

1 Rfx6 directs islet formation and insulin production in mice and humans

Stuart B. Smith¹, Hui-Qi Qu^{2*}, Nadine Taleb^{2*}, Nina Y. Kishimoto^{1*}, David W. Scheel^{1*}, Yang Lu², Ann-Marie Patch³, Rosemary Grabs², Juehu Wang¹, Francis C. Lynn^{1†}, Takeshi Miyatsuka¹, John Mitchell², Rina Seerke¹, Julie Désir⁴, Serge Vanden Eijnden⁴, Marc Abramowicz⁴, Nadine Kacet⁵, Jacques Weill⁵, Marie-Ève Renard⁵, Mattia Gentile⁶, Inger Hansen⁷, Ken Dewar⁸, Andrew T. Hattersley³, Rennian Wang⁹, Maria E. Wilson¹⁰, Jeffrey D. Johnson¹⁰, Constantin Polychronakos² & Michael S. German^{1,11}

Insulin from the β -cells of the pancreatic islets of Langerhans controls energy homeostasis in vertebrates, and its deficiency causes diabetes mellitus. During embryonic development, the transcription factor neurogenin 3 (Neurog3) initiates the differentiation of the β -cells and other islet cell types from pancreatic endoderm, but the genetic program that subsequently completes this differentiation remains incompletely understood. Here we show that the transcription factor Rfx6 directs islet cell differentiation downstream of Neurog3. Mice lacking Rfx6 failed to generate any of the normal islet cell types except for pancreatic-polypeptide-producing cells. In human infants with a similar autosomal recessive syndrome of neonatal diabetes, genetic mapping and subsequent sequencing identified mutations in the human *RFX6* gene. These studies demonstrate a unique position for Rfx6 in the hierarchy of factors that coordinate pancreatic islet development in both mice and humans. Rfx6 could prove useful in efforts to generate β -cells for patients with diabetes.

During embryonic development, the pancreas first appears as clusters of cells on the dorsal and ventral aspects of the gut endoderm. The exocrine and endocrine cells that form the adult pancreas differentiate from this pool of pancreatic progenitors¹. A single transcription factor, the pro-endocrine bHLH factor Neurog3, is both necessary and sufficient to drive these progenitor cells to differentiate into the endocrine cells that form the islets of Langerhans^{2–4}. Transient activation of Neurog3 expression in scattered progenitor cells initiates expression of additional transcription factors, including Neurod1, Pax4, Nkx2-2, Nkx6-1, Arx and others, which then direct the differentiation of those cells into the distinct islet cell subtypes and the activation of the mature islet transcription factors such as Mafa, Pax6 and Isl1 (ref. 1). Mutations in many of these genes can cause diabetes, highlighting the pathway's importance in human β -cell formation and insulin production⁵. Understanding and controlling this process of differentiation may ultimately provide us with the cells needed to treat diabetes mellitus.

Expression of Rfx6

In independent screens for genes co-expressed with Neurog3 in islet progenitor cells⁶, activated by Neurog3 (ref. 7) and uniquely expressed in islets (M.E.W. and J.D.J., unpublished data), we identified Rfx6, a member of the RFX (regulatory factor X-box binding) family of winged-helix transcription factors^{8,9}. *Rfx6* transcripts could be detected in mouse and human embryonic pancreas by PCR with reverse transcription (RT-PCR; Fig. 1a, b) but, unlike all other known islet transcription factors, not in brain (Fig. 1a, c), and not in mouse embryonic pancreas lacking Neurog3 (Fig. 1d). In contrast,

