2011; Taylor et al., 2003). As a result the zebrafish genome lacks many of the CNEs identifiable in other vertebrate species. We hypothesized that the increased divergence has occurred primarily at the level of the *cis*-regulatory elements themselves, as large changes in the transcriptional machinery involved in the reading of the elements are much less well tolerated because of the far-reaching effect on the pleiotropic functions of transcription factors. Our results indicate that putative regulatory sequences that are anciently conserved between

mammals and a cartilaginous species, such as the elephant shark, can be tested in zebrafish with confidence. Most elements of this type assayed in our study produced reproducible expression patterns in transgenic zebrafish. Three of the agCNE containing fragments were also tested in transgenic mice and produced largely similar patterns of expression, confirming the tissue-specificity observed in the fish. Nevertheless subtle variations between the mouse and fish reporter patterns are found for some elements, which likely reflect differences in the trans-acting environment between the species (Ariza-Cosano et al., 2013). We also show that our zebrafish transgenic approach can be used to dissect the individual contributions to expression patterns generated by larger, multi-CNE containing transgenic reporter constructs in the mouse. Transgenic mice for a construct containing three closely spaced CNEs, called E-55, showed specific expression in the hindbrain and along the spinal cord. By testing the three elements individually in zebrafish we could assign the hindbrain and spinal cord activity to element agCNE6, while agCNEs 7 and 8 were unable to drive expression on their own and may therefore serve a different function that is not uncovered by this type of reporter transgenic experiments. Two further elements that are well conserved in mammals and also present in chicken, E-250 and E-72, but do not exhibit the deep conservation to elephant shark, coelacanth or zebrafish, were also tested. Transgenic reporter fish with these elements failed to reveal any consistent expression pattern, while a transient transgenic assay at E13.5 indicated that at least one of the elements, E-250, has enhancer activity in the mouse. In combination these results suggest that these elements have arisen more recently in the tetrapod lineage after the divergence of the fish lineages and that zebrafish may lack or differ in some aspects of the transcriptional machinery that enable the interpretation of this element. We do not exclude the possibility that other young CNEs can be functional in zebrafish transgenic reporters, but our results suggest that the chances of an informative result will be reduced in comparison with ancient gnathostome CNEs.

Unlike the fully sequenced *C. Milii Pax6* locus, full genomic sequence for the elephant shark is incomplete at present (Venkatesh et al., 2007), but it is envisaged that in time the complete genome of this species will be available for the identification of ancient gnathostome CNEs. Meanwhile our data suggests that the genome of the coelacanth can serve as a useful alternative (Amemiya et al., 2013). All eleven agCNEs conserved in the elephant shark and characterised here were also present in coelacanth. An obvious next step in tracing the evolutionary origins of *PAX6/Pax6* regulatory landscapes will be the comparison with the lamprey and hagfish *Pax6* loci. Although sequencing of the lamprey genome was recently reported (Smith et al., 2013) the status of its assembly is currently insufficient to make reliable conclusions. However, as non-coding element conservation is conspicuously absent in comparisons between the mammalian and the ciona or amphioxus *Pax6* loci (Ravi et al., 2013), the comparative analysis with the lamprey and hagfish loci will be of great interest and may bring new insight into the hypothesis that the rapid emergence of a multitude of regulatory elements was triggered by the two whole genome duplication events that occurred at the base of the vertebrate radiation (Dehal and Boore, 2005; Lowe et al., 2011; McEwen et al., 2009).

