Supporting Information

Ragvin et al. 10.1073/pnas.0911591107

S1 Materials and Methods

HCNE Selection for Experimental Testing. We identified human-zebrafish nonexonic (i.e., noncoding and non-UTR) evolutionary conserved sequences using the HCNE identification procedure described below, complemented with the Vertebrate Multiz Alignment and PhastCons Conservation and Chained Alignments tracks in the UCSC Genome Browser (1) for the human March 2006 Assembly. Sequences were extracted from the UCSC Genome Browser for the human March 2006 assembly (hg18) or the X. tropicalis Aug. 2005 assembly (xenTro2).

Data Sources and LD Block Definition. We used sequence information from the following genome assemblies: Human March 2006 (hg18; International Human Genome Sequencing Consortium), Mouse February 2006 (mm8; National Center for Biotechnology Information and Mouse Genome Sequencing Consortium), Chicken May 2006 (galGal3; Genome Sequencing Center at Washington University), Zebrafish July 2007 (danRer5, zv7) and Zebrafish March 2006 (danRer4, zv6; WT Sanger Institute, Max Planck Institute for Developmental Biology and Netherlands Institute for Developmental Biology), Medaka April 2006 (oryLat1; National Institute of Genetics and the University of Tokyo), Tetraodon February 2004 (tetNig1; Genoscope and The Broad Institute), and Stickleback February 2006 (gasAca1; The Broad Institute). Alignments and annotations were obtained from the UCSC Genome Browser database (2) (http://genome.ucsc.edu). We defined LD blocks as the regions between flanking Hapmap recombination hotspots. Recombination hotspots were obtained from Hapmap release 21 (http://www.hapmap.org) and converted to hg18 coordinates using the UCSC liftOver tool.

Generation of Transgenic Zebrafish. Candidate HCNEs were amplified by PCR on human and Xenopus genomic DNA using Advantage 2 PCR Enzyme System (Clontech). The final enhancer test vector contained an HCNE in front of the zebrafish gata2 promoter coupled to the EGFP gene and a polyA signal, all flanked by Tol2 transposition sequences (3).

Preparation of Tol2 Transposase mRNA. Tol2 transposase mRNA was synthesized as described (4) and purified using MEGAclear Purification Kit (Ambion). Purified mRNA was eluted in nuclelease-free water and stored at −80°C.

Microinjections and Screening of Transgenic Lines. Enhancer test vectors were mixed with Tol2 transposase mRNA to a final concentration of 25 ng/μl each and co-injected into one-cell stage WT TAB zebrafish embryos. Injected embryos were screened under a fluorescent microscope (TE2000-S inverted microscope (Nikon) equipped with x10 and x20 lenses, a 500–20-mm excitation filter and a 515-nm BP emission filter (Chroma)] at 1 and 2 dpf for EGFP expression and positive candidates raised to 2 dpf for EGFP expression and positive candidates were identified using the cloned HCNE using the Gateway LR Clonase II Enzyme Mix (Invitrogen). We determined extent of synteny by joined BLASTZ net alignments (6) and investigated conservation of gene order among human, chicken, Xenopus, fugu, Tetraodon, medaka, stickleback, and zebrafish using BLASTZ and tBLASTn alignments obtained from the UCSC Browser (1). We identified highly conserved elements by scanning pairwise BLASTZ net whole-genome alignments (7) between human and each of the other genomes for regions with at least N identities over 50 alignment columns, using the identity thresholds (N/50) stated in Fig. 1 in the main text. For each pairwise comparison, we scanned two sets of net alignments (one from the perspective of each genome) to not miss elements with one too many orthologous relationships. Overlapping highly conserved elements were merged into one. We discarded elements whose genome coordinates overlapped by one or more bp with known repeats or any exon from the following gene annotations: human and mouse RefSeq Genes, Ensembl Genes, GenScan Genes, and UCSC Known Genes; chicken and frog RefSeq Genes, Ensembl Genes, and GenScan Genes; zebrafish RefSeq Genes, Ensembl Genes, zebrafish mRNAs, and human UCSC Known Genes aligned to the zebrafish assembly; and stickleback Ensembl Genes and stickleback mRNAs. To remove remaining repetitive sequences, we realigned each element with the two respective genomes using BLAT (8), counted all mapping positions with a sequence identity of at least 90%, and discarded any element with more than four mapping locations in a mammalian genome or eight mapping locations in a teleost genome and considered remaining elements HCNEs.