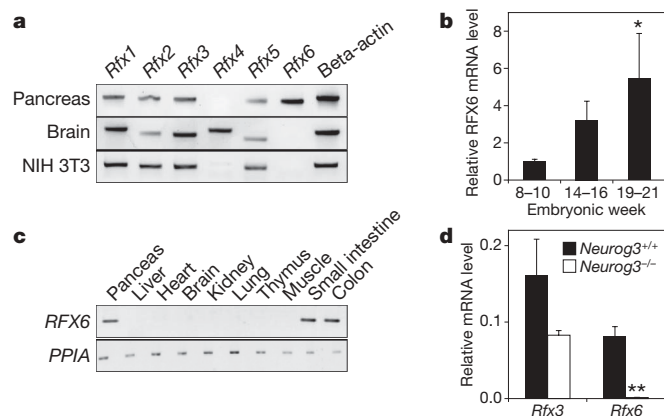


Figure 1 | Expression of Rfx6 in mice and human tissues. **a**, mRNA for *Rfx* genes 1–6 were amplified by RT-PCR from RNA isolated from the pancreas and brain of mouse embryos at e17.5 and from NIH 3T3 fibroblasts. **b**, Levels of *RFX6* mRNA were determined by real-time PCR of RNA from whole pancreas of human fetuses at the ages shown. $n = 5$ samples per fetal age group. * $P = 0.017$, weeks 8–10 vs 19–21, by two-tailed Student's *t*-test. **c**, mRNA for *RFX6* and control gene cyclophilin A (*PPIA*) genes were amplified by RT-PCR from RNA isolated from the human adult tissues shown. **d**, Levels of *Rfx3* and *Rfx6* mRNA were determined by real-time RT-PCR (TaqMan) of RNA isolated from the pancreata of wild-type and *Neurog3*^{-/-} mouse embryos at e17.5 values and expressed relative to the level of *Gusb*. $n = 3$ samples per group. ** $P = 0.0025$, wild type vs mutant, by two-tailed Student's *t*-test.

¹Diabetes Center, University of California San Francisco, San Francisco, California 94143, USA. ²Departments of Paediatrics and Human Genetics, McGill University, Montreal, Québec, Canada. ³Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, UK. ⁴Department of Neonatology, Hôpital Erasme-ULB, Brussels, Belgium. ⁵Department of Neonatology, Hôpital Calmette, Lille, France. ⁶Medical Genetic Unit, Di Venere General Hospital, Bari, Italy. ⁷Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, USA. ⁸Department of Human Genetics, McGill University, and Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada. ⁹Departments of Physiology, Pharmacology & Medicine, Child Health Research Institute, the University of Western Ontario, London, Ontario, Canada. ¹⁰Metabolex Inc., Hayward, California 94545, USA. ¹¹Department of Medicine, University of California San Francisco, San Francisco, California 94143, USA. †Current address: Departments of Surgery and Cellular and Physiological Sciences, Faculty of Medicine, University of British Columbia, British Columbia, Canada.

*These authors contributed equally to this work.

mouse *Rfx4*, the mammalian RFX gene with the highest homology to *Rfx6*, was amplified from brain and not from pancreas, and the other RFX genes were more widely expressed (Fig. 1a).

To explore the pattern of Rfx6 protein expression, we used anti-serum generated against recombinant Rfx6 protein for immunofluorescence studies. In mice, Rfx6 was detected as early as embryonic day 7.5 (e7.5) throughout the definitive, but not extra-embryonic, endoderm and persisted broadly in gut endoderm at e9.0, after which it becomes progressively restricted to the pancreas and scattered cells in the gut (Fig. 2a–c, Supplementary Fig. 1 and data not shown). At e10.0, immunofluorescence staining detected Rfx6 in foregut/midgut epithelium and in scattered cells in the nascent pancreatic buds, as indicated by staining for pancreatic transcription factor Pdx1 (Fig. 2d–f). Most of these scattered Rfx6-expressing cells did not co-express Pdx1, but many co-expressed Nkx2-2 and Neurog3 (Supplementary Figs 1 and 2). By e12.5, the Rfx6-expressing cells were generally distinct from the Pdx1-expressing progenitor cells, but most co-expressed glucagon, demonstrating restricted expression of Rfx6 in the endocrine lineage even at this early stage (Supplementary Fig. 3). In pancreata from *Neurog3*^{-/-} embryos¹⁰, there were no Rfx6-expressing cells (Supplementary Fig. 4).