The main aim of our study was the identification of novel tissue-specific enhancers for the medically important *PAX6* gene. The gene is expressed in a variety of tissues, both during embryonic and postnatal development and is also be involved in adult tissue maintenance (Hart et al., 2013). A number of medical conditions are associated with inadequate expression levels of PAX6, including some affecting non-ocular tissues such as diabetes, epilepsy and anosmia. Identification of the *cis*-elements for expression in the relevant tissues will provide crucial information for better diagnosis of disease susceptibility. Our characterisation of the regulatory activities residing in agCNE1-11 has revealed novel olfactory system, pineal gland and nervous system enhancers that can now be investigated for functional sequence variation in selected patient groups. Interestingly, several of the agCNEs drive overlapping expression patterns with each other and with previously characterised Pax6 enhancers (Fig. 7). Further analysis may resolve whether these patterns are fully overlapping or whether subtle differences exist in spatial pattern or temporal activity. Largely overlapping patterns of enhancer activity have been reported before for *Pax6* (McBride et al., 2011) and other genes (Montavon et al., 2011). Multiple limb enhancers around the *HoxD* locus have been suggested to form a regulatory archipelago (Montavon et al., 2011). In accordance with this nomenclature the agCNEs defined in our study can be viewed as ancient regulatory islands that form constituent parts of multiple separate archipelagos for a range of tissues. The elucidation of the precise role of the elements within the tissue-specific archipelagos will require further experiments to determine their individual role and careful experimental design will be crucial to capture both unique and redundant functions. The use of dual fluorescence reporter transgenic assays allows detailed analysis of the overlap between expression patterns driven by different enhancer elements in the same animal, and may thus reveal subtle differences in timing of expression or in the identity of positive cells within a tissue or organ.

Careful assessment of reporter expression patterns driven by regulatory elements from a chromosomal locus may also lead to a greater appreciation of gene activity in tissues where its low expression levels or limited number of positive cells had previously overlooked a putative role for the gene. Both these arguments are exemplified by our identification of two distal enhancers for the trigeminal nerve system. The trigeminal ganglion has been little appreciated as a site of Pax6 expression, though its expression in this structure has been noted before (Parisi, 2011; Pieper et al., 2011; Wakamatsu, 2011). Our newly identified enhancers may aid in further experiments to elucidate tissue-specific functions: the role of Pax6 in the trigeminal system is currently not well understood and specific functions are difficult to dissect due to the systemic misdevelopment associated with Pax6 disruption. Using the dual fluorescence approach to investigate overlap in expression between the two trigeminal nerve *cis*-elements we show that both elements drive highly similar expression patterns, suggesting that in their case the enhancers simply add to the robustness of the regulatory architecture, similar to the role fulfilled by shadow enhancers in Drosophila (Frankel et al., 2010; Perry et al., 2010). The presence of multiple separate enhancers activated by different regulatory input signals can provide robustness to the expression of the target gene against variety of variable external conditions. Nevertheless it is remarkable that both elements, as well as other CNEs with functional overlap at the *PAX6* locus, have survived over 450 million years of selective pressure, indicating that maintenance of regulatory robustness has been of crucial importance in vertebrate evolution.

The observation of a non-coding RNA transcript derived from the region upstream of murine *Pax6*, near the agCNE9 and 10 elements is intriguing. Non-coding transcripts have been found at a number of enhancers, but a role has been found for very few of them (Natoli and Andrau, 2012). They have been suggested variably as true functional agents or as unintentional by-products of the recruitment of parts of the transcriptional machinery by the enhancers. In this case a number of arguments suggest a *bona-fide* function for the transcript: the presence of a CpG island near its transcriptional start, the expression in a range of tissues beyond the enhancer's specificity, and the conservation of the putative promoter, agCNE10. However, apart from a potential equivalent in the zebrafish *pax6b*

locus (CT684153 and CT68415), homologous transcripts have so far not been annotated for other species. The AK032637 transcript is separate from the previously described PAX6OS opposite strand transcript (Alfano et al., 2005). Currently the role of both transcripts is unknown, but functions in modulation of *Pax6* transcription or in an independent trans-acting role are conceivable.

In summary, our survey of anciently conserved elements in the 280 kb upstream region between *PAX6* and its nearest protein encoding neighbour *RCN1* has revealed several novel tissue-specific enhancers, whose role in genetic disease and disease susceptibility can now be further assessed. The distinct tissue-specificity of the elements opens the opportunity for directed experimentation to investigate the role of *PAX6* in some lesser-known target tissues. The ancient conservation of the elements is a measure of their crucial functional importance and their organisation into regulatory archipelagos (Fig. 7C) highlights the fact that much remains to be learned about the evolution of the *cis*-regulatory genome.

