Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to *HHEX*, *SOX4*, and *IRX3*

Anja Ragvin[a,b,1], Enrico Moro[c,1], David Fredman[d,e,2], Pavla Navratilova[e], Øyvind Drivenes[e], Pär G. Engström[d,e,3], M. Eva Alonso[f], Elisa de la Calle Mustienes[g], José Luis Gómez Skarmeta[g], Maria J. Tavares[g], Fernando Casares[g], Miguel Manzanares[f], Veronica van Heyningen[h], Anders Molven[i,j], Pål R. Njølstad[a,b], Francesco Argenton[c], Boris Lenhard[d,e,4], and Thomas S. Becker[e,2,5]

[a]Department of Clinical Medicine, University of Bergen, N-5020 Bergen, Norway; [b]Pediatrics, Haukeland University Hospital, N-5021 Bergen, Norway; [c]Dipartimento di Biologia, Universita' degli Studi di Padova, I-35121 Padua, Italy; [d]Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, N-5008 Bergen, Norway; [e]Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway; [f]Fundación Centro Nacional de Investigaciones Cardiovasculares Carlos III, E-28029 Madrid, Spain; [g]Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide, Consejo Superior de Investigaciones Científicas, Sevilla 41013, Spain; [h]Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; [i]The Gade Institute, University of Bergen, N-5020, Bergen, Norway; and [j]Department of Pathology, Haukeland University Hospital, N-5021 Bergen, Norway

Communicated by Michael S. Levine, University of California, Berkeley, CA, November 10, 2009 (received for review January 15, 2009)

Genome-wide association studies identified noncoding SNPs associated with type 2 diabetes and obesity in linkage disequilibrium (LD) blocks encompassing *HHEX-IDE* and introns of *CDKAL1* and *FTO* [Sladek R, et al. (2007) *Nature* 445:881–885; Steinthorsdottir V, et al. (2007) *Nat. Genet* 39:770–775; Frayling TM, et al. (2007) *Science* 316:889–894]. We show that these LD blocks contain highly conserved noncoding elements and overlap with the genomic regulatory blocks of the transcription factor genes *HHEX, SOX4*, and *IRX3*. We report that human highly conserved noncoding elements in LD with the risk SNPs drive expression in endoderm or pancreas in transgenic mice and zebrafish. Both *HHEX* and *SOX4* have recently been implicated in pancreas development and the regulation of insulin secretion, but *IRX3* had no prior association with pancreatic function or development. Knockdown of its orthologue in zebrafish, *irx3a,* increased the number of pancreatic ghrelin-producing epsilon cells and decreased the number of insulin-producing β-cells and glucagon-producing α-cells, thereby suggesting a direct link of pancreatic *IRX3* function to both obesity and type 2 diabetes.

hb9 | hlxb9 | nkx2.2 | pancreatic islet development | pancreatic peptides

Genome-wide association studies of SNPs are a powerful tool in the genetic analysis of common diseases. Several studies have shown significant associations of individual SNPs and haplotypes with type 2 diabetes (T2D) (1–5) and obesity (6, 7). With few exceptions, the identified SNPs fall into noncoding regions, either intronic or intergenic, suggesting a regulatory effect of the disease-associated haplotypes. Regulatory elements controlling developmental expression of transcription factors often coincide with highly conserved noncoding elements (HCNEs) (8–10).

We have recently shown that vertebrate genomes contain functional segments termed genomic regulatory blocks (GRBs), which are chromosomal domains that control the expression of developmental regulator genes, here termed target genes (11). GRBs often include unrelated genes termed bystander genes and can frequently be recognized through the conservation of synteny among all sequenced vertebrate genomes. Conserved synteny of bystander and target genes is maintained by the presence of enhancers within introns of the bystander genes that are specifically required for the regulation of the target gene (11, 12). A consequence of this type of arrangement is that intronic variants in one gene can affect a target gene further away, potentially leading to erroneous disease–gene association. To assess whether GRBs play a role in human genetic disease, and whether GRB analysis can help to link genomic regions that contain disease-causing variants to target genes, we used computational analysis on three regions, namely the *HHEX-IDE*, *CDKAL1*, and *FTO* LD blocks, which harbor SNPs recently demonstrated by genome-wide association studies to be significantly associated with risk for T2D and obesity (Table 1). We demonstrate that comparative genomic analysis of risk regions can quickly confirm or disqualify a candidate disease gene by showing that it is *HHEX*, *SOX4*, and *IRX3* that likely underlie T2D and obesity in the three loci analyzed in this study. We show that expression patterns driven by HCNEs within the risk regions can rapidly be assessed through transgenesis in the zebrafish or in the mouse to identify the organ system or cell type affected by a common disease. We further demonstrate by knockdown of the zebrafish orthologue of *IRX3*, *irx3a*, that this transcription factor is involved in a conserved pathway that specifies pancreatic islet cell identities.

Results

Risk SNPs Identify Regions Regulating *HHEX* and *SOX4*. For the *HHEX-IDE* region, the associated SNPs lie in a 295-kb block of LD that includes three genes, *HHEX, KIF1* and *IDE,* encoding a transcriptional regulator involved in pancreatic development (13), a kinesin interacting factor, and an insulin-degrading enzyme, respectively (14). The risk allele has been associated with decreased pancreatic β-cell function (15). Throughout vertebrates, *HHEX* is in conserved synteny with the neighboring *EXOC6* gene. The *HHEX* conserved synteny block overlaps with a small part of the risk allele–containing LD block. This part of the LD block contains the SNPs with the highest association scores, as well as the *HHEX* gene (Fig. 1*A* and Table 1). In contrast, *KIF1* and *IDE* are located outside the conserved synteny block, strongly suggesting

Author contributions: A.R., E.M., D.F., P.N., Ø.D., P.G.E., M.M., A.M., P.R.N., F.A., B.L., and T.S.B. designed research; A.R., E.M., D.F., P.N., Ø.D., P.G.E., M.E.A., E.d.l.C.-M., J.-L.G.-S., M. J.T., F.C., B.L., and T.S.B. performed research; A.R., E.M., D.F., P.R.N., Ø.D., P.E., M.E.A., J.-L. G.-S., F.C., M.M., V.v.H., F.A., B.L., and T.S.B. analyzed data; and D.F., F.A., B.L., and T.S.B. wrote the paper.

