Int. J. Dev. Biol. 53: 935-946 (2009)
doi: 10.1387/ijdb.082819ck

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

Identification of regulatory elements in the *Isl1* gene locus

CLAUDIA KAPPEN* and J. MICHAEL SALBAUM

Pennington Biomedical Research Center, Baton Rouge, LA, USA

ABSTRACT Isl1 is a LIM/homeodomain transcription factor with critical roles for the development of the heart, the nervous system and the pancreas. Both deficiency and mis-expression of Isl1 cause profound developmental defects, demonstrating the importance of proper regulation of *Isl1* gene expression during development. In order to understand the mechanisms that control *Isl1* expression during embryogenesis and in tissue differentiation, we initiated a screen for gene regulatory elements in the *Isl1* locus using a novel dual reporter gene vector that allows screens of large genomic regions through reporter gene assays *in vitro* and *in vivo*. We identified regions from the *Isl1* gene locus that confer transcriptional activity in pancreatic cell lines *in vitro*. Using transgenic mice, we furthermore discovered an enhancer with *in vivo* specificity for the developing heart, as well as visceral and posterior mesoderm. Our findings further suggest that Foxo1 as well as Gata4 contribute to the activity of this enhancer in the developing embryo. We conclude that *Isl1* gene expression is controlled in modular fashion by several elements with distinct functionality. Embryonic *Isl1* expression in several tissues of mesodermal origin is driven by a specific enhancer that is located 3-6kb downstream of the gene.

KEY WORDS: *transgenic reporter, enhancer, heart mesoderm, cardiac crescent, lateral mesoderm*

Introduction

Gene transcription initiates at the core promoter and is often controlled by distant DNA sequence elements that act as enhancers or silencers of transcription and control the temporal and cell-type-specific gene expression. The identification of such regulatory elements relies on two principal experimental approaches: (i) Transfection studies in cultured cells in which DNA fragments are tested for capacity to activate transcription of a reporter gene in the relevant cell type. Advantages are quantitative readouts and scalability to high-throughput; limitations are imposed by the availability of suitable cell lines, and the need for validation *in vivo*, typically by targeted recombination or transgenesis. (ii) Regulatory elements with activity *in vivo* are frequently ascertained by a reporter gene approach in transgenic mice. Large genomic regions of the size of Bacterial Artificial Chromosomes (BAC) or smaller DNA fragments are tested for tissue- and stage-specific activation of reporter gene transcription. Compelling advantages are the immediate *in vivo* validation, and concomitant assessment of tissue-specificity as well as temporal regulation. However, the major drawback lies in the serial nature of this approach, as it either requires successive deletions (Yaworsky *et al.*, 1997; Yaworsky and Kappen, 1999) or consecutive testing of multiple DNA fragments (DiLeone *et al.*, 1998). Typically, both approaches also employ different reporter genes.

To simplify construct preparation for the discovery of regulatory elements, we have developed a unique dual reporter construct that allows parallel assay of any given DNA fragment for enhancer activity *in vivo* as well as *in vitro*. With a library of the genomic locus encompassing the *Isl1* gene in this vector, we were able to screen for and discover regulatory elements in a region spanning over 200kb. This scope exceeds the current limitations of screens for gene control elements, and suggests that genome-wide approaches to identify such elements may eventually become feasible.

The LIM-homeodomain transcription factor Islet 1 (Isl1) was initially identified as an insulin-promoter binding protein (Karlsson *et al.*, 1990). It is known to be required for development of the

Abbreviations used in this paper: da, dorsal aorta; ECR, evolutionary conserved region; fg, foregut; fgd, foregut diverticulum; g, gut; h, heart; hb, hind brain; IRES, internal ribosome entry site; IRS, insulin-responsive sequence; lacZ, β-galactosidase; lb, limb bud; lpm, lateral plate mesoderm; lv, left ventricle; mb, midbrain; mn, motor neurons; nd, nephric duct; nt, neural tube; ofp, olfactory pit; oft, outflow tract; op, optic pit; pe, pharyngeal endoderm; vm, visceral mesoderm.

***Address correspondence to:** Dr. Claudia Kappen.Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, LA 71010, USA. Tel: +1-225-763-2781. Fax: +1-225-763-0261. e-mail: Claudia.Kappen@pbrc.edu

Accepted: 26 January 2009. *Published online:* 24 June 2009.

ISSN: Online 1696-3547, Print 0214-6282

dorsal pancreas and for differentiation of insulin-producing cells (Ahlgren *et al.*, 1997). In addition, Isl1 controls differentiation of motor neurons in the ventral spinal cord (Ericson *et al.*, 1992; Pfaff *et al.*, 1996). Isl1 deficiency leads to defects of the dorsal aorta and the heart (Cai *et al.*, 2003) in developing embryos. *Isl1* is also expressed in specific regions of the brain (Thor *et al.*, 1991), in the eye (Galli-Resta *et al.*, 1997), and in teeth (Mitsiadis *et al.*, 2003).

We have recently shown that over-expression of *Isl1* in transgenic mice causes growth defects that resemble caudal regression/sacral agenesis (Muller *et al.*, 2003). This birth defect in humans is highly associated with maternal diabetes during pregnancy (Goto and Goldman, 1994), constituting one of the phenotypes of diabetic embryopathy, which encompasses heart, neural tube and caudal growth defects. From the similarity of the *Isl1*-induced abnormalities in mice to the caudal growth defects in human diabetic embryopathy, we derived the hypothesis that *Isl1* could be involved in the embryonic response to maternal diabetes (Muller *et al.*, 2003). One implication is that exposure to diabetes would lead to misregulation of *Isl1* in the developing embryo, and hence, the *Isl1* locus would have to contain regulatory elements that sense metabolic status and become activated in the specific tissues affected in diabetic embryopathy: the posterior mesoderm, the neural tube and the heart.

Regulatory elements for normal transcription of the *Isl1* gene have been identified for cells in the developing nervous system (Higashijima *et al.*, 2000; Uemura *et al.*, 2005; Bejerano *et al.*, 2006), but elements for the control of *Isl1* expression in many other tissues, such as heart, pancreatic cells, and posterior mesoderm have not been described to date. The present study was undertaken to identify such regulatory control elements for future investigations whether their temporal and tissue-specific activity could account for a pathogenic role of *Isl1* in diabetic embryopathy.

Results

Isl1 gene function has been explored predominantly in the pancreas, motor neurons of the spinal cord, and recently in the heart. To define the regulatory mechanisms for expression of *Isl1*, we initially created a *LacZ* reporter construct containing 3kb of genomic DNA upstream of the transcription start site. We generated 14 independent transgenic mouse embryos harboring this construct, but observed no reporter activity at 10.5 days of gestation (E10.5) in any of the embryonic cell types that express *Isl1*. A reporter construct with an additional 7kb of upstream DNA also did not yield consistent β-galactosidase activity in 20 independent transgenic embryos, indicating that the immediate 5'-upstream region of *Isl1* did not contain the necessary DNA control elements to drive expression. We concluded from these results that promoter-proximal regions of *Isl1* do not by themselves control cell-specific *Isl1* gene expression in the embryo.

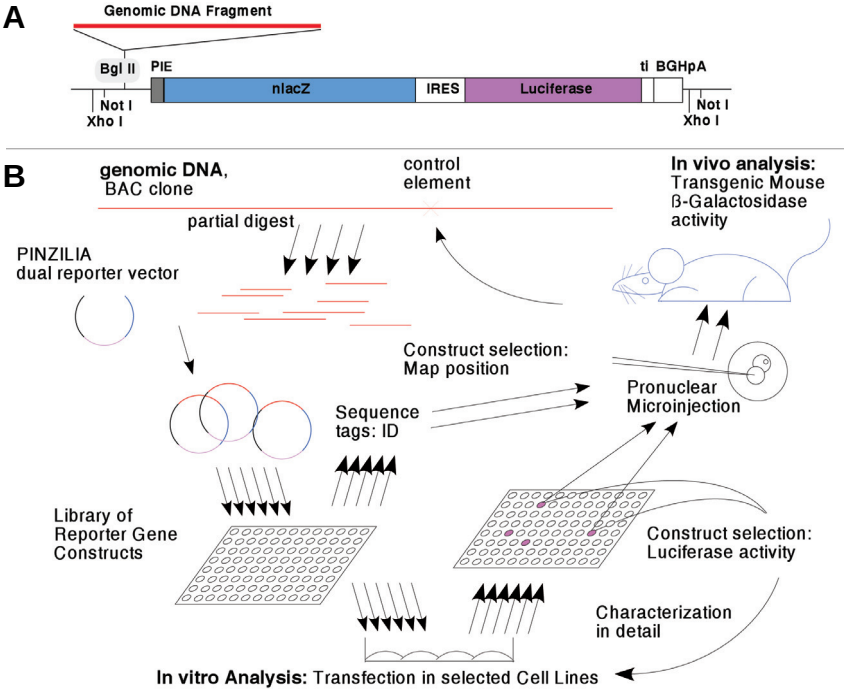


Fig. 1. Dual reporter library approach applied to study the regulation of *Isl1* gene expression. (A) *The dual reporter gene vector PINZILIA consists of a BglII cloning site for genomic DNA fragments, a minimal promoter from the HSV ICP4 gene (PIE), a LacZ gene encoding a nuclear-targeted β-galactosidase reporter, an internal ribosome entry site, a firefly Luciferase reporter gene, the intron from the SV40 small t-antigen gene, and a polyadenylation site from the Bovine Growth Hormone gene. Transcription driven by enhancer elements on the genomic DNA insert leads to translation of both reporter genes, with β-galactosidase used for transgenic mouse experiments* in vivo, *and Luciferase for transient transfection experiments* in vitro. **(B)** *A reporter library approach allows to convert larger genomic regions, e.g. as represented on BAC clones, into reporter constructs for mapping and functional analysis* in vivo *or* in vitro.