Acknowledgements

We would like to thank Jacek Mendrychowski, Janice Young and Harris Morrison for technical assistance. Shipra Bhatia was in part supported by NIH grant 1RO1 EY01 8000-02. Work at the IGMM Human Genetics Unit was supported by the Medical Research Council UK. Work in BV's lab was supported by the Biomedical Research Council of A*STAR, Singapore.

Appendix A. Supplementary information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ydbio.2014.01.007.

References

Alfano, G., et al., 2005. Natural antisense transcripts associated with genes involved in eye development. Hum. Mol. Genet 14, 913–923.
Altschul, S.F., et al., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Amemiya, C.T., et al., 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496, 311–316.
Ariza-Cosano, A., et al., 2013. Differences in enhancer activity in mouse and zebrafish reporter assays are often associated with changes in gene expression. BMC Genomics 13, 713.
Ashery-Padan, R., Gruss, P., 2001. Pax6 lights-up the way for eye development. Curr. Opin. Cell Biol 13, 706–714.
Ashery-Padan, R., et al., 2004. Conditional inactivation of Pax6 in the pancreas causes early onset of diabetes. Dev. Biol 269, 479–488.
Bailey, T.L., et al., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208.
Bamiou, D.E., et al., 2004. Deficient auditory interhemispheric transfer in patients with PAX6 mutations. Ann. Neurol. 56, 503–509.
Bessa, J., et al., 2009. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of *cis*-regulatory regions in zebrafish. Dev. Dyn 238, 2409–2417.
Carbe, C., et al., 2012. The functional role of the Meis/Prep-binding elements in Pax6 locus during pancreas and eye development. Dev. Biol. 363, 320–329.
Chanas, S.A., et al., 2009. Effects of elevated Pax6 expression and genetic background on mouse eye development. Invest. Ophthalmol. Vis. Sci 50, 4045–4059.
Davidson, E.H., 2006. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Academic Press, San Diego.
Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3, e314.
Delporte, F.M., et al., 2008. Expression of zebrafish pax6b in pancreas is regulated by two enhancers containing highly conserved *cis*-elements bound by PDX1, PBX and PREP factors. BMC Dev. Biol 8, 53.
Fantes, J., et al., 1995. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum. Mol. Genet 4, 415–422.
Favor, J., et al., 2001. Molecular characterization of Pax6(2Neu) through Pax6(10Neu): an extension of the Pax6 allelic series and the identification of two possible hypomorph alleles in the mouse Mus musculus. Genetics 159, 1689–1700.
Fisher, S., et al., 2006a. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312, 276–279.
Fisher, S., et al., 2006b. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat. Protoc 1, 1297–1305.
Frankel, N., et al., 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466, 490–493.
Frazer, K.A., et al., 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32, W273–W279.
Gehring, W.J., 2005. New perspectives on eye development and the evolution of eyes and photoreceptors. J. Hered 96, 171–184.
Glaser, T., et al., 1994. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. Nat. Genet 7, 463–471.
Gosmain, Y., et al., 2012. Pax6 is crucial for beta-cell function, insulin biosynthesis, and glucose-induced insulin secretion. Mol. Endocrinol 26, 696–709.
Griffin, C., et al., 2002. New 3′ elements control Pax6 expression in the developing pretectum, neural retina and olfactory region. Mech. Dev 112, 89–100.
Grindley, J.C., et al., 1995. The role of Pax-6 in eye and nasal development. Development 121, 1433–1442.
Hammond, C.J., et al., 2004. A susceptibility locus for myopia in the normal population is linked to the PAX6 gene region on chromosome 11: a genome-wide scan of dizygotic twins. Am. J. Hum. Genet 75, 294–304.