Cloning of HCNEs. The following primers were used: element 1 Xenopus (3,179 bp): forward primer F1 5′ cat ggt ggt acg ctg cta tg, reverse primer R6 5′ tgt gct gca ccc att aac tc; element 2 (999 bp): forward primer F2 5′ agt ttt tig taa aag agg ttgg, reverse primer R2 5′ gtt att aac caa tac tca tca aat aat gtt; element 3 (1,192 bp): forward primer F3 5′ gga gcc ctc caa atg tgc tta t, reverse primer R3 5′ cca ctc acc ttg gag tca aca a; element 4 (401 bp): forward primer F4 5′ gtt tgc agg gtt ctt cat ctct, reverse primer R4 5′ aag tgc ctt ctt ttg tgc aat gtt t; element 5 (600 bp): forward primer F5 5′ ttc acc ttt gcc ctc ctt cgg tgc tga t, reverse primer R5 5′ tgc gct gca ccc aat taa ctc; element 6 (543 bp): forward primer F6 5′ cat ggt gtt acg ctt cta gtt, reverse primer R6 5′ ctc cca aag tcc tgg caa a. To increase cloning efficiency, PCR products were subcloned into the standard Gateway entry vector pCR8GW/TOPO (Invitrogen) before insertion into the destination enhancer test vector (derived from the Tol2 T2KHG plasmid and containing the Gateway C1 cassette) by LR recombination where the Gateway C1 cassette of the destination vector was replaced by the cloned HCNE using the Gateway LR Clonase II Enzyme Mix (Invitrogen).

Histology. Cryosections (20–40 μm) of formaldehyde-fixed GFP-expressing embryos were stained with anti-GFP (A11122; Molecular Probes). Anti-rabbit Alexa488 (Molecular Probes) was used to detect the primary anti-GFP antibody. Sections were counterstained with rhodamine–phalloidin and DAPI to visualize the cells’ contours and nuclei, respectively, and imaged on a Zeiss SPE confocal system. Images were processed with Adobe Photoshop CS2.

Generation of Transgenic Mice. A genomic fragment from the HCNE containing rs1118875 from human HHEX was amplified from human BACs obtained from the BACPAC Resource Center (http://bacpac.chori.org) with the Expand Long Template PCR System (Roche), using the following oligonucleotides: HHEX.1 forward, 5′ gca tgg ttc ctc aag gac tct gaa, HHEX.1 reverse, 5′ gca tgc ata aat ttc ctt gtc. PCR fragments were

Ragvin et al. www.pnas.org/cgi/content/short/0911591107 1 of 4
checked by sequencing and cloned into a vector containing the human minimal β-globin promoter, lacZ, and an SV40 polyadenylation signal. Vectors were linearized, the vector backbone removed, and the construct microinjected into one cell mouse embryos according to standard techniques (9). F0 embryos of the desired stage were dissected and stained for lacZ activity.

**Morpholinos.** To knock down *in3a* expression, we used the following morpholinos: MO1-*in3a* (AGCTGTTGGAAAGACATTGTGTG) and MO2-*in3a* (GTGCTTCTTAAAAACACAGAAACATTGTGTG), targeting the ATG codon and the second intron–exon boundary, respectively. Solutions were prepared and microinjected into the yolk of one-cell stage embryos according as previously described (10). The morpholino for Nkx2.2 (MO2-Nkx-5UTR, 5′-TGGAGCATTTGATGCAGTCAAGTTG) was the one reported by Pauls et al. (11). For the controls we used an unrelated/unspecific morpholino (GTtAATACcAGtATAgATTgATAGTTG).