A dual reporter vector for analysis of gene regulation

To further avoid consecutive testing of DNA fragments, we developed an alternative strategy for construction and assay of reporter constructs. Since it would be desirable to evaluate any given reporter construct qualitatively in transgenic mouse embryos as well as quantitatively in cultured cells, we constructed a vector that contained two reporter genes: (i) a *LacZ* gene suitable for histological analyses in transgenic mice, and (ii) a firefly *Luciferase* gene for quantitative studies in cultured cells. Both reporter genes are located on a single transcript under the control of a minimal promoter and are joined by an internal ribosome entry site. The structure of the vector is shown in Fig. 1A. Cloning of genomic DNA in front of the minimal promoter allows testing for transcription-activating activity using the readout from either reporter gene.

A reporter construct library for the mouse Isl1 *gene*

We generated a library of reporter constructs with an average insert size of 8-10kb from two BAC clones covering the *Isl1* gene locus (Fig. 1B). This size is almost two orders of magnitude larger than typical individual regulatory elements, which tend to be in the range of several hundred base

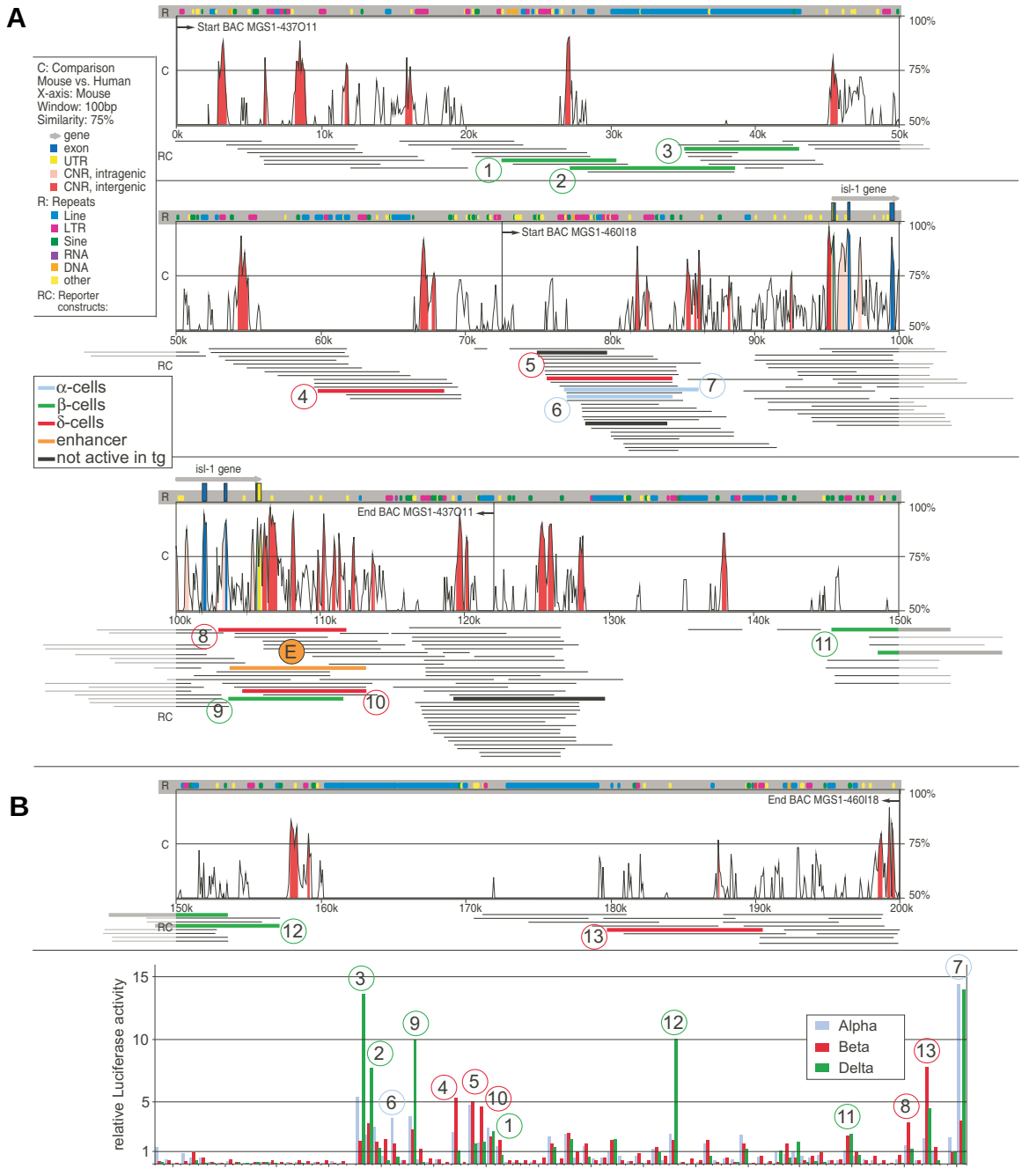


Fig. 2. Reporter library for the 200kb *Isl1* genomic region. (A) *Analysis of the murine* Isl1 *gene locus for evolutionary conservation using VISTA. 200kb of DNA sequence representing two BAC clones that each contains the* Isl1 *gene are represented in four tiers of 50kb each. Conservation (C) and color code for repetitive sequences (R) are indicated. Reporter constructs (RC) are shown as thin lines below the conservation plot. Thick black lines represent constructs tested* in vivo *in transgenic mice; the thick orange line (labeled "E") indicates the construct 3cF10 with enhancer activity in transgenic mice. The library consists of 202 clones with an average insert size of 8.5kb; 131 constructs could be mapped to the genomic sequence on the basis of sequence tags; 39 clones were mapped only on one end due to presence of repetitive sequences on one end of the genomic insert; and 32 constructs contain mouse genomic DNA derived from the BAC clones but could not be mapped unambiguously due to repetitive sequences in both sequence tags. Colored thick lines (see color code) and circled numbers refer to constructs with activity in transfection experiments shown in B.* **(B)** *Evaluation of the* Isl1 *reporter library by transfection in pancreatic cell lines. 94 reporter constructs (together with positive and negative controls) were transfected into three different pancreatic cell lines: Alpha TC-6 for glucagon-producing alpha cells, BetaTC-3 for insulin-producing beta cells and TGP61 for somatostatin-producing delta cells. Transfections were carried out in 96-well format in triplicate, and Luciferase activity was normalized to the mean of the library to yield relative activity for each construct. Activity was plotted for alpha (blue), beta (green) and delta cells (red). Circled numbers refer to the position of the construct on the map in A.*

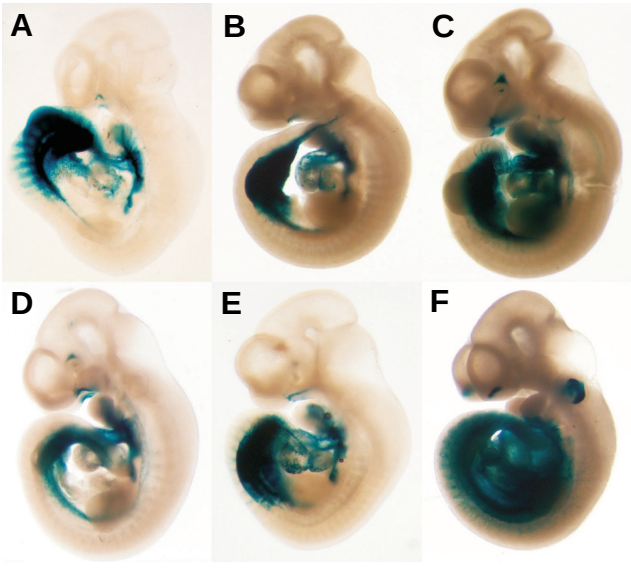


Fig. 3. The reporter construct 3cF10 contains an enhancer from the *Isl1* gene. (A-D). *Analysis of reporter activity at 10.5 days of gestation demonstrated that independent transgenic mouse embryos harboring the construct 3cF10 and stained for β-galactosidase show a highly similar pattern of reporter expression. This indicates that the staining pattern is due to a biological function of the transgene, and suggests the presence of a transcriptional enhancer on the construct. This common pattern of staining is found in the heart and the outflow tract region, the visceral mesoderm region between and posterior to the hind limb buds, and the olfactory pit.* **(A)** *Transient transgenic embryo.* **(B)** *Embryo sired from founder #3.* **(C)** *From founder #5.* **(D)** *From founder #9.* **(F)** *From founder #11.* **(J)** *From founder #21.*

pairs. However, regulatory activity is often exerted through multiple cooperative regulatory elements (Davidson, 2006), and we chose a size class of 8-10kb in order to keep potential regulatory superstructures intact. The resulting DNA constructs were identified via sequence tags from both ends of the genomic insert. We used sequence comparisons to eliminate constructs containing BAC vector DNA or *E.coli* genomic DNA; a total of 200 constructs with inserts of mouse genomic DNA were used for all further studies.

Sequence analysis and annotation of the 129 Sv/J mouse Isl1 *gene locus*

The sequence covered by the two original BAC clones starts approximately 95kb 5'-upstream of the */s/1* gene and extends about 97kb 3'-downstream; it is colinear with genomic DNA sequence from the human */s/1* gene locus on chromosome 5: Fig. 2A shows a VISTA plot of a mouse/human sequence comparison revealing conserved regions, location of repetitive elements, and location of reporter constructs. As expected, we obtained high sequence coverage in the regions where the two BAC clones overlap, including the */s/1* gene itself. We also observed regions with apparent low or no construct coverage. In part, this was due to the high content of repetitive DNA sequence elements: unequivocal positional assignment of reporter constructs was not possible when both ends of the genomic insert of a given construct originated from regions of repetitive DNA. Gene regulatory elements are often composed of sequences that are conserved between mammals, and possibly other vertebrates. All regions of human-mouse conservation were well covered by our library, with the exception of one conservation peak (at ~158kb, Fig. 3) close to an extended region of repetitive DNA that was not captured.