The authors declare no conflict of interest.

[1]A.R. and E.M. contributed equally to this work.

[2]Present address: Department of Molecular Evolution and Development, University of Vienna, 1090 Wien, Austria.

[3]Present address: European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, CB10 1SD, UK.

[5]Present address: The Brain & Mind Research Institute, Sydney Medical School, University of Sydney, Camperdown, NSW 2050, Australia.

[4]To whom correspondence may be addressed. E-mail: boris.lenhard@bccs.uib.no or tsbecker@med.usyd.edu.au.

This article contains supporting information online at www.pnas.org/cgi/content/full/0911591107/DCSupplemental.

Table 1. T2D/obesity risk SNPs fall into GRBs

SNP	GRB/target gene	Bystander gene	Reference
rs1111875* rs7923837	*HHEX*†	–	1–4
rs7754840 rs7756992 rs10946398	*SOX4*	*CDKAL1*†	2, 3, 5
rs8050136* rs9939609* rs1421085* rs17817449*	*IRX3*	*FTO*†	6, 7

*SNPs located within HCNEs.
†Gene reported as disease-associated.

that any *cis*-regulatory elements within this part of the LD block regulate *HHEX*, and not *KIF1* or *IDE*. The second risk region is a 200-kb LD block that includes the proximal promoter and exons and introns 1 to 5 of the gene *CDKAL1*. The SNP rs10946398 in this block is associated with decreased insulin secretion (15) but, in contrast to *HHEX*, *CDKAL1* had no prior connection to either pancreatic development or β-cell function. Comparative genomic analysis revealed *CDKAL1* to lie within a mammal:zebrafish synteny block containing *CDKAL1* and *SOX4* in tetrapods and *cdkal1* and *sox4b* in zebrafish (Fig. 1*B*). Of these genes, *SOX4*/*sox4b* has been linked to both pancreatic development (16, 17) and insulin secretion in the mouse (18), and as such represents a plausible candidate for T2D, whereas *cdkal1* in zebrafish is expressed ubiquitously and at low levels (Fig. S1). The other teleost fish with sequenced genomes (e.g., stickleback, medaka, fugu) retained only one copy of *sox4* after whole-genome duplication, and the conserved synteny is larger, including the loci of *E2F3* and *MBOAT1*. *E2F3*, a cell cycle regulator, is not expressed in pancreas (19), and its fate after whole-genome duplication in zebrafish indicates that it is not a GRB target gene. The third risk region, an LD block of 49 kb encompassing a large part of the first intron of *FTO* (fat mass- and obesity-associated gene), also contains numerous HCNEs and lies within an extended region of conserved synteny with the *IRX3*/*5*/*6* cluster, the closest gene of which is the transcription factor–encoding *IRX3* (Fig. 1). Although our computational analysis and GRB visualization suggested that the target gene of the HCNEs within *FTO* is *IRX3*, no connection of *IRX3* to pancreatic development or insulin secretion was apparent in the literature. To evaluate both genes as candidates, we performed in situ hybridization for both *fto* and *irx3a* in zebrafish larvae, revealing barely detectable activity for *fto* (Fig. S1) and expression in endoderm, colocalized with insulin transcripts, for *irx3a* (Fig. S2).

To provide further functional evidence for our hypothesis that the aforementioned T2D-associated SNPs are in genomic regions devoted to long-range regulation of *HHEX*, *SOX4*, and *IRX3* (the GRB targets), we examined the general landscape in the LD blocks containing reported SNPs and tested HCNEs within these LD blocks for the ability to drive expression consistent with that of the GRB target gene. Enhancer activity of human HCNEs from healthy, lean subjects was tested experimentally using a GFP reporter assay in zebrafish, whose validity in one case was confirmed by mouse transgenic reporter assays. A summary of tested elements is given in Table 2.

Reporter Assays in Mouse and Zebrafish Support Long-Range Regulation of *HHEX* and *SOX4*. Patients carrying the risk-associated C-allele of rs1111875 in the LD block overlapping with the *HHEX* GRB show lower acute insulin response (15). This highly associated SNP maps to a conserved region that, when tested in enhancer reporter assays in mouse or zebrafish, shows pancreatic islet expression (tested element 1; Fig. 2 *A* – *C*). Although this does not prove that rs1111875 is the variant conferring disease risk, our finding that the normal version of the HCNE containing rs1111875 directs reporter expression to the islet is highly suggestive of an effect on β-cell function.