During the peak of endocrine cell differentiation at e15.5, Rfx6 co-localized with Neurog3 in the nuclei of a subset of the endocrine progenitor cells (Fig. 2g), and overlapped with the islet transcription factors Nkx2-2, Nkx6-1 and Pdx1 (Supplementary Fig. 5). At e18.5, Rfx6 could be found in the nuclei of cells expressing each of the major

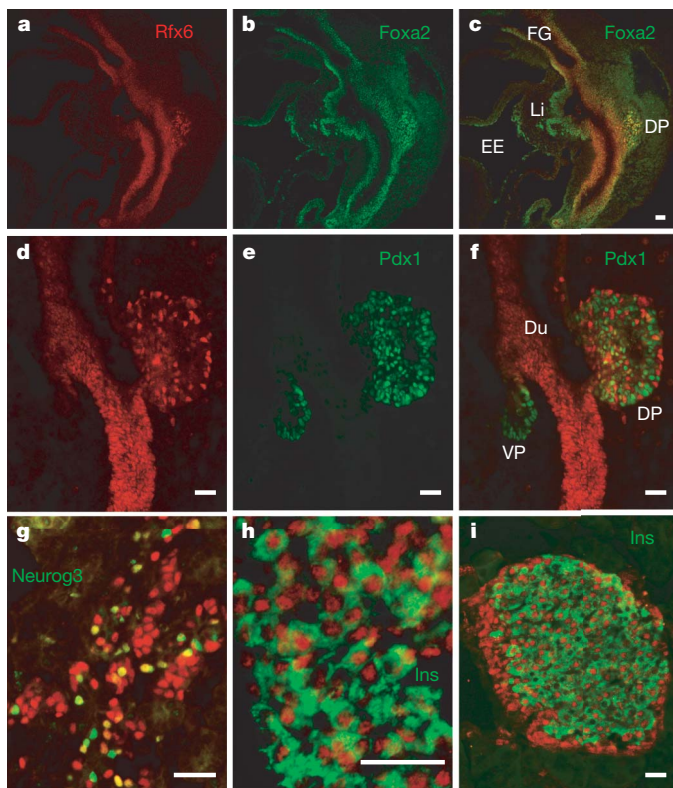


Figure 2 | Expression of Rfx6 in mice. Immunofluorescence staining was performed for Rfx6 (red) in mouse embryos. **a–c**, Rfx6 staining overlaps with Foxa2 in the gut epithelium (including foregut, FG) and nascent dorsal pancreatic bud (DP) at e9, but Foxa2 is expressed alone in the liver bud (Li) and extra-embryonic endoderm (EE). Separate colour channels are shown for red (**a** and **d**) and green (**b** and **e**). **d–f**, co-staining was performed with Pdx1 (green) in gut (duodenum, Du), dorsal pancreas (DP) and ventral pancreas (VP) at e10. **g**, e15.5 pancreas co-stained for Neurog3 (green). Co-staining nuclei appear yellow. **h**, e18.5 pancreas co-stained for insulin (green). **i**, Adult pancreas co-stained for insulin (green). Higher resolution photomicrographs from additional dates with additional markers can be found in Supplementary Figs 1–7. Scale bars, 25 μ m.

2

Nature nature08748.3d 13/1/10 16:04:47

pancreatic endocrine hormones (Fig. 2h and Supplementary Fig. 6). In the adult pancreas, Rfx6 expression was restricted to the islets where it could be detected in all endocrine lineages (Fig. 2i and Supplementary Fig. 7).