Transfection of the Isl1*-reporter library into pancreatic cell lines*

Since Isl1 contributes to the regulation of the genes for the pancreatic hormones insulin, glucagon and somatostatin, we chose cell lines representing alpha-, beta-, and delta-cells of pancreatic islets for our transfection analysis. A series of transfection experiments with 94 constructs is shown in Fig. 2B. We were able to identify several constructs with distinct activities in the three different cell types, which indicates that these clones harbor specific regulatory elements of the */s/1* gene. Although */s/1* itself is active in all three islet cell types, we did not detect any constructs that were active in all three cell lines. We did observe several constructs that had specific activities in one

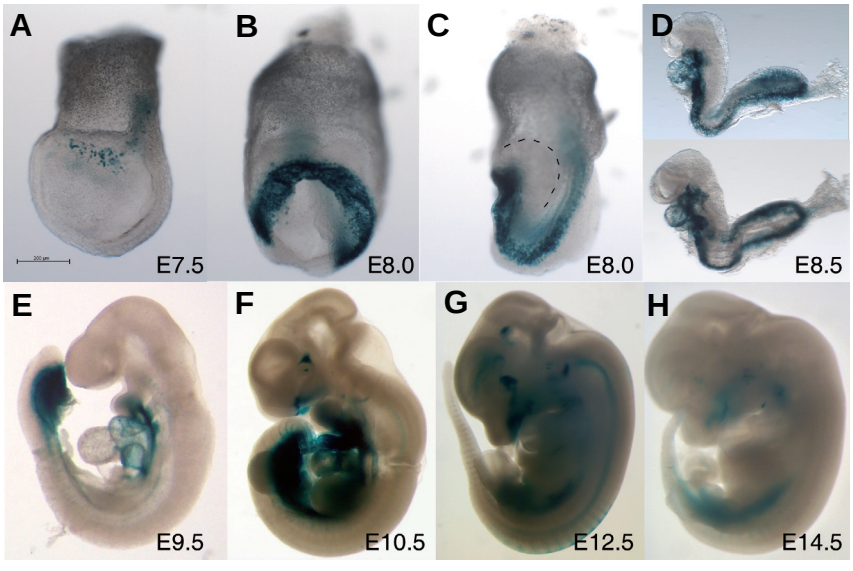


Fig. 4. 3cF10 reporter activity at different stages of development. *Embryos were generated from offspring of founder #9 and # 21.* **(A)** *The first signs of reporter activity appear at E7.5 in a region of migrating mesodermal cells.* **(B)** *frontal and* **(C)** *lateral view of an embryo at E8.0 with staining in heart and lateral plate mesoderm. The stippled line (C) indicates the outline of the neural folds.* **(D)** *two embryos at E8.5 showing staining in the lateral plate mesoderm, the developing heart and the foregut diverticulum.* **(E)** *Embryo at E9.5, with continuing reporter activity in heart and visceral mesoderm.* **(F)** *Strong reporter activity is observed at E10.5 in the heart and outflow tract and in visceral mesoderm.* **(G)** *The reporter pattern at E12.5 begins to lose definition and specificity.* **(H)** *At E14.5, the reporter staining patters has very much degraded, indicating that the enhancer element on the 3cF10 construct is not able to sustain Isl-1 expression in late stages of development, and suggesting that additional regulatory elements are required for proper control of Isl-1 expression.*

or two cell lines, which lead us to conclude that controlled pancreatic expression of *Isl1* is brought about by combinatorial action of several control elements.

An enhancer of Isl1 *gene expression located downstream of* Isl1

To scan the Isl1 gene and its close vicinity for potential regulatory elements, we tested several reporter constructs that covered the *Isl1* gene itself as well as adjacent regions in transient transgenic mouse reporter gene experiments. Reporter construct DNAs were used for pronuclear injection, and founder embryos were analyzed directly for LacZ staining at embryonic day E10.5. We obtained 4 transgenic embryos for a construct spanning a region from –20kb to –16kb (relative to the transcription start of *Isl1*); 7 transgenic embryos for a construct ranging from –17kb to –12kb; only 2 transgenic embryos with a construct covering the Isl1 coding region itself (+0.5kb downstream of the transcription start to 0.9kb upstream of the transcription stop); and 4 transgenic embryos with a construct covering the region from +24.4kb to +35kb downstream relative to the transcription start. While more exhaustive tests may be desired, we did not observe lacZ reporter staining from any of these constructs in the initial experiments that would suggest that the regions covered by these constructs might contribute to the control of *Isl1* expression at E10.5.

However, this set of experiments did reveal a single construct with strong and consistent reporter gene activity in transgenic mouse embryos at E10.5. The construct 3cF10 with an insert of 10kb covered a region beginning 8.5kb downstream of the transcription start site and ending 7kb downstream of the transcription stop of the *Isl1* gene. This construct is indicated in Fig. 2A, and the genomic insert is shown in more detail in Fig. 6. The initial transgenic experiment yielded five embryos containing the 3cF10 transgene, and four of these embryos showed LacZ staining in a pattern common to all four specimen, indicating the presence of a functional enhancer element on this construct. We therefore generated stable transgenic mouse lines with the 3cF10 reporter construct for a more detailed analysis. Five more expressing transgenic founders were obtained; Fig. 3 shows the β- galactosidase staining of one transient transgenic embryo from the initial experiment (Fig. 3A), and of embryos sired from each founder (Fig. 3 B,C,D,F,J). All embryos exhibited consistent reporter activity in the dorsal aorta, the outflow tract and the heart, the visceral mesoderm in the region of the hind limb bud, and in the olfactory pit.

Fig. 4 shows a temporal analysis of 3cF10 reporter activity in transgenic embryos at different stages of development. The temporal sequence of reporter activity began at E7.5, with staining detectable in a small number of cells that were found in locations consistent with migrating heart mesoderm cells. At E8.0, strong staining was observed in the cardiac crescent and the lateral plate mesoderm. At E8.5, staining in the heart tube itself was spotty, consistent with the previously described loss of *Isl1* expression in differentiating cardiomyocytes (Prall *et al.*, 2007; Sun *et al.*, 2007), whereas strong staining was again visible in the lateral plate mesoderm and in the foregut diverticulum. Reporter activity at E9.5 remained present in heart and outflow tract as well as the pharyngeal endoderm, whereas visceral mesoderm staining was mostly found at the level of the hindlimb and no longer along the entire gut tract. This pattern continued to be sharply delineated at E10.5. However, analysis of later stages at E12.5 and E14.5 revealed that the reporter pattern lost definition, and slowly degraded.

We found that the β- galactosidase staining pattern elicited by the 3cF10 construct was not only common between independent transgenic events at one given developmental time point, it was also consistent between independent transgenic lines through a developmental time series, although the lines exhibited slightly different levels of intensity of the staining. We therefore concluded

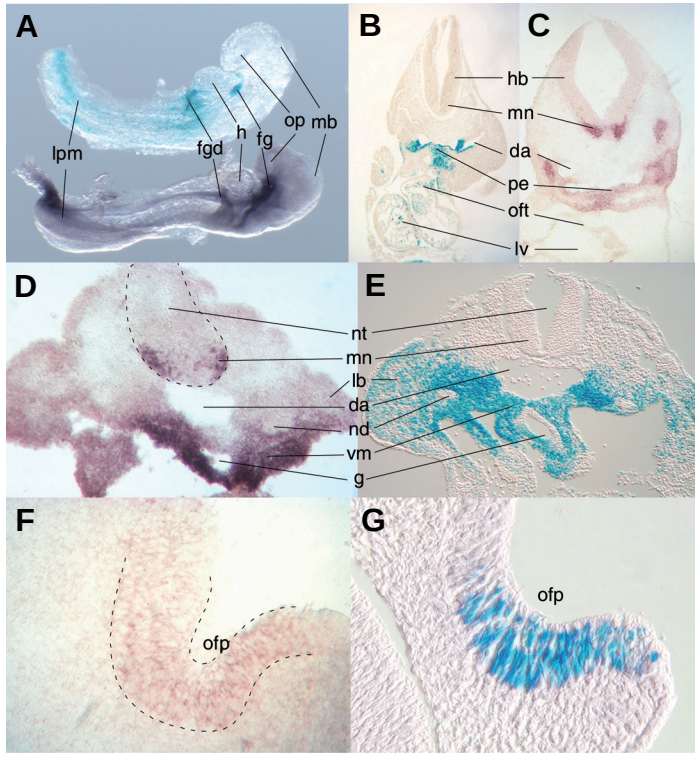


Fig. 5. Comparison of normal *Isl1* expression and reporter gene activity in 3cF10-transgenic embryos. (A) *Embryos at E8.5 stained for lacZ reporter activity (top) or processed for whole-mount* in situ *hybridization to demonstrate* Isl1 *expression (bottom). Expression and reporter signals coincide for the foregut pocket, the foregut diverticulum, the lateral plate mesoderm, and at lower levels, in the heart (fg, foregut, fgd, foregut diverticulum, h, heart, lpm, lateral plate mesoderm, mb, midbrain, op, optic pit). Section from embryos at E10.5 show 3cF10 reporter gene activity* **(B)** *matching Isl1 expression* **(C)** *in the wall of the dorsal aorta, the pharyngeal endoderm, the outflow tract, and in the heart. Reporter activity in hind brain motor neurons (where Isl1 is expressed) is noticeably absent, indicating that the 3cF10 construct does not contain regulatory elements that control* Isl1 *expression in the nervous system (da, dorsal aorta, hb, hind brain, lv, left ventricle, mn, motor neurons, oft, outflow tract, pe, pharyngeal endoderm).* **(D)** *A section at E10.5 from the hind limb level reveals* Isl1 *expression in motor neurons of the neural tube, in parts of the wall of the dorsal aorta, in the mesenchyme underlying the dorsal aorta and surrounding the gut, and weaker expression in the limb bud mesenchyme.* **(E)** *With the exception of spinal cord motor neurons, reporter activity in 3cF10 transgenic mice reflects the endogenous* Isl1 *gene expression seen at the hind limb level (da, dorsal aorta, g, gut, lb, limb bud, mn, motor neurons, nd, nephric duct, nt, neural tube, vm, visceral mesoderm).* **(F)** Isl1 *expression matches 3cF10 reporter activity* **(G)** *in the olfactory pit (ofp).*

that this common pattern of reporter expression in multiple, independently generated transgenic embryos reflects a functional biological property of the genomic DNA sequence contained on the construct. We further conclude that this pattern is generated by a bona fide gene regulatory element - presumably an enhancer - since it could activate a heterologous promoter on the 3cF10 reporter construct.