Even though many functional enhancers are not conserved beyond mammalian genomes (20), the enhancer-containing regions are often recognizable through conserved gene order in all vertebrate genomes. We found *CDKAL1* to be within the GRB of *SOX4* and in conserved synteny with it across vertebrate genomes. The risk-associated SNPs, rs7754840 and rs7756992 (2, 3, 5), fall into its large fifth intron, and the corresponding LD block contains noncoding elements conserved from human to frog genomes. The density pattern of HCNEs and the conservation of synteny between *CDKAL1*/*SOX4* and zebrafish *cdkal1*/*sox4b* suggest that noncoding elements in this region likely regulate *SOX4* transcription. When tested in zebrafish, the enhancer activity of the HCNE (element 2) closest to rs7754840 showed a complex hindbrain expression pattern,

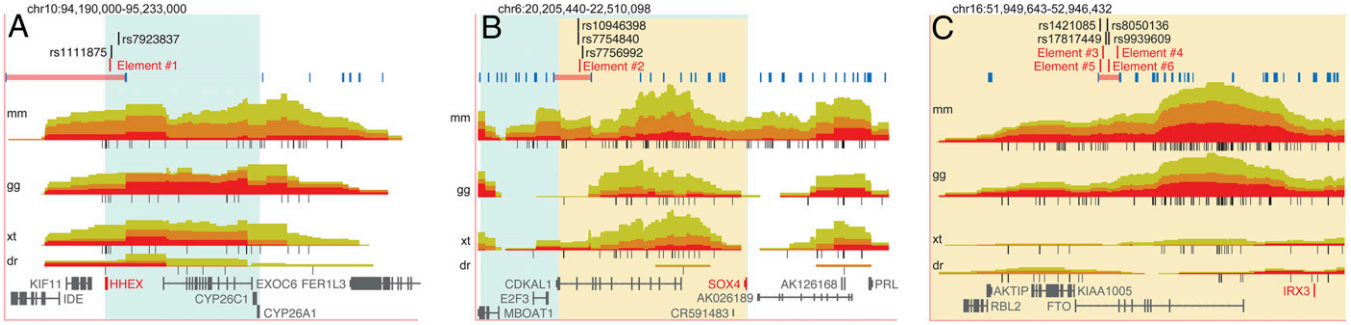

Fig. 1. Disease-associated SNPs (*Top*) and tested elements (elements 1–6) in the context of vertebrate GRBs around *HHEX* (*A*), *SOX4* (*B*), and *IRX3* (*C*). GRBs were defined by minimal synteny blocks to zebrafish (tan shading) or stickleback (light blue shading). Hapmap recombination hotspots (blue bars) mark the edges of LD blocks (blocks of interest shaded red). HCNE density plots for pairwise comparisons are colored yellow, orange, and red, based on the minimum percentage identity of the elements ≥50 bp in size used to derive them: human vs. mouse (*mm*: ≥95%, ≥98%, 100% identity), chicken (*gg*: ≥90, ≥95, ≥98), *Xenopus* (*xt*: ≥70, ≥80, ≥90) and zebrafish (*dr*: ≥70, ≥80, ≥90). HCNE densities were calculated as number of bases in HCNEs in sliding windows of 300 kb, and curves for different species are drawn to the same scale. HCNEs (black bars) are shown at the lowest threshold for each pairwise comparison. All features, including the suggested regulatory target gene (red) and bystander genes (gray) are shown in human coordinates. Gene annotations were adapted from the UCSC Known Genes set.

Table 2. Summary of tested elements

Tested HCNE-containing element	Human coordinates (hg18)	Relation to SNP position	Relation to closest gene	GRB target gene
1*	chr10:94452612–94453112	Contains rs1111875	10 kb downstream of *HHEX*	*HHEX*
2	chr6:20769594–20770392	Closest HCNE to rs7754840 in the corresponding LD block	In fourth intron of *CDKAL1*	*SOX4*
3	chr16:52364635–52365826	Contains rs1477196	In first intron of *FTO*	*IRX3*
4	chr16:52397232–52397632	Most conserved HCNE in the corresponding LD block	In first intron of *FTO*	*IRX3*
5	chr16:52358243–52358842	Contains rs1421085	In first intron of *FTO*	*IRX3*
6	chr16:52377745–52378287	Contains rs9939609	In first intron of *FTO*	*IRX3*

The genomic positions of the elements are indicated in Fig. 1. Reporter expression patterns driven by each of the elements are shown in Fig. 2.
*The equivalent element in *Xenopus* is contained within the tested genomic interval scaffold_444:319497–322675 in genome assembly xenTro2.

which corresponds to a subdomain of *sox4* expression (Fig. 2D) unrelated to that of the broadly expressed *cdkal1* (Fig. S1). In addition, the human element directed expression to the primordium of the swim bladder, an endodermal derivative adjacent to the pancreas in fish that also expresses the islet marker *hlxb9* (21). These findings suggests that the tested element, which lies within the risk LD block, indeed regulates *SOX4* and is, together with prior findings implicating mouse and zebrafish *Sox4*/*sox4b* in pancreatic development/insulin secretion, highly suggestive of *SOX4* being the gene underlying the disease, rather than *CDKAL1*. Further work will be necessary to identify the functional SNP underlying disease susceptibility.

SNPs in *FTO* Are Within a Region Regulating *IRX3*. Variants rs8050136, rs9939609, rs1421085, and rs17817449 were found to be highly associated with T2D and obesity and lie in an LD block of 49 kb in the *FTO* gene (4, 6, 7). *FTO* contains numerous conserved noncoding elements in its introns and is located adjacent to a gene desert next to *IRX3*, a transcriptional regulator expressed in the kidney, notochord, forebrain, hypothalamus, and endodermal derivatives (22). We tested the expression driven by the 2 most deeply conserved HCNEs from the obesity-associated LD block: in both cases, we obtained expression patterns consistent with that of the *IRX3* gene (Fig. 2 *E* and *F*). Element 3 drove GFP expression in pronephric duct: *Irx3* is expressed in the developing kidney (22) and directs nephron segment identity (23). Element 4 drove reporter expression in the notochord, also a subdomain of the *Irx3* expression pattern (22). These results show that HCNEs located within the *FTO* obesity-risk LD block are most likely acting on *IRX3*, whereas zebrafish *fto* is at these stages not significantly expressed in the embryo. Neither of these subdomains, however, suggested a link to obesity, and no disease-associated SNPs map to those HCNEs. Upon testing the elements that overlap with risk SNPs rs1421085 and rs9939609, we noted that the majority of transgenes of the former (5 of 8) and a single transgenic line of the latter directed reporter expression to the pancreas (Fig. 2 *G* and