To generate Rfx6 null mice, we used homologous recombination to replace the first five exons of the Rfx6 gene, including the sequences encoding the DNA-binding domain, with a cassette encoding an enhanced green fluorescent protein (eGFP)–cre fusion protein (Supplementary Fig. 8). By crossing mice heterozygous for the mutant allele with mice carrying the marker gene *ROSA26* (also known as *Gt(ROSA)26Sor*) *loxP-stop-loxP lacZ* (*R26R*) (ref. 11), we generated *Rfx6*^{+/*eGFPcre*}/*R26R* double heterozygous mice in which Rfx6-expressing cells and their descendants are marked by the expression of β -galactosidase and can be visualized with the 5-bromo-4-chloro-3-indoyl- β -D-galactoside (X-gal) substrate (Fig. 3a–d and Supplementary Fig. 9). β -galactosidase expression was detected in all embryonic tissue derived from the endoderm germ layer, but not in other embryonic or in extra-embryonic tissues, demonstrating that Rfx6 is broadly expressed in and restricted to the definitive endoderm before the formation of the endoderm-derived organs. Taken together, the immunohistochemistry data and lineage tracing demonstrate that Rfx6 is expressed initially broadly in the definitive endoderm after gastrulation, becomes restricted to the gut and pancreatic bud at mid gestation, is reactivated by Neurog3 in islet progenitor cells and is ultimately restricted to pancreatic islets in the mature pancreas.

Rfx6^{eGFPcre/eGFPcre} mice

From heterozygous crosses, homozygous *Rfx6*^{eGFPcre/eGFPcre} mice were born at the expected Mendelian ratio, but failed to feed normally, exhibited gross bowel distension due to small bowel obstruction

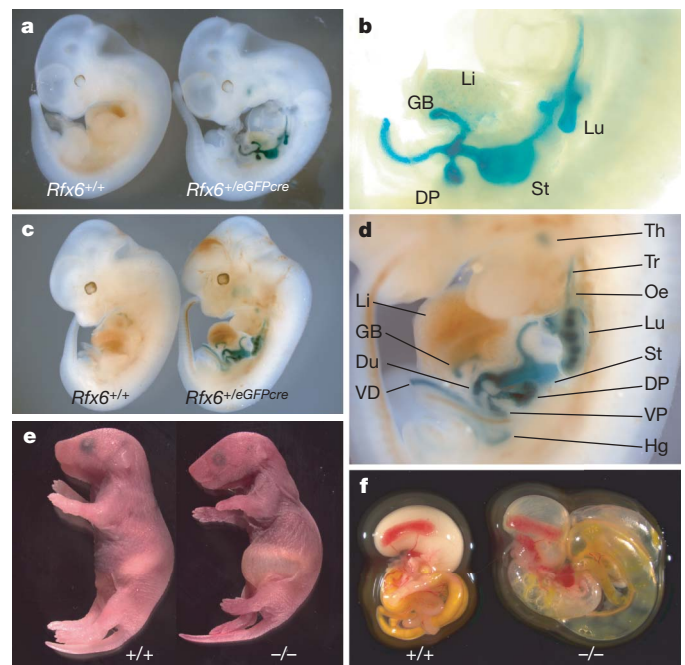


Figure 3 | Targeting of the Rfx6 gene in mice. **a–d**, Lineage tracing performed on *Rfx6*^{+/*R26R*} (left in **a** and **c**) or *Rfx6*^{+/*eGFPcre*}/*R26R* (**b** and **d**, and right in **a** and **c**) mice at e10.5 (**a**, **b**) and e12.5 (**c**, **d**) by staining for β -galactosidase activity with X-gal (blue). **b**, Close-up view of the animal on the right in panel **a**, and **d** shows a close-up view of the animal on the right in panel **c**. **e**, an *Rfx6*^{eGFPcre/eGFPcre} pup at postpartum day 2 (p2) is shown on the right, with a wild-type litter mate on the left. **f**, Dissected abdominal viscera are shown for wild-type (left) and *Rfx6*^{eGFPcre/eGFPcre} (right) pups at p0.5. Additional photographs of the lineage tracing and mutant animals can be found in Supplementary Fig. 10. Li, liver; Du, duodenum; GB, gall bladder; VD, vitelline duct; Th, thymus; Tr, trachea; Oe, oesophagus; Lu, lung; St, stomach; DP, dorsal pancreas; VP, ventral pancreas; Hg, hindgut.

(Fig. 3e–f) and died within 2 days post partum. Some, but not all, of the *Rfx6* null animals also had reduced pancreas size (data not shown).