Isl1 *gene expression and reporter activity*

To assess the specificity of the 3cF10 reporter pattern relative to */s/1* gene expression, we compared reporter activity to *in situ* hybridization results obtained with an */s/1* antisense riboprobe on normal embryos of comparable stage. We consistently found a high congruence between reporter activity and */s/1* gene expression (Fig. 5) in tissues of mesendodermal origin, and the sites labeled by 3cF10 reporter activity at various stages of development were all sites with expression of the endogenous */s/1* gene. One notable exception were the neuronal elements of */s/1* expression (Dong *et al.*, 1991; Ericson *et al.*, 1992): motor neurons of the spinal cord and the dorsal root ganglia were clearly labeled by the antisense riboprobe for */s/1*, but did not exhibit reporter gene activity in 3cF10-transgenic embryos. This finding confirmed published observations (Uemura *et al.*, 2005; Bejerano *et al.*, 2006) that the particular enhancer element(s) responsible for neuronal */s/1* gene expression are located elsewhere. Furthermore, the congruence between */s/1* gene expression and 3cF10 reporter activity was not maintained at later stages of development (e.g. E12.5 or E14.5), indicating that the 3cF10 capabilities for temporal-specific control of expression are limited to early stages of development. Together, these data indicated that the 3cF10 reporter construct faithfully recapitulated several (but not all) aspects of */s/1* gene expression, suggesting that sequences on the 3cF10 construct contribute to the regulation of */s/1* gene expression in the mouse embryo.

Sequence analysis of the 3cF10 reporter construct

Given the evolutionary conservation of */s/1* gene function and expression (Ericson *et al.*, 1992; Korzh *et al.*, 1993; Gong *et al.*, 1995; Yuan and Schoenwolf, 2000), we reasoned that the enhancer element on the 3cF10 construct might be conserved, and that such sequence conservation could be used to delineate the specific DNA sequences on the 9751bp genomic insert that might contribute to the enhancer activity of 3cF10. We therefore determined which regions were evolutionarily conserved by comparing the 3cF10 sequence to genomic sequences from human, dog, rat, chick, and *Xenopus* using the Genome Alignment function of the ECR browser (Ovcharenko *et al.*, 2004). A VISTA plot (Frazer *et al.*, 2004) of the resulting alignment (Fig. 6) showed that there are many conserved sequence elements on the 3cF10 sequence when compared to other mammalian species. However, only three such regions of conservation remained if chick genomic data were taken into consideration; two of these regions were also conserved in frog. The third region was also present in frog but fails the standard cutoff criteria for VISTA analyses (100bp, >75% similarity); nevertheless, as this smaller conservation peak clearly contains sequence similarity to all other species, we included it in further analyses. Highly conserved regions may confer gene regulatory activity – here as an enhancer - and since enhancers consist of transcription factor binding sites, we used rVISTA (Loots and Ovcharenko, 2004) to predict potential transcription factor binding sites on the three conserved sequence regions. This examination revealed conserved transcription factor binding sites in all three regions (Fig. 6). Based on the level of deep conservation, we concluded that these DNA regions of the 3cF10 construct were excellent candidates for functional enhancer sequences. Intriguingly, potential transcription factors that could

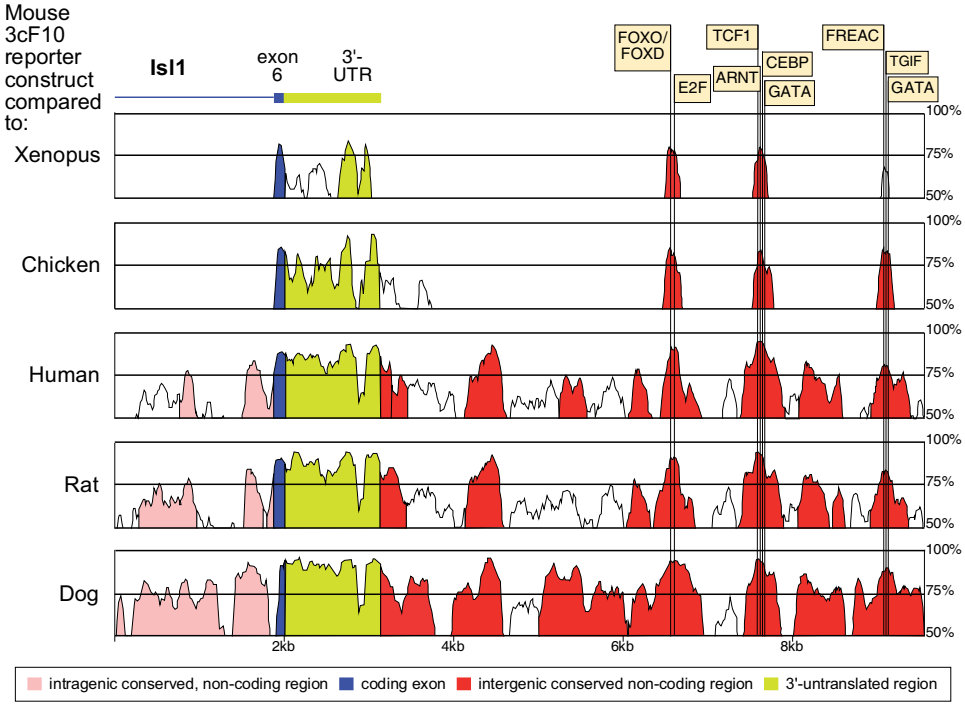


Fig. 6. Conserved sites in the 3cF10 enhancer. *Multispecies conservation analysis of the 10kb genomic DNA sequence from the 3'-end of the* Isl1 *gene that is contained on the 3cF10 reporter construct compared to the respective genomic sequences from* Xenopus*, chicken, human, rat, and dog, using VISTA. Degree of conservation between 50 and 100% is plotted on the x-axis; pink shading highlights intragenic conserved sequences; blue indicates exons; yellow marks transcribed untranslated regions; and red represents conserved intergenic sequences. Comparison to bird and amphibian sequences revealed three regions of conservation that were analyzed for the presence of conserved transcription factor binding sites using rVISTA. Potential transcription factor binding sites that were found to be conserved through all species are shown as flags above the conservation peaks.*

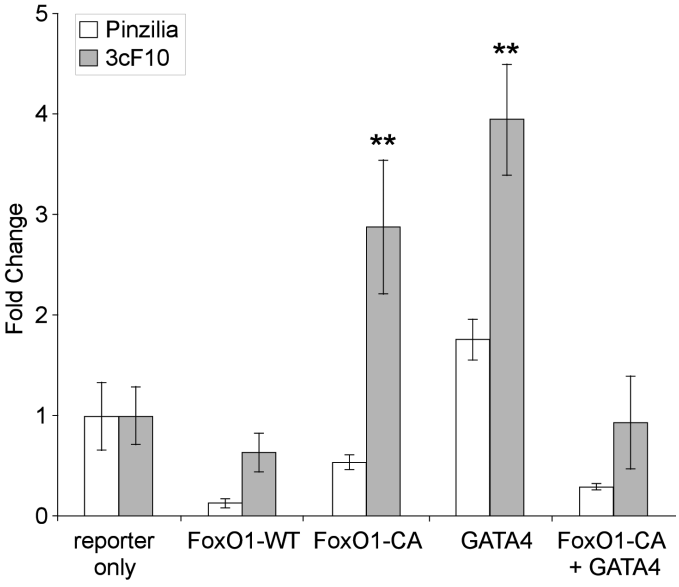


Fig. 7. Regulation of the 3cF10 enhancer. *Co-transfection of BHK-21 fibroblast cells with reporter constructs and plasmids expressing transcription factors. The reporter vector (Pinzilia), and the 3cF10 reporter construct were transfected into BHK-21 cells either with or without a plasmid expressing a wildtype version of the Foxo1 transcription factor, a plasmid expressing a constitutively active form of Foxo1, a plasmid expressing the transcription factor Gata4, or a mixture of the constitutively active form of Foxo1 and of Gata4. Positive controls (not shown) were SV40-Luc, and IRS3TK-Luc, which responded as expected to the presence of constitutively active FoxO1. The 3cF10 reporter construct from the* Isl1 *gene locus shows activation by either the constitutively active form of Foxo1, or by Gata4, but not by a combination of both factors. ** represents a p-value below 0.001.*

bind to motifs in the conserved DNA regions of 3cF10, namely Foxo1 and Gata4, have previously been implicated in cardiovascular development (Kelley *et al.*, 1993; Laverriere *et al.*, 1994), and 3cF10 enhancer activity was indeed observed in the developing heart, the outflow tract and the ventral part of the dorsal aorta (see Panels B and E in Fig. 5). Our results thus suggest that these factors might participate in enhancer function of the 3cF10 construct and in the regulation of the *Isl1* gene in the developing vasculature.

Cotransfection of transcription factor genes and the 3cF10 reporter

To test the hypothesis that the transcription factors identified by sequence conservation on a DNA fragment with proven enhancer activity *in vivo* might participate in the biological function of the enhancer, we carried out reporter gene assays in cultured cells. We chose to use BHK-21 fibroblasts that do not express *Isl1* endogenously, and therefore provide the opportunity of testing transcription factor effects on the reporter construct in an ectopic context and without interference from the endogenous gene. BHK-21 cells also represent a heterologous context with respect to Gata4 (Ritz-Laser *et al.*, 2005). A cotransfection analysis (Fig. 7) in BHK-21 fibroblasts demonstrated that the 3cF10 construct is activated by either Foxo1 or by Gata4. As expected, the unmodified form of Foxo1 did not activate either the control construct (not shown) or 3cF10. The relatively moderate (3- to 4-fold) activation of the 3cF10 reporter by either a constitutively active form of Foxo1 (Guo *et al.*, 1999), or by Gata4, may be attributable to the fact that the plasmid contains approximately 10kb of genomic DNA, and drives *Luciferase* via an IRES sequence. The size of the genomic insert may affect the efficacy of transcriptional activation when compared to e.g. a synthetic IRS response element, and the IRES sequence affects translation of the *Luciferase* gene. For reference, a synthetic promoter with 3 Foxo1-responsive IRS elements driving a *Luciferase* reporter directly responded to a co-transfection with the constitutively active form of Foxo1 (Guo *et al.*, 1999) with approximately 20-fold higher activity (not shown). Interestingly, co-transfection of the 3cF10 reporter with both the constitutively active form of Foxo1 and Gata4 together negated the effect that either transcription factor exerted individually. We conclude from these experiments that both Foxo1 and Gata4 are candidate transcription factors that could contribute to the mechanism of *Isl1* gene regulation, but may not simply act by activation of expression.