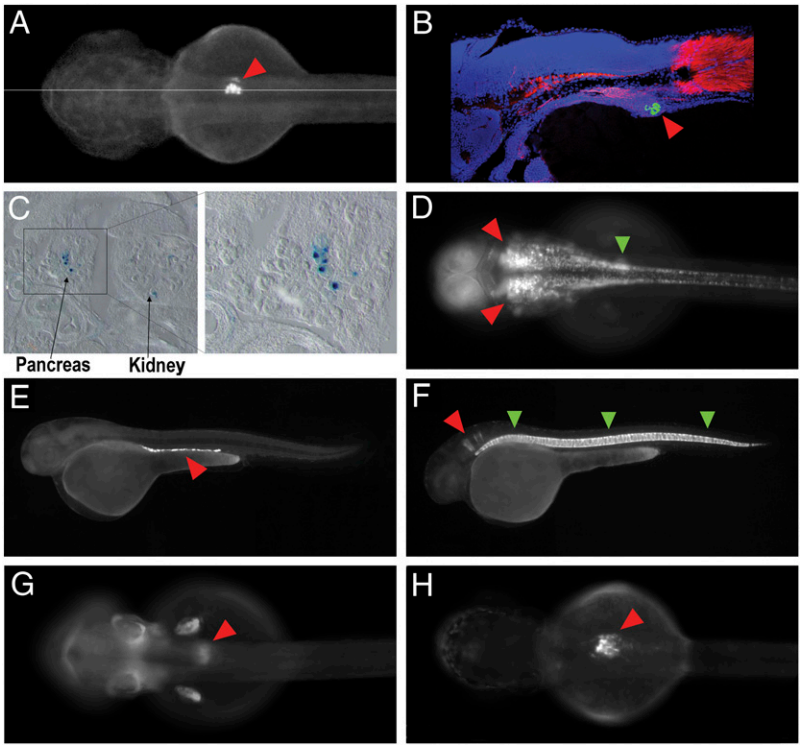

Fig. 2. Reporter expression patterns directed by noncoding elements from T2D risk regions. (*A* and *B*) Live dorsal image (*A*) and cryosection (*B*) of a 72-hpf transgenic zebrafish showing GFP-reporter expression driven by the *Xenopus* element equivalent to the rs1111875-containing human HCNE downstream of *HHEX* (element 1). Red arrowheads point to the pancreatic anlagen. The horizontal line in (*A*) shows the level of section in *B*. The cryosection (*B*) was stained for DNA using DAPI (blue), for muscle actin using rhodamine-phalloidin (red) and for GFP using anti-GFP (green). (*C*) Sections through the pancreas of a 14.5-d mouse transgenic for rs1111875-containing human HCNE downstream of HHEX (element 1). LacZ staining is seen in the pancreatic anlagen. (*D*) Live dorsal image of a 48-hpf transgenic zebrafish showing GFP-reporter expression driven by the human HCNE closest to rs7754840 in *CDKAL1* (element 2). Arrowheads point to expression in the hindbrain (red) and the primordium of the swim bladder (green). (*E* and *F*). Live lateral images of 48-hpf transgenic zebrafish showing GFP-reporter expression driven by the most deeply conserved HCNEs in the obesity-associated LD block of *FTO*. Element 3 (*E*) drives expression in the pronephric duct (red arrowhead), whereas element 4 (*F*) drives expression in the notochord (green arrowheads) and in hindbrain rhombomeres (red arrowhead). (*G* and *H*) Live dorsal images of 48-hpf transgenic zebrafish showing GFP-reporter expression driven by the rs1421085-containing HCNE and the rs9939609- containing HCNE in the *FTO* intron 1. Red arrowheads point to expression in the pancreatic area.

GENETICS

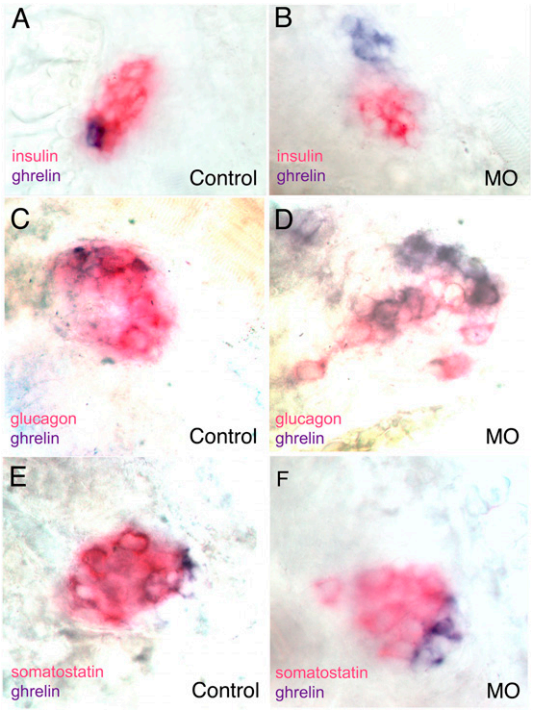

Fig. 3. Pancreatic buds of 48-hpf zebrafish embryos showing a reduction in the number of insulin-expressing cells and an increase of ghrelin-expressing cells in *irx3a* morphants. In situ hybridization with insulin (red) and ghrelin (blue) probes on control (*A*) and *irx3a* morphant embryos (*B*). Statistical support for the observation is given in Table 3. (*C* and *D*) *irx3a* morphants display a reduction in pancreatic α-cells. In situ hybridization with glucagon (red in *C* and *D*), somatostatin (red in *E* and *F*), and ghrelin (blue in *C*–*F*) probes on control (*C* and *E*) and *irx3a* morphant embryos (*D* and *F*). showing reduction in the number of α-cells (*D*) and increase of ghrelin-expressing cells (*D* and *F*) in the pancreatic bud of 48-hpf *irx3a* morphant embryos. The number of δ-cells is unaffected. Embryos are in ventral view. Statistical support for the observation is given in Table 4.