To test for effects on gene expression before birth, we harvested RNA from e17.5 pancreata, and used low density TaqMan arrays⁶ to measure the levels of a set of pancreatic genes (Supplementary Table 1). *Rfx6*^{eGFPcre/eGFPcre} pancreata had almost no expression of the islet hormones genes, except for pancreatic polypeptide (*Ppy*). Several other islet/ β -cell genes, such as the zinc transporter *Slc30a8* and G-protein coupled receptor *GPR40* (also known as *Ffar1*) genes, were similarly reduced; but other β -cell genes were more modestly reduced, including the glucose sensing genes *Gck*, *Slc2a2* and *Kcnj11*. Immunofluorescence staining with the endocrine markers chromogranin A or synaptophysin at e17.5 demonstrated that the *Rfx6*^{eGFPcre/eGFPcre} pancreata still contained a large number of endocrine cells, but confirmed that none of these cells expressed insulin, glucagon, somatostatin or ghrelin (Fig. 4). Although the number of *Ppy*-expressing cells (PP cells) was increased in the *Rfx6*^{eGFPcre/eGFPcre} pancreata, they only accounted for a subset of the endocrine cells (Fig. 4i–l), leaving the identity of the remaining endocrine cells unknown.

We also tested whether *Rfx6* regulates other transcription factor genes (Supplementary Table 2). The absence of *Rfx6* did not affect *Neurog3* expression, and this was confirmed at the protein level (data not shown). In sharp contrast, expression of genes downstream of *Neurog3* encoding factors involved in α -cell development, including *Irx2* and *Arx*, was markedly reduced in *Rfx6*^{eGFPcre/eGFPcre} pancreata. Interestingly, genes encoding several factors involved in insulin gene transcription (*Pax6*, *Mafa*, *Neurod1* and *Pdx1*) also had reduced expression, but some key genes involved in β -cell specification either did not significantly change (*Nkx2-2* and *Nkx6-1*), or increased (*Pax4*). Immunofluorescence staining at e17.5 revealed that the field of *Nkx6-1* expression expanded from β -cells alone in wild-type pancreata to include all of the chromogranin A positive endocrine cells, including the PP cells in *Rfx6*^{eGFPcre/eGFPcre} pancreata (Fig. 4i–l and Supplementary Table 3). These studies indicate that whereas *Rfx6* regulates the transcription factors involved in β -cell maturation and function, it restricts the expression of the β -cell differentiation and specification genes, and thus the β -cell fate choice.

Mice with a targeted disruption of the *Rfx3* gene have an islet phenotype that is similar to, but less extreme than, the *Rfx6*^{eGFPcre/eGFPcre}

mice, with reductions in the numbers, but not complete loss, of insulin- and glucagon-expressing cells and an increase in pancreatic polypeptide-expressing cells¹². In a proteome-wide screen of protein–protein interactions, *Rfx6* was found to interact with *Rfx2* and *Rfx3* (ref. 13). Since the RFX transcription factors generally bind to their target DNA sites (the ‘X box’) as dimers¹⁴, we tested whether *Rfx6* and *Rfx3* form a heterodimeric DNA binding complex in an electromobility shift assay (EMSA). We found that both full-length *Rfx3* and truncations that retain the DNA-binding and dimerization domains bound to an X-box site together with *Rfx6* (Fig. 5a and Supplementary Fig. 10), and that the two factors cooperated in activating a promoter containing multimers of this X-box site (Fig. 5b).

It has been proposed that the islet phenotype of the *Rfx3*^{-/-} mice results from defects in primary cilia formation on islet cells¹², although islets lacking any primary cilia develop fairly normally^{15,16}. Unlike in the *Rfx3*^{-/-} mice, we found that primary cilia formation was unaffected in the *Rfx6*^{eGFPcre/eGFPcre} islets (Supplementary Fig. 11a–d). In addition, expression of the cilia genes *Ift88* and *Dync2li1*, which are reduced in the pancreas of *Rfx3*^{-/-} mice¹², was not reduced in the pancreas of *Rfx6*^{eGFPcre/eGFPcre} mice (Supplementary Fig. 11e). We conclude that *Rfx3* and *Rfx6* cooperate in regulating a set of genes involved in islet development but not in cilia formation. The more modest islet phenotype in the *Rfx3*^{-/-} mice may be due to the ability of *Rfx2* to compensate partially for the loss of *Rfx3* (Fig. 1a), or the ability of *Rfx6* to direct gene expression as a homodimer (Fig. 5a, b).