**In Situ Hybridization.** Double whole-mount in situ hybridizations were performed as previously described (12). Ghrelin- and insulin-labeled antisense probes were prepared and RNA in situ hybridizations were conducted as previously described (11). The following fluorescein-labeled antisense riboprobes were used: *insulin*, *somatostatin* (PPS2), and *glucagon* (13). The digoxigenin labeled antisense riboprobes were *ghrelin* (11) and *irx3a* (provided by Zhiyuang Gong, Singapore).

**RT-PCR.** Extraction of total RNA from embryos at the larval stage (48 hpf) was carried out as previously described (14). All RNAs were treated by RNase-free DNase I (Roche) and purified by RNAeasy columns. Approximately 1.5 μg of RNA was used as template for reverse transcription using primers 3AP(T)17 and (T)16, and M-MLV reverse transcriptase (Promega) in a 20-μL reaction mixture. For the RT-PCR, one tenth of the cDNA synthesis product was amplified with the following oligos: in3exon1for (5′-CCCA-CACGTGGGATATCAGT), in3exon1rev (5′-TGGACAGATCATTGTGAAATAAGGAA-3′), actinfor (5′-TGGAGCATTTGATGCAGTCAAGTTG), and actintrev (5′-TTCTCC TTAGTGTCAACCCAGGAC-3′).


**Fig. S1.** In situ hybridization on day 2 zebrafish embryos. (A) fto sense control and (B) fto antisense probe: fto is not significantly expressed at this stage of embryogenesis. (C) irx3a; arrowhead points to endodermal expression. (D) cdkal1 is expressed ubiquitously at a low level.

**Fig. S2.** *irx3a* expression in the pancreatic islet during late somitogenesis (20-somite stage). *irx3a* digoxigenin labeled riboprobes colocalize with insulin-expressing cells marked by fluorescently tagged antisense riboprobes (outlined by the rectangle). Lateral view, with anterior to the left.
Fig. S4. irx3a and nkx2.2a knockdown severely perturb hlxb9 expression in pancreatic endocrine cells. Representative confocal projections of 30 hpf hlxb9:GFP larvae injected with control morpholino (A–C), irx3a morpholino (D and E), and irx3a/nkx2.2a morpholinos (F, G, and H). (B) irx3a and nkx2.2a knockdown increases hb9 expression in spinal cord motor neurons (MNs). Representative confocal projections of 30 hpf hb9:GFP larvae injected with control morpholino (A), irx3a morpholino (B), and irx3a/nkx2.2a morpholinos (C).

Fig. S3. Exon 2 skipping shown by the RT-PCR amplification of a 266-bp amplicon and absence of the 1,296 band (lane 8). Lanes 1 and 10, molecular weight marker; lanes 2 and 6, negative control (mock); lanes 3–5, exon 1 amplification (positive control); lanes 7–9, exon 1–3 amplifications; lane 7, 0.5 g/L MO2-irx3a–injected embryos; lane 8, 1 g/L MO2-irx3a–injected embryos; lane 9, uninjected embryos.
Table S1. Fluorescence in the pancreatic islet and fluorescent MNs in the spinal cord

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Decreased fluorescence in pancreas</th>
<th>Increased fluorescence in spinal cord</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>80</td>
<td>1/80 (1.2)</td>
<td>0/80 (0)</td>
</tr>
<tr>
<td>MO2-irx3a</td>
<td>59</td>
<td>24/59 (40.7)</td>
<td>30/59 (50.8)</td>
</tr>
<tr>
<td>MO2-irx3a+MOnk-5UTR</td>
<td>26</td>
<td>16/26 (61.5)</td>
<td>17/26 (65.4)</td>
</tr>
</tbody>
</table>

Twenty-four of 59 and 30 of 59 irx3a-morphants displayed a strong decrease in the fluorescence at the pancreatic level and an increase of fluorescent MNs in the spinal cord, respectively. In double irx3a/nkx2.2a morphants a more pronounced decrease (16 of 26 larvae) in the hlxb9-positive endocrine cells and increase (17 of 26 larvae) in fluorescent MNs is seen. Absolute numbers and percentages are given for each single morpholino or combination of morpholinos. For the decrease of the hlxb9 fluorescent signal in the pancreatic compartment, morphant embryos are scored as “with a decreased fluorescence” when the fluorescent signal was unambiguously reduced (>25%) compared with controls.