H), suggesting that these HCNEs might regulate *IRX3* expression in pancreas, and that its effect upon T2D might act through changes in glucose homeostasis. As we detected transcripts of the zebrafish *IRX3* orthologue *irx3a* in the embryonic area with the earliest insulin expression (Fig. S2), we next tested whether *irx3a* affects cell identity in the pancreas, akin to the transcriptional regulator *Nkx2.2*, which has been shown to regulate the number of ghrelin- and insulin- producing cells in the pancreatic islet in both mouse and zebrafish (24, 25).

Zebrafish *irx3a* Affects the Ratios of Insulin-, Ghrelin-, and Glucagon-Producing Islet Cells. The risk allele mapping to the LD block in *FTO* is linked to the development of T2D in obese subjects and a deregulation of the relative numbers of β- (i.e., insulin-producing) and ε-(i.e., ghrelin-producing) cells would be a plausible explanation for the co-occurrence of obesity and T2D. We therefore reasoned that *IRX3* might serve a role related to that of *Nkx2.2* during pancreas development.

To test this hypothesis, we generated morpholinos against the translation start site and the first splice junction of *irx3a* in zebrafish. The knockdown efficiency was tested by RT-PCR for skipping of the second exon of *irx3a* and at optimal conditions reached >99% (Fig. S3). Subsequent in situ hybridization for ghrelin and insulin revealed a significant increase for the former and a large decrease for the latter (Fig. 3 and Table 3 and Table 4), establishing the involvement of zebrafish *irx3a* in regulating the ratio of β- to ε-cells. In addition, knockdown of *irx3a* also diminished the number of α-cells producing glucagon, but not of those producing somatostatin (Fig. 3 and Table 4).

Neither *IRX3* nor any of its vertebrate orthologues had any prior association with pancreatic development, but mouse *Irx3*, together with *Nkx2.2*, acts as a potent repressor of *Hlxb9* during the development of spinal cord interneurons (26). *Hlxb9* encodes a transcriptional regulator whose loss of function also affects pancreas development (27, 28). We therefore tested how knockdown of *irx3a* combined with that of *nkx2.2a* would affect expression of *hlxb9* in a *hlxb9*:GFP transgenic zebrafish line (29). Knockdown of *irx3a* and *nkx2.2a* resulted in reduction of the numbers of *hlxb9*:GFP–expressing cells in the islet (Fig. S4*A*), whereas in the spinal cord, the expression domain of *hlxb9* predictably expanded dorsally in *irx3a* knockdowns and ventrally in *nkx2.2a* knockdowns (Fig. S4*B* and Table S1). This indicates that, although *irx3a*, *nkx2.2a* and *hlxb9* are involved in regulating the development of both spinal cord neurons and pancreatic islet cells, the interaction of *irx3a* and *nkx2.2a* with *hlxb9* appears to be different in these tissues.

Both the conservation of the genic neighborhood of *FTO*/*IRX3* across vertebrates and the finding that the strongest genome-wide signals for obesity map to an LD block containing HCNEs driving reporter expression in the pancreas strongly suggest that *FTO* is a bystander gene harboring HCNEs that drive *IRX3* expression. In the case of lower *IRX3* expression, the number of ghrelin-producing epsilon cells may increase at the expense of the pancreatic β- and α-cells, suggesting that a deregulation of pancreatic peptide hormones can underlie obesity and T2D with regard to the *FTO* association variants.

Discussion

Our analysis of 3 LD blocks containing T2D and obesity risk variants identified regulatory regions highly likely to belong to GRBs targeting developmental transcriptional regulators. This led us to link the regions containing the disease-associated SNPs to candidate genes that had a prior association with pancreatic development. None of our observations represents an unequivocal identification of a functional disease-causing SNP, but this was not our aim: instead, our results provide strong evidence that the enclosing LD blocks are within evolutionary conserved GRBs devoted to the regulation of particular developmental genes (*HHEX, SOX4*, and *IRX3*). Although, for both *HHEX* and *SOX4*, evidence was already available for their involvement in pancreas development and/or regulation of insulin secretion, the case of *IRX3* elucidates the power of GRB analysis of the results of GWA

Table 3. *irx3a* knockdown affects the numbers of pancreatic ε-, β-, and α-cells

Condition	*N*	Embryos with increased ghrelin expression (%)	Relative ghrelin area staining (average, *n* = 18)	Embryos with decreased insulin expression	Relative insulin area staining (average, *n* = 18)
MO2-irx3a	39	25/39 (64)	3.2	39/39	0.3
Controls	47	0/47	1	0/47	1

Twenty-five of 39 morphants (64.1%; $P < 0.0001$) displayed a dramatic increase of ghrelin expression and a reduction in the amount of insulin transcripts compared to uninjected embryos. By ImageJ-assisted area calculations on 18 morphant embryos, we estimated a mean reduction to 32.7% of the β-cell mass and an increase of the total number of ghrelin-expressing cells of approximately threefold.