Hart, A.W., et al., 2013. The developmental regulator pax6 is essential for maintenance of islet cell function in the adult mouse pancreas. PLoS One 8, e54173.
Hill, R.E., et al., 1991. Mouse small eye results from mutations in a paired-like homeobox-containing gene. Nature 354, 522–525.
Hogan, B.L., et al., 1986. Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. J. Embryol. Exp. Morphol 97, 95–110.
Ishibashi, M., et al., 2013. Using zebrafish transgenesis to test human genomic sequences for specific enhancer activity. Methods 62, 216–225.
Kammandel, B., et al., 1999. Distinct *cis*-essential modules direct the time-space pattern of the Pax6 gene activity. Dev. Biol 205, 79–97.
Kim, J., Lauderdale, J.D., 2006. Analysis of Pax6 expression using a BAC transgene reveals the presence of a paired-less isoform of Pax6 in the eye and olfactory bulb. Dev. Biol 292, 486–505.
Kimmel, C.B., et al., 1995. Stages of embryonic development of the zebrafish. Dev. Dyn. 203, 253–310.
Kioussi, C., et al., 1999. Pax6 is essential for establishing ventral-dorsal cell boundaries in pituitary gland development. Proc. Nat. Acad. Sci. U.S.A. 96, 14378–14382.
Kleinjan, D.A., et al., 2004. Conserved elements in Pax6 intron 7 involved in (auto) regulation and alternative transcription. Dev. Biol 265, 462–477.
Kleinjan, D.A., et al., 2006. Long-range downstream enhancers are essential for Pax6 expression. Dev. Biol 299, 563–581.
Kleinjan, D.A., et al., 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., van Heyningen, V., 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am. J. Hum. Genet 76, 8–32.
Lauderdale, J.D., et al., 2000. 3′ deletions cause aniridia by preventing PAX6 gene expression. Proc. Nat. Acad. Sci. U.S.A. 97, 13755–13759.
Lee, A.P., et al., 2011. Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes. Mol. Biol. Evol 28, 1205–1215.
Levine, M., Davidson, E.H., 2005. Gene regulatory networks for development. Proc. Nat. Acad. Sci. U.S.A 102, 4936–4942.
Lowe, C.B., et al., 2011. Three periods of regulatory innovation during vertebrate evolution. Science 333, 1019–1024.
Manuel, M., et al., 2008. Overexpression of Pax6 results in microphthalmia, retinal dysplasia and defective retinal ganglion cell axon guidance. BMC Dev. Biol 8, 59.
McBride, D.J., et al., 2011. DNaseI hypersensitivity and ultraconservation reveal novel, interdependent long-range enhancers at the complex Pax6 *cis*-regulatory region. PLoS One 6, e28616.
McEwen, G.K., et al., 2009. Early evolution of conserved regulatory sequences associated with development in vertebrates. PLoS Genet 5, e1000762.
McGaughey, D.M., et al., 2008. Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. Genome Res 18, 252–260.
Mitchell, T.N., et al., 2003. Polymicrogyria and absence of pineal gland due to PAX6 mutation. Ann. Neurol 53, 658–663.
Montavon, T., et al., 2011. A regulatory archipelago controls Hox genes transcription in digits. Cell 147, 1132–1145.
Natoli, G., Andrau, J.C., 2012. Noncoding transcription at enhancers: general principles and functional models. Annu. Rev. Genet. 46, 1–19.
Navratilova, P., et al., 2009. Systematic human/zebrafish comparative identification of *cis*-regulatory activity around vertebrate developmental transcription factor genes. Dev. Biol 327, 526–540.
Newburger, D.E., Bulyk, M.L., 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res. 37, D77–D82.
Nishi, M., et al., 2005. A case of novel de novo paired box gene 6 (PAX6) mutation with early-onset diabetes mellitus and aniridia. Diabetic Med 22, 641–644.
Nomura, T., et al., 2007. Role of a transcription factor Pax6 in the developing vertebrate olfactory system. Dev. Growth Differ 49, 683–690.
Osumi, N., et al., 2008. Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. Stem Cells 26, 1663–1672.