Discussion

Reporter construct library in a dual-reporter vector

With the goal to understand the mechanisms that control expression of *Isl1*, we applied a reporter gene approach to identify DNA regulatory elements in the *Isl1* gene locus. The strategy to initially generate reporter libraries from larger genomic regions is designed to address the problem that regulatory elements are often not located in the promoter region proximal to a gene. The *Isl1* gene is a case in point: all regulatory elements that have been described for this gene (this study and Uemura *et al.*, 2005; Bejerano *et al.*, 2006) are located at significant distances from the transcription start site. Generating a library together with sequence identification tags for each construct parallelizes the task of covering a larger area with reporter constructs. While the dual-reporter vector doesn't address the intrinsic caveats of biochemical transfection results *in vitro* for biology *in vivo*, it nevertheless streamlines the evaluation of reporter constructs *in vivo* and *in vitro*: combining the advantages of *LacZ* as well as firefly *Luciferase* avoids the typical need to reclone a genomic inserts into different vectors. As we have demonstrated with the 3cF10 enhancer of the *Isl1* gene, a construct with activity in transgenic assays can be used directly for further characterization by co-transfection assays. In similar fashion, constructs with activity *in vitro* could be used directly in transgenic mice.

One limitation inherent in our current dual reporter vector design is that reporter gene expression requires transcription from the TATA box-containing minimal promoter on the vector. However, if a potential regulatory element has preference for interactions with non-TATAA core promoters, such as e.g. Initiator elements, its activity may be reduced in our assays. The mouse *Isl1* gene does not display a readily identifiable TATA box, but contains a sequence of 5'-TCAGACC-3' at position –42 relative to the annotated transcript start (chr13:117430367, UCSC mm8). This sequence is a perfect match to the Initiator consensus 5'-YYANWYY-3' (Schug *et al.*, 2005). The same sequence is present in the human *Isl1* gene at position –67 relative to the annotated transcription start site of the human *ISL1* RefSeq gene (NM_002202; chr5:50715026, UCSC hg18). If the regulation of *Isl1* expression were to involve control elements that act specifi-

cally upon Initiator-type core promoter elements, it would be possible that some regulatory elements from the *Isl1* gene may have been missed in our assay. The rationale for choosing a TATA box-containing minimal promoter for the dual reporter vector was twofold: (i) we had previously shown in transgenic mouse experiments that it produced very little background activity by itself in the absence of an enhancer (Yaworsky *et al.*, 1997; Yaworsky and Kappen, 1999), and (ii) computational analyses of genes with high specificity of expression in pancreas showed that such genes preferentially contained a TATA box as core promoter element (Schug *et al.*, 2005). Thus, for the initial proof-of-principle, the choice of a TATAA-containing core promoter was well justified. Nevertheless, it may be desirable to include non-TATAA configurations in the future, either in the same vector backbone, or in new reporter libraries from the same gene locus.

We included DNA from a 202kb genomic region around the mouse *Isl1* gene locus in our study. Choice of insert size for the library leads to a compromise between library size, intended coverage of the region, and integrity of regulatory structures. The insert size of the library, and therefore the resolution of the approach, was about 8kb. Since typical enhancers are only several hundred base pairs in size, reporter activity of a construct with an average insert size of 8kb will not identify enhancer sequences directly. Rather, at this resolution, the strategy is designed to survey a large genomic area for regions with positive regulatory – i.e. transcription-activating – capacity; such regions must then be subjected to detailed analysis. In contrast to methods that rely on computational prediction of potential regulatory elements (Pennacchio and Rubin, 2001; Pennacchio *et al.*, 2006), the reporter library represents an unbiased approach within the limits of library construction that allows testing an entire region of genomic DNA without prior knowledge of potential function. Genomic origin of constructs can easily be assessed by end-sequencing, and even short stretches of end-sequence are usually sufficient to map constructs with high precision and confidence. Mapped constructs can then be arrayed for minimal construct coverage of a region of interest, thereby reducing the number of transfections compared to random testing. With automation of clone picking, DNA extraction, transfection, and reporter evaluation, it is conceivable that the reporter library strategy can be extended to whole-genome scale. Once the library is in place, the approach can easily be applied to different cell lines in order to identify transcription-activating DNA elements across the genome by functional criteria.

Isl1 *reporter construct library in pancreatic cell lines*

Interestingly, although *Isl1* is expressed in all three endocrine cell types, we did not recover constructs that were equally active in all pancreatic cell lines assayed here. This indicates that our assays provided a measure of cell-type specificity rather than detecting 'generic' promoters or enhancers that have universal or widespread activity. In fact, our transfection assays encompassed five constructs that carry the transcription start site of the *Isl1* gene in the context of various lengths of upstream and downstream DNA. Those fragments covered an area from -8.0kb to +6.2kb relative to the transcription start site (chr13:117430367, UCSC mm8), including the *Isl1* promoter as well as any promoter-proximal control elements in addition to the transcription start site. We find that none of these five constructs exhibited reporter activity in any of the cell lines tested. These results support our initial observation in transgenic mice that basal promoter activity is insufficient to direct cell type-specific transcription of the *Isl1* gene, yet seem to conflict with results by Cai *et al.* (Cai *et al.*, 2005) reporting modulation of *Isl1* promoter activity by Tbx20. However, those experiments were conducted in a heterologous context of HEK293 cells, a human embryonic kidney cell line. In contrast, our results were obtained in transgenic mice, and may indicate a stricter requirement for cooperation from other regulatory elements – such as the 3cF10 enhancer described here – to yield a specific expression pattern in the embryo.

A total of 14 constructs with activity in pancreatic cell lines contained DNA fragments from eight distinct regions of the *Isl1* locus; typically, a construct would be active in one or two cell types. When classified by highest activity, we found four DNA constructs that were active in alpha cells, five in beta cells, and five in delta cells. The differential activity of reporter constructs in some, but not all, cell lines indicates that the transfection assay discriminated cell-specific activities, and suggests that the *Isl1* locus contains several physically separate elements that regulate transcription of *Isl1* in different pancreatic cell lines. All eight regions carry sequences that are conserved between human and mouse: conserved putative transcription factor binding sites for Areb6, Foxj2, Gata, Lef1b, Mef2, Nkx6.2, Oct1, and Stat5a were common to all constructs with activity in beta cells; alpha cell-active constructs shared putative sites for Freac7 and Usf; and no common sites could be found for delta cell-active constructs. GATA factors (Xu and Murphy, 2000; Ketola *et al.*, 2004; Ritz-Laser *et al.*, 2005), Stat5a (Friedrichsen *et al.*, 2003), and Oct1 (Darville *et al.*, 2004) are all expressed in beta cells, and could therefore contribute to the beta cell line activity observed from the reporter constructs. These DNA regions are excellent candidates to test whether the predicted binding sites are functionally relevant for control of *Isl1* gene expression *in vivo*. In summary, our results reveal several candidate regions in the *Isl1* gene locus that confer reporter activity in transfected pancreatic cell lines, and therefore may be involved in the regulation of *Isl1* gene expression in pancreas.

Isl1 *enhancer activity* in vivo

With the focus on the *Isl1* expression pattern at mid-gestation, we initially tested several constructs from our dual reporter vector library in transgenic mouse embryos for reporter activity *in vivo*. Neither the construct containing the *Isl1* promoter and 3kb of upstream region, nor the upstream region between –3 kb and –10 kb showed any reporter activity in E10.5 mouse embryos. This result affirmed our earlier conclusion that *Isl1* promoter sequences were insufficient to generate a proper expression pattern *in vivo*. In contrast, the construct covering the 3'-end of the *Isl1* gene showed reporter activity, with excellent match to the expression pattern of the *Isl1* gene itself at E10.5. This construct essentially reproduced *Isl1* expression in the heart and outflow tract as well as in the posterior, visceral, and limb bud mesoderm. Since the activity of this construct arose in the absence of the *Isl1* gene promoter, and the reporter vector is based on a heterologous promoter, we suggest that the construct contains a bona fide *Isl1* gene enhancer

The activity of the enhancer was highly similar to the expression of the gene between E8.0 and E10.5, with initial expression

in heart mesoderm, followed by the cardiac crescent, and subsequent loss of expression in most cells of the heart tube itself. The enhancer therefore adheres to a pattern of Isl1 expression in cardiomyocyte precursors at early stages, followed by a down-regulation once the heart tube forms from the crescent (Prall *et al.*, 2007). Beginning at E8.0, the enhancer also recapitulates */s/1* gene expression in the foregut and hindgut diverticula. We found it notable that the enhancer generated reporter activity in the lateral plate mesoderm, which was in congruence with the Isl1 mRNA pattern. However, the enhancer generated a very strong signal along the entire length of the lateral plate. In contrast, */s/1* mRNA expression as well as */acZ* reporter activity from a */s/1 /acZ* knock-in allele (Prall *et al.*, 2007; Sun *et al.*, 2007) show a strong signal only in the posterior region, with weaker signals at the flanks of the E8.0 embryo. This suggested that the 3cF10 sequence might require additional regulatory elements for an exact reproduction of Isl1 expression in lateral plate and splanchnic mesoderm. The conclusion that the 3cF10 sequence represents a functional part but not the entire regulatory mechanism for */s/1* expression in mesendodermal derivatives is also supported by the finding that, while there is good agreement between enhancer activity between E8.0 and E10.5, the enhancer pattern degrades shortly after E10.5 and loses congruence to */s/1* mRNA expression. This indicates that maintenance of expression in dorsal aorta, heart, and visceral mesoderm requires additional control sequences that are separate from the 3cF10 sequence.