Table 4. *irx3a* knockdown affects the numbers of pancreatic ε-, β-, and α-cells

Condition	*N*	Embryos with increased ghrelin expression (%)	Relative ghrelin area staining (average, *n* = 18)	Embryos with decreased glucagon expression	Relative glucagon area staining (average, *n* = 18)
MO2-irx3a	29	19/29 (65)	1.6	39/39	0.6
Controls	29	0/29	1	0/47	1

Nineteen of 29 morphants (65.5%; $P < 0.0001$) displayed a significant increase of ghrelin expression and a reduction in the amount of glucagon transcripts compared to mismatch control injected embryos. By ImageJ-assisted area calculations on 18 morphant embryos, we estimated a mean reduction to 65.3% of the α-cell mass and an increase of the total number of ghrelin-expressing cells of approximately 1.6-fold.

studies for the identification of hitherto unsuspected players in vertebrate development and human disorders. Although *IRX3* had no prior association with pancreatic development, we found among possible clues in the literature that mouse *Irx3*, together with *Nkx2.2*, acts as a potent repressor of *Hlxb9* during the development of spinal cord neurons (26). *Hlxb9* encodes a transcriptional regulator whose loss of function also affects pancreas development (27, 28). In the mouse, *Hlxb9* is necessary for early dorsal pancreas development as well as for β-cell development (27), whereas in zebrafish, *hlxb9* deficiency affects only β-cells (21). As extension of the early expression window of *Hlxb9* is also detrimental for murine pancreas development (30), one might speculate that *Hlxb9* repression during this period is also dependent upon *Irx3* and *Nkx2.2*, but this could not be ascertained in the zebrafish. Additionally, a direct effect of *irx3a*/*nkx2.2a* in the pancreas on *hlxb9* would not explain the effect of their knockdown upon α-cells, which, in the mouse, do not express *Hlxb9* (27). Any interaction of *nkx2.2a* and *irx3a* with *hlxb9* in pancreatic development therefore remains speculative at this point and will be the subject of future studies.

It is of note in this regard that 3 genes encoding transcriptional regulators, namely *Nkx2.2*, *Pax4*, and *Pax6*, were already identified as regulators of β- versus ε-cell fate (24, 31), with *Nkx2.2* and *Pax6* acting in the same pathway. It will be interesting to see whether any of these genes are also involved in the development of obesity, and how *IRX3* will fit into these pathways. We note that the principal site of ghrelin production in the mouse is the gastrointestinal tract, where there is no expression in zebrafish at the stages of our assay. However, loss of function of *Nkx2.2* also affects the identity of peptide hormone producing cells in the mouse gut (32), suggesting that *Irx3* may serve a related role in mammals. Interestingly, mouse *Irx1* and *Irx2* are also expressed in the pancreatic islets and in glucagon-producing cells, and act downstream of *Ngn3* (33). Together with our finding that glucagon-producing cells are also affected in the zebrafish *irx3a* morphant, these results suggest the elucidation of the regulation of islet cell identity as a promising avenue for future research in human metabolic disorders.

Although additional evidence will be required for any of the SNPs associated to disease to be functional, we note that the expression patterns driven in transgenic zebrafish by human HCNEs can yield useful clues whether they could be functional in a given expression context, in addition to ascertaining their regulatory function through potential changes in transcription factor–binding sites (7). It is of interest that the tested human elements are not conserved in the zebrafish genome, but still can be interpreted by the zebrafish embryo, similar to what was previously reported by Fisher and colleagues (34). While the present work was under consideration, a knockout of the mouse *Fto* gene was described, reporting postnatal growth retardation and increased energy expenditure (35). Notably, the targeted deletion removed exons 2 and 3, and thus did not touch upon the interaction of HCNEs in *Fto* intron 1 with the *Irx3* target gene. This knockout therefore did not test the function of the risk LD block, nor do these results contradict our findings. A second publication reported human FTO to be a gene causing recessive multiple malformations and lethality in early childhood, also accompanied by growth retardation (36). Interestingly, the authors noted the absence of obesity in the heterozygous parents.

It is striking that the risk sequence variants fall within GRBs of regulatory genes associated with multiple phenotypes (i.e., pleiotropic genes), which is not the case for T2D-associated SNPs that affect the coding region of genes, such as that of *SLC30A8* (1–4). In the former case, the affected genes we have suggested have complex spatiotemporal expression and multiple roles in development and differentiation, so different SNPs within large regions can affect different roles of the same gene, whereas the latter represents an islet-specific gene whose coding mutation is not expected to affect other tissues. Elucidation of the functional relationship between HCNEs and their target genes may hold a key to understanding the disease mechanism behind the T2D- and obesity-associated genomic variation. As the risk SNPs in *FTO* are common variants, it is tempting to speculate that they confer a selective advantage under specific environmental conditions. Such an advantage may be the "thrifty genotype" postulated by Neel (37), whereby natural selection during famine may enrich human populations for variants decreasing energy expenditure and increasing adiposity in times of plenty (38).

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 later, complemented with the Vertebrate Multiz Alignment and PhastCons Conservation and Chained Alignments tracks in the UCSC Genome Browser (39) for the human March 2006 assembly. Sequences were extracted from the UCSC Genome Browser for the human March 2006 assembly (hg18) or the *Xenopus tropicalis* August 2005 assembly (xenTro2).

Generation of Transgenic Zebrafish. Candidate HCNEs were amplified by PCR on human and *Xenopus* genomic DNA using the 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 an polyA signal, all flanked by Tol2 transposition sequences. Microinjection and screening were done as described (12). For description of bioinformatic methods, cloning of HCNEs, generation of transgenic mice, in situ hybridization, histology, and RT-PCR, see *SI Materials and Methods*.

Morpholinos. To knock down *irx3a* expression, we have used the following morpholinos: MO1-*irx3a* (AGCTGTGGGAAAGACATTGTTGTGG) and MO2-*irx3a* (GTGCTCCCTTAAAAACACAGAACAT), 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 (40). The morpholino for nkx2.2a (MOnk-5UTR, 5′-TGGAGCATTTGATGCAGTCAAGTTG) was the one reported by Pauls et al. (25). For the controls we used an unrelated/unspecific morpholino (GTtAATACcAGgATTAgATTgATTG).