Parisi, I., The Roles of Pax6 in the Development of the Sensory System of the Mammalian Head. Ph.D. University of Aberdeen, Aberdeen, 2011.
Perry, M.W., et al., 2010. Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. 20, 1562–1567.
Pieper, M., et al., 2011. Origin and segregation of cranial placodes in Xenopus laevis. Dev. Biol 360, 257–275.
Plaza, S., et al., 1995. Identification and characterization of a neuroretina-specific enhancer element in the quail Pax-6 (Pax-QNR) gene. Mol. Cell Biol 15, 892–903.
Ravi, V., et al., 2013. Sequencing of pax6 Loci from the elephant shark reveals a family of pax6 genes in vertebrate genomes, forged by ancient duplications and divergences. PLoS Genet 9, e1003177.
Schedl, A., et al., 1996. Influence of PAX6 gene dosage on development: over-expression causes severe eye abnormalities. Cell 86, 71–82.
Schwartz, S., et al., 2003. MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Res 31, 3518–3524.
Sievers, F., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol 7, 539.
Simpson, T.I., Price, D.J., 2002. Pax6; a pleiotropic player in development. Bioessays 24, 1041–1051.
Sisodiya, S.M., et al., 2001. PAX6 haploinsufficiency causes cerebral malformation and olfactory dysfunction in humans. Nat. Genet 28, 214–216.
Smith, J.J., et al., 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat. Genet 45 (415-21), 421e1–422e.
St-Onge, L., et al., 1997. Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. Nature 387, 406–409.
Stoykova, A., Gruss, P., 1994. Roles of Pax-genes in developing and adult brain as suggested by expression patterns. J. Neurosci 14, 1395–1412.
Strug, L.J., et al., 2009. Centrotemporal sharp wave EEG trait in rolandic epilepsy maps to Elongator Protein Complex 4 (ELP4). Eur. J. Hum. Genet 17, 1171–1181.
Taylor, J.S., et al., 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res 13, 382–390.
Thisse, C., Thisse, B., 2008. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat. Protoc 3, 59–69.
van Heyningen, V., Williamson, K.A., 2002. PAX6 in sensory development. Hum. Mol. Genet. 11, 1161–1167.
Venkatesh, B., et al., 2006. Ancient noncoding elements conserved in the human genome. Science 314, 1892.
Venkatesh, B., et al., 2007. Survey sequencing and comparative analysis of the elephant shark (Callorhinchus milii) genome. PLoS Biol 5, e101.
Venkatesh, B., et al., 2005. A compact cartilaginous fish model genome. Curr. Biol. 15, R82–R83.
Visel, A., et al., 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res 35, D88–D92.
Wakamatsu, Y., 2011. Mutual repression between Pax3 and Pax6 is involved in the positioning of ophthalmic trigeminal placode in avian embryo. Dev. Growth Differ 53, 994–1003.
Walther, C., Gruss, P., 1991. Pax-6, a murine paired box gene, is expressed in the developing CNS. Development 113, 1435–1449.
Warren, N., Price, D.J., 1997. Roles of Pax-6 in murine diencephalic development. Development 124, 1573–1582.
Westerfield, M., 2007. The Zebrafish Book: A Guide for the Laboratory use of zebrafish Danio rerio. University of Oregon Press, Eugene, OR
Williams, S.C., et al., 1998. A highly conserved lens transcriptional control element from the Pax-6 gene. Mech. Dev 73, 225–229.
Xu, P.X., et al., 1999. Regulation of Pax6 expression is conserved between mice and flies. Development 126, 383–395.
Yasuda, T., et al., 2002. PAX6 mutation as a genetic factor common to aniridia and glucose intolerance. Diabetes 51, 224–230.
Zhang, X., et al., 2003. Cre-loxp fate-mapping of Pax6 enhancer active retinal and pancreatic progenitors. Genesis 35, 22–30.