Human mutations in RFX6

The phenotype of the *Rfx6*^{eGFPcre/eGFPcre} mice is remarkably similar to human patients born with neonatal diabetes and small bowel obstruction due to bowel atresia^{17,18}. Despite some reduction in pancreatic size, these cases were not deficient in enzymes of the exocrine pancreas, and autopsies of two cases (proband 1 and case 3 in ref. 17) revealed normal-appearing exocrine pancreata with clusters of chromogranin A-positive cells but total absence of cells staining for insulin, glucagon or somatostatin¹⁷. In addition, the syndrome involves hypoplastic gall bladder, and intractable diarrhoea unresponsive to pancreatic enzyme replacement.

The disease locus was mapped using overlapping homozygosity in probands 1 and 2 (see Supplementary material for pedigree information

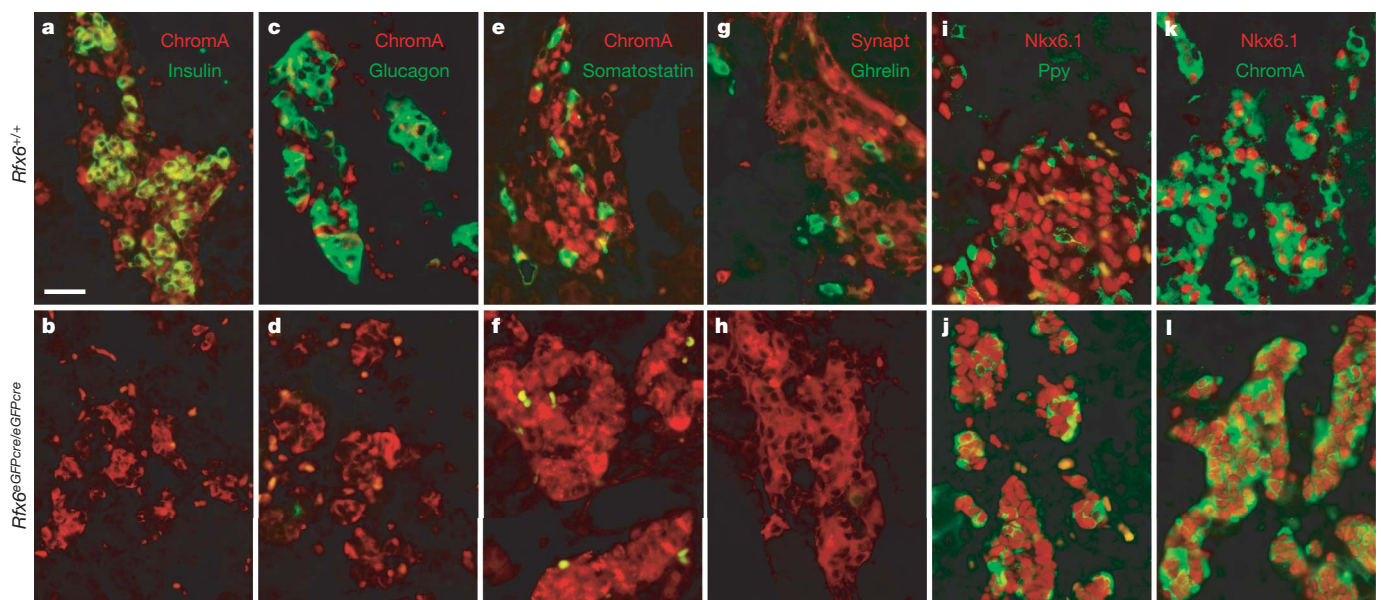


Figure 4 | Expression patterns of islet markers in wild-type and *Rfx6*^{eGFPcre/eGFPcre} mice at e17.5. a–f, Immunofluorescence co-staining was performed for chromogranin A (ChromoA, red, a–f, green k, l), synaptophysin (Syn, red, g, h), insulin (green, a, b), glucagon (green, c, d), somatostatin (Sst, green, e, f), ghrelin (green g, h), *Nkx6-1* (red in

i–l), and pancreatic polypeptide (*Ppy*, green, k, l) on pancreas sections from e17.5 embryos with the genotypes shown on the left. Quantification of cells expressing *Ppy* and *Nkx6-1* is shown in Supplementary Table 3. Scale bars, 25 μ m.