ACKNOWLEDGMENTS. We gratefully acknowledge imaging help by Mary Laplante, and the Sars Centre zebrafish facility for animal maintenance. We thank Dr. Koichi Kawakami for the tol2 vector, Dr. Robb Krumlauf for the lacZ vector, Dr. Dirk Meyer for the hlxb9:GFP transgenic line, and Albert V. Smith for discussion. This work was funded in part by grants from the Sars Centre (Ø.D., P.N., B.L., and T.S.B.), grants from the University of Bergen (A.M., P.R.N., and T.S.B.), a grant from the Institut du Cerveau et de la Moelle épinière, Paris, France (T.S.B.), the FUGE Program at the Research Council of Norway (A.R., A.M., P.R.N. and B.L.), Helse Vest (A.M. and P.R.N.), Innovest (A.M. and P.R.N.), The Translational Fund (A.M. and P.R.N.), the YFF Program of the Research

Council of Norway and Bergen Forskningsstiftelse (B.L.), European Commission Grant LSHG-CT-2003-503469 as part of the ZF-Models integrated project in the 6th framework program (to T.S.B., F.A., and E.M.), grants BFU2007-60042/BMC, BFU2006-00349/BMC, and CSD2007-00008 from Ministerio de Educación y Ciencia of Spain (co-funded by Feder) (to J.L.G.S. and F.C.), Junta de Andalucía Grants CVI00260 and CVI 2658 (to J.L.G.S. and F.C.), Grant BFU2005-00025 from the Spanish Ministry of Education and Science (to M.E.A. and M.M.), the ProCNIC Foundation (M.E.A. and M.M.), CELDEV-CAM Grant S-SAL-0190-2006 from the Regional Government of Madrid (to M.E.A. and M.M.), CONSOLIDER-INGENIO Programme Grant 25120 (to J.L.G.S., F.C., M.E.A., and M.M.). M.J.T. was supported by Fundação para a Ciência e Tecnologia, Portugal.

1. Sladek R, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885.
2. Saxena R, et al.; Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336.
3. Zeggini E, et al. (2007) Replication of genome-wide association signals in U.K. samples reveals risk loci for type 2 diabetes. *Science* 317:1035–1036.
4. Scott LJ, et al. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345.
5. Steinthorsdottir V, et al. (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775.
6. Frayling TM, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889–894.
7. Dina C, et al. (2007) Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 39:724–726.
8. Bejerano G, et al. (2004) Ultraconserved elements in the human genome. *Science* 304: 1321–1325.
9. Sandelin A, et al. (2004) Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* 5:99.
10. Woolfe A, et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 3:e7.
11. Kikuta H, et al. (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* 17:545–555.
12. 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.
13. Bort R, Martinez-Barbera JP, Beddington RS, Zaret KS (2004) Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* 131:797–806.
14. Doria A, Patti ME, Kahn CR (2008) The emerging genetic architecture of type 2 diabetes. *Cell Metab* 8:186–200.
15. Pascoe L, et al.; RISC Consortium; U.K. Type 2 Diabetes Genetics Consortium (2007) Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. *Diabetes* 56:3101–3104.
16. Mavropoulos A, et al. (2005) sox4b is a key player of pancreatic alpha cell differentiation in zebrafish. *Dev Biol* 285:211–223.
17. Wilson ME, et al. (2005) The HMG box transcription factor Sox4 contributes to the development of the endocrine pancreas. *Diabetes* 54:3402–3409.
18. Goldsworthy M, et al. (2008) Role of the transcription factor sox4 in insulin secretion and impaired glucose tolerance. *Diabetes* 57:2234–2244.
19. Iglesias A, et al. (2004) Diabetes and exocrine pancreatic insufficiency in E2F1/E2F2 double-mutant mice. *J Clin Invest* 113:1398–1407.
20. Visel A, Bristow J, Pennacchio LA (2007) Enhancer identification through comparative genomics. *Semin Cell Dev Biol* 18:140–152.
21. Wendik B, Maier E, Meyer D (2004) Zebrafish mnx genes in endocrine and exocrine pancreas formation. *Dev Biol* 268:372–383.
22. Houweling AC, et al. (2001) Gene and cluster-specific expression of the Iroquois family members during mouse development. *Mech Dev* 107:169–174.
23. Alarcón P, Rodríguez-Seguel E, Fernández-González A, Rubio R, Gómez-Skarmeta JL (2008) A dual requirement for Iroquois genes during Xenopus kidney development. *Development* 135:3197–3207.
24. Prado CL, Pugh-Bernard AE, Elghazi L, Sosa-Pineda B, Sussel L (2004) Ghrelin cells replace insulin-producing beta cells in two mouse models of pancreas development. *Proc Natl Acad Sci USA* 101:2924–2929.
25. Pauls S, Zecchin E, Tiso N, Bortolussi M, Argenton F (2007) Function and regulation of zebrafish nkx2.2a during development of pancreatic islet and ducts. *Dev Biol* 304: 875–890.
26. Lee SK, Jurata LW, Funahashi J, Ruiz EC, Pfaff SL (2004) Analysis of embryonic motoneuron gene regulation: derepression of general activators function in concert with enhancer factors. *Development* 131:3295–3306.
27. Li H, Arber S, Jessell TM, Edlund H (1999) Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. *Nat Genet* 23:67–70.
28. Harrison KA, Thaler J, Pfaff SL, Gu H, Kehrl JH (1999) Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. *Nat Genet* 23:71–75.
29. Flanagan-Steet H, Fox MA, Meyer D, Sanes JR (2005) Neuromuscular synapses can form in vivo by incorporation of initially aneural postsynaptic specializations. *Development* 132:4471–4481.
30. Li H, Edlund H (2001) Persistent expression of Hlxb9 in the pancreatic epithelium impairs pancreatic development. *Dev Biol* 240:247–253.
31. Heller RS, et al. (2005) Genetic determinants of pancreatic epsilon-cell development. *Dev Biol* 286:217–224.
32. Desai S, et al. (2008) Nkx2.2 regulates cell fate choice in the enteroendocrine cell lineages of the intestine. *Dev Biol* 313:58–66.
33. Petri A, et al. (2006) The effect of neurogenin3 deficiency on pancreatic gene expression in embryonic mice. *J Mol Endocrinol* 37:301–316.
34. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS (2006) Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* 312:276–279.
35. Fischer J, et al. (2009) Inactivation of the Fto gene protects from obesity. *Nature* 458: 894–898.
36. Boissel S, et al. (2009) Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. *Am J Hum Genet* 85:106–111.
37. Neel JV (1962) Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J Hum Genet* 14:353–362.
38. Diamond J (2003) The double puzzle of diabetes. *Nature* 423:599–602.
39. Karolchik D, et al. (2008) The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* 36 (Database issue, Database issue) D773–D779.
40. Nasevicius A, Ekker SC (2000) Effective targeted gene 'knockdown' in zebrafish. *Nat Genet* 26:216–220.