Notably, except for the olfactory pit, the 3cF10 enhancer does not contain elements for neuronal expression of */s/1*. Expression of */s/1* in neurons is mediated by several distinct enhancers (Uemura *et al.*, 2005; Bejerano *et al.*, 2006). All these regulatory elements are located outside the area covered by the two BAC clones used for this study. Our findings therefore complement those in the literature, as they characterize a novel enhancer for aspects of the */s/1* expression pattern that are not controlled by the neuronal enhancers. The complete lack of neuronal reporter activity of the 3cF10 enhancer in the nervous system demonstrates that the control of */s/1* gene expression involves distinct and independent regulatory modules that together generate the expression pattern of the gene.

Embryos lacking */s/1* die at mid-gestation, with defects at the dorsal aorta (Pfaff *et al.*, 1996), and failure to form the outflow tract (Cai *et al.*, 2003). As we demonstrated (Fig. 5B), the 3cF10 reporter labels the dorsal aorta, raising the possibility that the regulatory element(s) present on this construct ultimately play(s) a role in formation of the vessel. Currently, we cannot make a functional claim for this regulatory element in the dorsal aorta; this would require a targeted deletion of the region covered by 3cF10, without altering the */s/1* gene itself. However, it may not be necessary to delete the entire 3cF10 region: our bioinformatics analyses of the 3cF10 sequence revealed three deeply conserved regions. These regions contain evolutionarily conserved potential binding sites for Foxo/Foxd, E2f, Tcf1, Arnt, Cebp, Gata, Freac, and Tgif transcription factors, or members of these transcription factor families. We conclude from the co-transfection studies of the enhancer with Foxo1 and Gata4 transcription factors that several of the factors implicated by bioinformatics indeed functionally contribute to 3cF10 activity in cells. We suggest that the enhancer activity of the 3cF10 construct locates to these conserved regions that are found between 3 and 6kb downstream of the */s/1* gene, and that Foxo1 and Gata4 are likely candidates for regulation of */s/1* gene expression via the enhancer element located on the 3cF10 construct.

Isl1 *as target gene for Foxo1 and Gata4*

Both Foxo1 and Gata factors were previously shown to be involved in angiogenesis and vascular development: (i) mice missing Foxo1 die in mid-gestation as a consequence of incomplete vascular development (Hosaka *et al.*, 2004); and (ii) Gata4, Gata5, and Gata6 have been reported to be involved in various aspects of the formation of the heart and the great vessels (Kelley *et al.*, 1993; Laverriere *et al.*, 1994; Morrissey *et al.*, 1997). Therefore, Foxo1 and Gata factors are expressed at the right time and place to potentially be involved in */s/1* gene regulation. Our finding that Foxo1 and Gata4 individually activated this */s/1* enhancer in cellular transfection assays strongly supports the notion that */s/1* is, in fact, a target gene of Foxo1 and Gata4. While we cannot rule out that other forkhead or Gata transcription factors may play a role in the regulation of */s/1*, our initial findings would place */s/1* in a regulatory cascade downstream of Gata4 and Foxo1. Furthermore, the role of Gata4 for */s/1* expression indicates that Gata4 may have multiple control functions during heart development, namely at the level of */s/1*, as well as a the level of the downstream Isl1-target Mef2c (Dodou *et al.*, 2004), a crucial factor for heart development (Lin *et al.*, 1997). In other tissues, the expression pattern of Gata4 suggested that */s/1* is not only a Gata4-target in the developing heart, but also in the lateral plate mesoderm (Rojas *et al.*, 2005).

Our finding that Foxo1 and Gata4 together did not exhibit additive or synergistic effects on the 3cF10 enhancer activity suggests that */s/1* regulation may not involve the type of cooperation of forkhead and Gata transcription factors seen at other enhancers (Cirillo and Zaret, 1999; Denson *et al.*, 2000; Ritz-Laser *et al.*, 2005). Typically, forkhead protein anchoring at nucleosomes leads to subsequent binding of the Gata protein to DNA (Cirillo and Zaret, 1999; Cirillo *et al.*, 2002). It is likely that the 1kb distance between the Foxo and Gata sites on 3cF10 may prevent this particular cooperative mechanism. Our results rather indicate that Foxo1 and Gata4 may participate in */s/1* regulation through mutually exclusive pathways. In contrast to Foxo1 and Gata4, binding sites for the forkhead-related factors Freac2 or Freac1 (Mahlapuu *et al.*, 1998), and Gata4, are located only 70bp apart within one conserved region of the */s/1* enhancer, thereby allowing for a potential nucleosome-based cooperation between the forkhead protein and Gata4. Both forkhead-related transcription factors that are expressed in the visceral mesoderm (Aitola *et al.*, 2000) (Rojas *et al.*, 2005), and Gata4 is also expressed in the visceral mesoderm of the mouse embryo (Rojas *et al.*, 2005) in a pattern that matches */s/1* expression, as well as the pattern of the 3cF10 */s/1* reporter activity. These factors are therefore excellent candidates for transcription factors that may regulate 3cF10 enhancer function in the mesenchyme surrounding the gut. Further experiments will be required to elucidate the contribution of forkhead and Gata proteins for the control of */s/1* expression in derivatives of the lateral plate.

Materials and Methods

Reporter constructs

The direct reporter constructs were made by ligating DNA pieces as follows: The proximal upstream 3kb of */s/1* were cloned as a HindIII-Mfe I

fragment (-2873 to +224 relative to the murine *Isl1* transcription start) upstream of a *LacZ* reporter gene. The next 7kb located upstream was isolated as a Hind III fragment (-9549 to -2867) and cloned into a vector containing a minimal IE promoter and a *LacZ* reporter gene (Gardner *et al.*, 1996). The min-IE-*LacZ* reporter construct is inactive in transgenic mice in the absence of a functional enhancer and is subject to ectopic activation in random integration sites in less than 7% of cases (Yaworsky *et al.*, 1997). In the present study, ectopic activity was only observed for the 7kb fragment.

Dual-reporter vector for analyzing regulatory capacity of DNA

A new vector (PINZILIA), containing two reporter genes on a single transcript, was created (using as the plasmid backbone pBluescript KS+; Stratagene, La Jolla, CA) by combining the minimal promoter of the HSV ICP4 gene (Gardner *et al.*, 1996) with the *E.coli LacZ* gene modified by a nuclear localization signal. An internal ribosome entry site was placed downstream of the *LacZ* gene; the coding sequence of firefly *Luciferase*, an intron from the small antigen gene of SV40, and a polyadenylation signal from the bovine growth hormone gene make up the remainder of the reporter array. A single Bgl II site for cloning DNA of interest precedes the promoter, and the whole assembly is flanked by restriction sites for NotI as well as XhoI to allow excision of reporter construct assemblies for generation of transgenic mice.

Generation of a reporter construct library

The genomic locus of the murine *Isl1* gene was isolated in the form of two BAC clones (MGS1-437O11, 122 kb, and MGS1-460I18, 127 kb) by screening a mouse 129Sv/J BAC library (Genome Systems, St. Louis, MO) with two PCR assays that detected the first and second exon of the mouse *Isl1* gene, respectively. BAC DNA was prepared (Clontech, Mountain View, CA), partially digested with Sau3AI restriction enzyme, and DNA fragments in the range from 8 to 10kb were isolated using gel electrophoresis and subsequently purified (Qiagen, Valencia, CA). The PINZILIA dual reporter vector was linearized with Bgl II, treated with Alkaline Phosphatase, and BAC DNA fragments were ligated to the vector. After transformation, glycerol stocks, as well as DNA (Qiagen), were prepared from individual colonies. 104 constructs were generated from BAC MGS1-460I18, and 98 constructs were obtained from BAC MGS1-437O11; constructs were arrayed, propagated and maintained in 96-well format. Sequence tags were obtained from each individual construct for identification and mapping purposes and to eliminate constructs that contained *E. coli* genomic DNA. The sequence tag data, together with end-sequences obtained directly from the BAC clones, and sequence data for the two BAC clones obtained through the NIH Trans Genome Sequencing Initiative (see below) resulted in the annotation of the region as shown in Fig. 2. Sequence tags were used to orient and order clone inserts on the map.

Cell transfections

Library clones were transfected (n=3) into cells seeded in 96-well plates 24 hours prior to transfection. 100 ng of DNA were added to each well following the Effectene transfection protocol (Qiagen). Cells were used for Luciferase assays 72 hours after transfection. Cells were lysed using Glo-Lysis Buffer (Promega, St. Luis Obispo, CA), and Luciferase activity was determined employing the Steady-Glo Luciferase assay system (Promega) on a Fluoroskan Ascent FL luminometer. For standardization of transfection efficiency, cells were co-transfected with a CMV-EGFP construct, and EGFP fluorescence was recorded, prior to cell lysis, on the same instrument. Constructs were designated positive if their mean activity exceeded the mean activity of the entire library by at least two-fold. Transcription factor co-transfection assays (n=6) were carried out in BHK-21 cells. The following reporter plasmids were used: (i) Pinzilia (the dual reporter gene vector without an insert); (ii) IRS3TkLuc, a reporter gene with 3 synthetic Foxo1-responsive IRS sites as positive control for activation by Foxo1; (iii) SV40Luc, a luciferase reporter plasmid carrying enhancer and promoter of SV40 as general positive control; and (iv) 3cF10, the reporter construct identified from the *Isl1* gene locus. Cells were transfected using Effectene reagent (Qiagen) with 100ng of reporter plasmid DNA, and either 40 ng of vector DNA, or 20ng of vector DNA and 20ng of transcription factor expression plasmid, or 20ng of each of two expression plasmids. Expression plasmids for transcription factors were as follows: a wildtype version of Foxo1; a constitutively active form of Foxo1; and Gata4. The plasmids TkLuc, IRS3TkLuc, Foxo1, and the constitutively active form of Foxo1 (Guo *et al.*, 1999) were generous gifts from Dr. Terry Unterman (University of Illinois at Chicago), and we are grateful to Dr. Ted Simon (Washington University School of Medicine, St. Louis) for the Gata4 expression plasmid (Divine *et al.*, 2004). Luciferase activities for each assay were normalized to the average Luciferase activity of the corresponding reporter plasmid without exogenous transcription factors; to determine statistical significance, a two-tailed t-test assuming equal variances was performed.