Corrections

GENETICS

Correction for "Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to *HHEX*, *SOX4*, and *IRX3*," by Anja Ragvin, Enrico Moro, David Fredman, Pavla Navratilova, Øyvind Drivenes, Pär G. Engström, M. Eva Alonso, Elisa de la Calle Mustienes, José Luis Gómez Skarmeta, Maria J. Tavares, Fernando Casares, Miguel Manzanares, Veronica van Heyningen, Anders Molven, Pål R. Njølstad, Francesco Argenton, Boris Lenhard, and Thomas S. Becker, which appeared in issue 2, January 12, 2010, of *Proc Natl Acad Sci USA* (107:775–780; first published December 22, 2009; 10.1073/pnas.0911591107).

The authors note that, due to a printer's error, the second affiliation for Anders Molven should instead appear as "Department of Pathology, Haukeland University Hospital, N-5021 Bergen, Norway." The corrected affiliation line appears below. The online version has been corrected.

[a]Department of Clinical Medicine, University of Bergen, N-5020 Bergen, Norway; [b]Pediatrics, Haukeland University Hospital, N-5021 Bergen, Norway; [c]Dipartimento di Biologia, Universita' degli Studi di Padova, I-35121 Padua, Italy; [d]Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, N-5008 Bergen, Norway; [e]Sars Centre for Marine Molecular Biology, University of Bergen, N-5008 Bergen, Norway; [f]Fundación Centro Nacional de Investigaciones Cardiovasculares Carlos III, E-28029 Madrid, Spain; [g]Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide, Consejo Superior de Investigaciones Científicas, Sevilla 41013, Spain; [h]Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; [i]The Gade Institute, University of Bergen, N-5020, Bergen, Norway; and [j]Department of Pathology, Haukeland University Hospital, N-5021 Bergen, Norway

www.pnas.org/cgi/doi/10.1073/pnas.1101890108

BIOCHEMISTRY

Correction for "In vitro and in vivo reconstitution of the cadherin–catenin–actin complex from *Caenorhabditis elegans*," by Adam V. Kwiatkowski, Stephanie L. Maiden, Sabine Pokutta, Hee-Jung Choi, Jacqueline M. Benjamin, Allison M. Lynch, W. James Nelson, William I. Weis, and Jeff Hardin, which appeared in issue 33, August 17, 2010, of *Proc Natl Acad Sci USA* (107:14591–14596; first published August 5, 2010; 10.1073/pnas.1007349107).

The authors note that in the Materials and Methods under the section "Actin Pelleting Assay," the 10× stock concentration was listed instead of the 1× final working concentration for the polymerization and reaction buffers. The correct concentrations are as follows: polymerization buffer (20 mM Imidazole pH 7.0, 100 mM KCl, 2 mM $MgCl_2$, 0.5 mM ATP, 1 mM EGTA) and reaction buffer (20 mM Imidazole pH 7.0, 150 mM NaCl, 2 mM $MgCl_2$, 0.5 mM ATP, 1 mM EGTA, 1 mM DTT).

www.pnas.org/cgi/doi/10.1073/pnas.1102189108

MEDICAL SCIENCES

Correction for "Prolonged duration local anesthesia with minimal toxicity," by Hila Epstein-Barash, Iris Shichor, Albert H. Kwon, Sherwood Hall, Michael W. Lawlor, Robert Langer, and Daniel S. Kohane, which appeared in issue 17, April 28, 2009, of *Proc Natl Acad Sci USA* (106:7125–7130; first published April 13, 2009; 10.1073/pnas.0900598106).

The authors note that on page 7129, left column, fourth full paragraph, sentences 3 through 5, "DSPC:DSPG:cholesterol or DMPC:DMPG-cholesterol (molar ratio 3:1:2) were dissolved in *t*-butanol. Dexamethasone was added in some samples before lyophilization. The lyophilized cake was hydrated with 250 mM ammonium sulfate or, in some groups, with 0.1 mg STX, at 55–60 °C." should instead appear as "DSPC:DSPG:cholesterol or DMPC:DMPG-cholesterol (molar ratio 3:1:2) were dissolved in 9:1 vol/vol chloroform:methanol solution. Dexamethasone was added in some samples. A thin lipid film was obtained by drying the lipid mixture in a rotary evaporator at 60 °C. The pressure in the evaporator was maintained at 480 torr for 20 minutes and then gradually decreased to 20 torr over a period of 30 min. The lipid film was reconstituted in *t*-butanol at 55–60 °C and immediately cooled by immersing in liquid nitrogen. The sample was then transferred to a lyophilizer for 24 h. The lyophilized cake was hydrated with 250 mM ammonium sulfate or, in some groups, with 1.0 mg STX, at 55–60 °C."

www.pnas.org/cgi/doi/10.1073/pnas.1102194108