Generation of transgenic mice

Constructs selected for *in vivo* analysis in transgenic mice were digested with NotI, and the assembly of genomic DNA/promoter/reporter genes was purified from bacterial vector sequences by gel electrophoresis and extraction from agarose. DNA concentration was adjusted to 3μg/ml in 5mM TrisCl, 0.1mM EDTA, pH 7.5 and used for pronuclear injection following published procedures (Hogan *et al.*, 1996); fertilized eggs were of FVB genetic background, and CD-1 females were used as foster mothers. The construct 3cF10 was chosen to generate lines; five independent founders were obtained and used for further breeding.

Isolation of mouse embryos

Normal mouse embryos were obtained from timed matings of FVB mice where the detection of a copulation plug in the morning was counted as 0.5 days of gestation. Embryos were isolated at different stages of gestation, embedded in O.C.T. compound (Sakura Finetek, Torrance, CA), frozen on dry ice and stored at –80°C. Sections of 30 μm thickness were produced on a Zeiss Microm cryostat and processed for *in situ* hybridization as described (Salbaum, 1998).

Reporter gene assays in transgenic mouse embryos

Reporter gene assays for β-galactosidase activity were performed as described previously (Yaworsky and Kappen, 1999). Noon of the day of embryo transfer (for direct analysis of founder embryos) or noon of the day of appearance of the vaginal plug (for matings of transgenic mice) was counted as gestational day 0.5. Embryos from lines were prepared at various gestational days. Genotyping was performed on genomic DNA from placenta, or – in the case of transgenic mothers – from the yolk sac, using a LacZ gene-specific PCR method (Yaworsky *et al.*, 1997). Embryos were fixed and processed for β-galactosidase staining as described before (Yaworsky *et al.*, 1997). Sections of paraffin-embedded stained embryos were cut at 10 μm thickness on a Leica microtome. Documentation of histological staining for β-galactosidase activity was done on a Leica M9.5 stereomicroscope using a Kodak MDS290 digital camera system as well as a Leica Z16 macroscope with a Leica DFC420 digital camera.

Isl1 *expression studies*

Expression of *Isl1* was assessed by *in situ* hybridization using previously established protocols (Salbaum, 1998; Yaworsky and Kappen, 1999; Salbaum and Kappen, 2000) and a DIG-labeled antisense RNA probe generated from the rat *Isl1* cDNA (Karlsson *et al.*, 1990).

Sequence analysis and annotation

Sequence data were generated (i) by our own laboratory in the form of sequence tags from the ends of the genomic insert of each construct (averaging 400bp in length), (ii) sequence tags from BAC ends, and (iii) additional sequence information was obtained from the Trans-NIH Mouse

Genome Initiative (Priority Sequencing Project performed at the Harvard Genome Sequencing Center providing 4x coverage). Sequence comparisons and contig assembly were performed using MacMolly software (SoftGene, Berlin, Germany). The final sequence had a total size of 202 kb with 12 small gaps. The structure of the *Isl1* gene locus in mouse genome information data (strain C57BL/6; NCBI m32 mouse assembly) remained in question (with half the *Isl1* gene missing); therefore, the *Isl1* genomic sequence from 129Sv/J mice obtained through this study was used for all further analyses.

Bioinformatics analyses

Sequence comparisons to other species were performed using the alignment function of the ECR browser (Ovcharenko *et al.*, 2004). The mouse 129Sv/J *Isl1* sequence was validated through colinearity with the mammalian counterparts human, rat, and dog. Evolutionary conserved sequences (either throughout the 202 kb locus or for reporter constructs only) were detected in comparisons to human, dog, rat, chicken, and *Xenopus* using the VISTA function of the ECR browser (Loots and Ovcharenko, 2004; Ovcharenko *et al.*, 2004) with a window size of 100 bp and a similarity level of 75%. Conserved regions were considered for further analysis only if conservation extended through all five vertebrate species. Potential transcription factor binding sites in these conserved sequences were predicted using rVISTA (Loots and Ovcharenko, 2004).

Author's contributions

JMS was responsible for the experimental design, the molecular and transfection experiments, histological evaluation, bioinformatics, and participated in drafting the manuscript. CK conceived of the study and was in charge of the transgenic mouse experiments, evaluation of data, and participated in drafting the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Y.L. Muller for cloning and sequencing of the upstream region and the first intron of the mouse Isl1 *gene and mapping of the transcription start site, and the Mayo Clinic Scottsdale Transgenic facility for microinjection services of transgene constructs. Kimberly Bynoté, Diane Costanzo, Dr. Gary Leisman, Matthew Northam and Tarra Wiggins as well as the UNMC Mouse Genome Engineering Facility provided technical assistance, and M. Anita Jennings contributed histology services. The assistance by Jacalyn MacGowan in finalizing this study at PBRC is gratefully acknowledged. We are also grateful for comments from Drs. Silvia Arber and Gary Schoenwolf, and to Saralyn Fisher for typing of the manuscript. This work was supported in parts by the Nebraska Research Initiative, R01-HD037804 (to C.K.), R21-DK59280 (to J.M.S.) and a Basil O'Connor Scholarship from the March of Dimes Birth Defects Foundation to J.M.S. High throughput sequencing of the murine* Isl1 *locus was performed by the Harvard Genome Center, courtesy of the Priority Sequencing Program of the Trans-NIH Mouse Genome Initiative.*

References

AHLGREN, U., PFAFF, S.L., JESSELL, T.M., EDLUND, T. and EDLUND, H. (1997). Independent requirement for isl1 in formation of pancreatic mesenchyme and islet cells. *Nature* 385: 257-260.

AITOLA, M., CARLSSON, P., MAHLAPUU, M., ENERBACK, S. and PELTO-HUIKKO, M. (2000). Forkhead transcription factor foxf2 is expressed in mesodermal tissues involved in epithelio-mesenchymal interactions. *Dev Dyn* 218: 136-149.

BEJERANO, G., LOWE, C.B., AHITUV, N., KING, B., SIEPEL, A., SALAMA, S.R., RUBIN, E.M., KENT, W.J. and HAUSSLER, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441: 87-90.

CAI, C.L., LIANG, X., SHI, Y., CHU, P.H., PFAFF, S.L., CHEN, J. and EVANS, S. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell* 5: 877-889.

CAI, C.L., ZHOU, W., YANG, L., BU, L., QYANG, Y., ZHANG, X., LI, X., ROSENFELD, M.G., CHEN, J. and EVANS, S. (2005). T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. *Development* 132: 2475-2487.

CIRILLO, L.A., LIN, F.R., CUESTA, I., FRIEDMAN, D., JARNIK, M. and ZARET, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors hnf3 (foxa) and gata-4. *Mol Cell* 9: 279-289.

CIRILLO, L.A. and ZARET, K.S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* 4: 961-969.

DARVILLE, M.I., TERRYN, S. and EIZIRIK, D.L. (2004). An octamer motif is required for activation of the inducible nitric oxide synthase promoter in pancreatic beta-cells. *Endocrinology* 145: 1130-1136.

DAVIDSON, E.H. (2006). *The regulatory genome.* Academic Press, San Diego.

DENSON, L.A., MCCLURE, M.H., BOGUE, C.W., KARPEN, S.J. and JACOBS, H.C. (2000). Hnf3beta and gata-4 transactivate the liver-enriched homeobox gene, hex. *Gene* 246: 311-320.

DILEONE, R.J., RUSSELL, L.B. and KINGSLEY, D.M. (1998). An extensive 3' regulatory region controls expression of bmp5 in specific anatomical structures of the mouse embryo. *Genetics* 148: 401-408.

DIVINE, J.K., STALOCH, L.J., HAVERI, H., JACOBSEN, C.M., WILSON, D.B., HEIKINHEIMO, M. and SIMON, T.C. (2004). Gata-4, gata-5, and gata-6 activate the rat liver fatty acid binding protein gene in concert with hnf-1alpha. *Am J Physiol Gastrointest Liver Physiol* 287: G1086-G1099.

DODOU, E., VERZI, M.P., ANDERSON, J.P., XU, S.M. and BLACK, B.L. (2004). Mef2c is a direct transcriptional target of isl1 and gata factors in the anterior heart field during mouse embryonic development. *Development* 131: 3931-3942.

DONG, J., ASA, S.L. and DRUCKER, D.J. (1991). Islet cell and extrapancreatic expression of the lim domain homeobox gene isl-1. *Mol Endocrinol* 5: 1633-1641.

ERICSON, J., THOR, S., EDLUND, T., JESSELL, T.M. and YAMADA, T. (1992). Early stages of motor neuron differentiation revealed by expression of homeobox gene islet-1. *Science* 256: 1555-1560.

FRAZER, K.A., PACHTER, L., POLIAKOV, A., RUBIN, E.M. and DUBCHAK, I. (2004). Vista: Computational tools for comparative genomics. *Nucleic Acids Res* 32: W273-W279.

FRIEDRICHSEN, B.N., RICHTER, H.E., HANSEN, J.A., RHODES, C.J., NIELSEN, J.H., BILLESTRUP, N. and MOLDRUP, A. (2003). Signal transducer and activator of transcription 5 activation is sufficient to drive transcriptional induction of cyclin d2 gene and proliferation of rat pancreatic beta-cells. *Mol Endocrinol* 17: 945-958.

GALLI-RESTA, L., RESTA, G., TAN, S.S. and REESE, B.E. (1997). Mosaics of islet-1-expressing amacrine cells assembled by short-range cellular interactions. *J Neurosci* 17: 7831-7838.

GARDNER, D.P., BYRNE, G.W., RUDDLE, F.H. and KAPPEN, C. (1996). Spatial and temporal regulation of a lacz reporter transgene in a binary transgenic mouse system. *Transgenic Res* 5: 37-48.

GONG, Z., HUI, C.C. and HEW, C.L. (1995). Presence of isl-1-related lim domain homeobox genes in teleost and their similar patterns of expression in brain and spinal cord. *J Biol Chem* 270: 3335-3345.

GOTO, M.P. and GOLDMAN, A.S. (1994). Diabetic embryopathy. *Curr Opin Pediatr* 6: 486-491.

GUO, S., RENA, G., CICHY, S., HE, X., COHEN, P. and UNTERMAN, T. (1999). Phosphorylation of serine 256 by protein kinase b disrupts transactivation by fkhr and mediates effects of insulin on insulin-like growth factor-binding protein-1 promoter activity through a conserved insulin response sequence. *J Biol Chem* 274: 17184-17192.

HIGASHIJIMA, S., Y., H. and OKAMOTO, H. (2000). Visualization of cranial motor neurons in live transgenic zebrafish expressing the green fluorescent protein under control of the isl-1 promoter/enhancer. *J. Neuroscience* 20: 206-218.

HOGAN, B.L., COSTANTINI, F. and LACY, E. (1996). *Manipulating the mouse embryo: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

HOSAKA, T., BIGGS III, W.H., TIEU, D., BOYER, A.D., VARKI, N.M., CAVENEE, W.K. and ARDEN, K.C. (2004). Disruption of forkhead transcription factor (foxo) family members in mice reveals their functional diversification. *Proc.Natl. Acad. Sci. USA* 101: 2975-2980.

KARLSSON, O., THOR, S., NORBERG, T., OHLSSON, H. and EDLUND, T. (1990). Insulin gene enhancer binding protein isl-1 is a member of a novel class of proteins containing both a homeo- and a cys-his domain. *Nature* 344: 879-882.

KELLEY, C., BLUMBERG, H., ZON, L.I. and EVANS, T. (1993). Gata-4 is a novel transcription factor expressed in the endocardium of the developing heart. *Development* 118: 817-827.

KETOLA, I., OTONKOSKI, T., PULKKINEN, M.A., NIEMI, H., PALGI, J., JACOBSEN, C.M., WILSON, D.B. and HEIKINHEIMO, M. (2004). Transcription factor gata-6 is expressed in the endocrine and gata-4 in the exocrine pancreas. *Mol Cell Endocrinol* 226: 51-57.

KORZH, V., EDLUND, T. and THOR, S. (1993). Zebrafish primary neurons initiate expression of the lim homeodomain protein isl-1 at the end of gastrulation. *Development* 118: 417-425.

LAVERRIERE, A.C., MACNEILL, C., MUELLER, C., POELMANN, R.E., BURCH, J.B.E. and EVANS, T. (1994). Gata-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *J. Biol. Chemistry* 269: 23177-23184.

LIN, Q., SCHWARZ, J., BUCANA, C. and OLSON, E.N. (1997). Control of mouse cardiac morphogenesis and myogenesis by transcription factor mef2c. *Science* 276: 1404-1407.

LOOTS, G.G. and OVCHARENKO, I. (2004). Rvista 2.0: Evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res* 32: W217-W221.

MAHLAPUU, M., PELTO-HUIKKO, M., AITOLA, M., ENERBACK, S. and CARLSSON, P. (1998). Freac-1 contains a cell-type-specific transcriptional activation domain and is expressed in epithelial-mesenchymal interfaces. *Dev. Biol.* 202: 183-195.

MITSIADIS, T.A., ANGELI, I., JAMES, C., LENDAHL, U. and SHARPE, P.T. (2003). Role of islet1 in the patterning of murine dentition. *Development* 130: 4451-4460.

MORRISSEY, E.M., IP, H.S., TANG, Z., LU, M.M. and PARMACEK, M. (1997). Gata-5: A transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Dev. Biol.* 183: 21-36.

MULLER, Y.L., YUEH, Y.G., YAWORSKY, P.J., SALBAUM, J.M. and KAPPEN, C. (2003). Caudal dysgenesis in islet-1 transgenic mice. *FASEB J* 17: 1349-1351.

OVCHARENKO, I., NOBREGA, M.A., LOOTS, G.G. and STUBBS, L. (2004). Ecr browser: A tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res* 32: W280-W286.

PENNACCHIO, L.A., AHITUV, N., MOSES, A.M., PRABHAKAR, S., NOBREGA, M.A., SHOUKRY, M., MINOVITSKY, S., DUBCHAK, I., HOLT, A., LEWIS, K.D. *et al.* (2006). *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* 444: 499-502.

PENNACCHIO, L.A. and RUBIN, E.M. (2001). Genomic strategies to identify mammalian regulatory sequences. *Nat Rev Genet* 2: 100-109.

PFAFF, S.L., MENDELSOHN, M., STEWART, C.L., EDLUND, T. and JESSELL, T.M. (1996). Requirement for lim homeobox gene isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* 84: 309-320.

PRALL, O.W., MENON, M.K., SOLLOWAY, M.J., WATANABE, Y., ZAFFRAN, S., BAJOLLE, F., BIBEN, C., MCBRIDE, J.J., ROBERTSON, B.R., CHAULET, H. *et al.* (2007). An nkx2-5/bmp2/smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* 128: 947-959.

RITZ-LASER, B., MAMIN, A., BRUN, T., AVRIL, I., SCHWITZGEBEL, V.M. and PHILIPPE, J. (2005). The zinc finger-containing transcription factor gata-4 is expressed in the developing endocrine pancreas and activates glucagon gene expression. *Mol Endocrinol* 19: 759-770.

ROJAS, A., DE VAL, S., HEIDT, A.B., XU, S.M., BRISTOW, J. and BLACK, B.L. (2005). Gata4 expression in lateral mesoderm is downstream of bmp4 and is activated directly by forkhead and gata transcription factors through a distal enhancer element. *Development* 132: 3405-3417.

SALBAUM, J.M. (1998). Punc, a novel mouse gene of the immunoglobulin superfamily, is expressed predominantly in the developing nervous system. *Mech Dev* 71: 201-204.

SALBAUM, J.M. and KAPPEN, C. (2000). Cloning and expression of nope, a new mouse gene of the immunoglobulin superfamily related to guidance receptors. *Genomics* 64: 15-23.

SCHUG, J., SCHULLER, W.P., KAPPEN, C., SALBAUM, J.M., BUCAN, M. and STOECKERT, C.J., JR. (2005). Promoter features related to tissue specificity as measured by shannon entropy. *Genome Biol* 6: R33.

SUN, Y., LIANG, X., NAJAFI, N., CASS, M., LIN, L., CAI, C.L., CHEN, J. and EVANS, S.M. (2007). Islet 1 is expressed in distinct cardiovascular lineages, including pacemaker and coronary vascular cells. *Dev Biol* 304: 286-296.

THOR, S., ERICSON, J., BRANNSTROM, T. and EDLUND, T. (1991). The homeodomain lim protein isl-1 is expressed in subsets of neurons and endocrine cells in the adult rat. *Neuron* 7: 881-889.

UEMURA, O., OKADA, Y., ANDO, H., GUEDJ, M., HIGASHIJIMA, S., SHIMAZAKI, T., CHINO, N., OKANO, H. and OKAMOTO, H. (2005). Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression. *Dev Biol* 278: 587-606.

XU, W. and MURPHY, L.J. (2000). Cloning of the mouse pax4 gene promoter and identification of a pancreatic beta cell specific enhancer. *Mol Cell Endocrinol* 170: 79-89.

YAWORSKY, P.J., GARDNER, D.P. and KAPPEN, C. (1997). Transgenic analyses reveal developmentally regulated neuron- and muscle-specific elements in the murine neurofilament light chain gene promoter. *J Biol Chem* 272: 25112-25120.

YAWORSKY, P.J. and KAPPEN, C. (1999). Heterogeneity of neural progenitor cells revealed by enhancers in the nestin gene. *Dev Biol* 205: 309-321.

YUAN, S. and SCHOENWOLF, G.C. (2000). Islet-1 marks the early heart rudiments and is asymmetrically expressed during early rotation of the foregut in the chick embryo. *Anat Rec* 260: 204-207.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

See our Special Issue ***Epigenetics & Development***, edited by Saadi Khochbin and Stefan Nonchev at:
http://www.ijdb.ehu.es/web/contents.php?vol=53&issue=2-3

See our Special Issue ***Fertilization***, in honor of David L. Garbers and edited by Paul M. Wassarman and Victor D. Vacquier at:
http://www.ijdb.ehu.es/web/contents.php?vol=52&issue=5-6

Understanding the regulatory genome
M. Eva Alonso, Bárbara Pernaute, Miguel Crespo, José Luis Gómez-Skarmeta and Miguel Manzanares
Int. J. Dev. Biol. (2009) doi: 10.1387/ijdb.072428ma

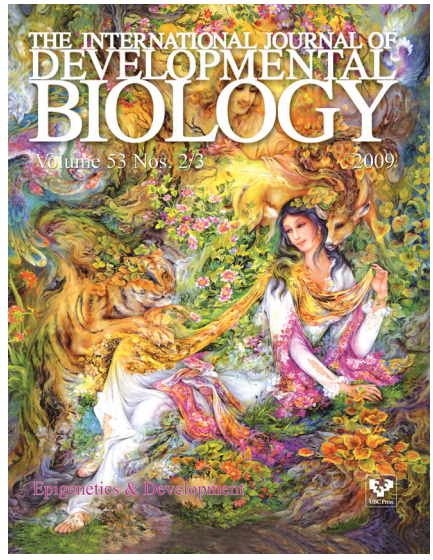
Genes controlling pancreas ontogeny.
Claire Bonal and Pedro L. Herrera
Int. J. Dev. Biol. (2008) 52: 823-835

Efficient Cre-mediated deletion in cardiac progenitor cells conferred by a 3'UTR-ires-Cre allele of the homeobox gene Nkx2-5.
Edouard G Stanley, Christine Biben, Andrew Elefanty, Louise Barnett, Frank Koentgen, Lorraine Robb and Richard P Harvey
Int. J. Dev. Biol. (2002) 46: 431-439

cdx4/lacZ and cdx2/lacZ protein gradients formed by decay during gastrulation in the mouse.
Stephen J. Gaunt, Deborah Drage and Richard C. Trubshaw
Int. J. Dev. Biol. (2005) 49: 901-908

Building the vertebrate heart - an evolutionary approach to cardiac development.
José M. Pérez-Pomares, Juan M. González-Rosa and Ramón Muñoz-Chápuli
Int. J. Dev. Biol. (2009) 52: 2409-2409

Analysis of transcriptional regulatory regions in vivo.
R J MacDonald and G H Swift
Int. J. Dev. Biol. (1998) 42: 983-994



5 yr ISI Impact Factor (2008) = 3.271