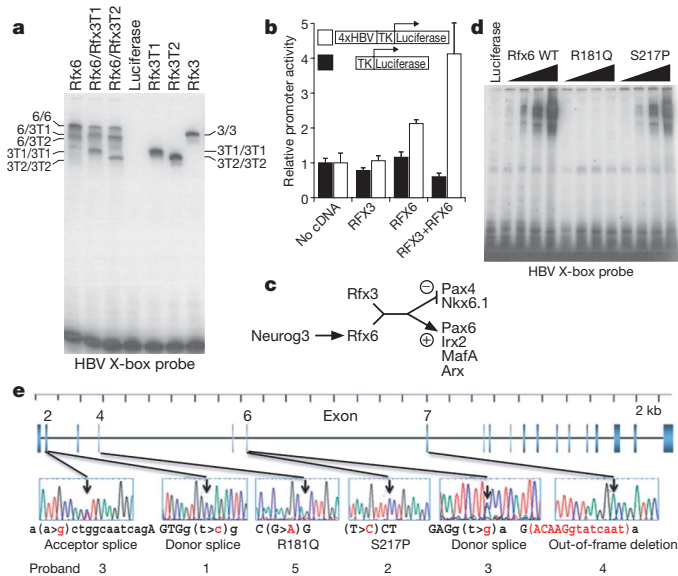


Figure 5 | Function of the human Rfx6 protein. **a**, DNA binding of the human *in vitro*-translated proteins shown above each lane to the double-stranded, radiolabelled oligonucleotide HBV X-box probe⁴² was tested by EMSA. Combined proteins were co-translated, and probe bound by the heterodimer partners has a mobility between that of the two homodimers. Truncated proteins Rfx3T1 and Rfx3T2 have the first 119 and 160 amino-terminal amino acids removed, respectively, but retain the DNA-binding and dimerization domains. *In vitro*-translated luciferase is included as a negative control. A close-up view of a longer gel is shown in Supplementary Fig. 10a. **b**, Mouse pancreatic ductal mPac L20 cells were co-transfected with a DNA plasmid containing the reporter constructs shown and another expressing the RFX cDNAs shown, luciferase reporter expression was assayed for each combination. **P* = 0.0026 vs no cDNA, 0.0024 vs Rfx3 alone, and 0.011 vs Rfx6 alone by two-tailed Student's *t*-test. **c**, Schematic shows the proposed interactions, either direct or indirect, of several transcription factors during pancreas development. **d**, Increasing amounts of the *in vitro*-translated human Rfx6 wild-type and 181 Arg→Gln (R181Q) and 217 Ser→Pro (S217P) mutant proteins were assayed for binding to the X-box DNA probe. Efficiency of mutant protein production is demonstrated in Supplementary Fig. 10b. **e**, Mutations found in patients are indicated on a map of the RFX6 gene. All mutations were homozygous except for proband 3.

and references to previous clinical case reports) respectively the offspring of first and second cousins. High-resolution homozygosity mapping identified 10 homozygosity-by-descent (HBD) segments >500 kb in proband 1 (after excluding those that overlapped with her unaffected sibling, Supplementary Table 4), and 25 HBD segments >500 kb in proband 2 (Supplementary Table 5). Only three HBD regions were common in the two probands, totalling 24 Mb (Table 1). Altogether, 194 RefSeq genes map to these regions. Of these genes, only *RFX6*, which falls in the largest segment at 6q21–22, had pancreas-enriched expression in the TiGER database¹⁹ (Supplementary Table 6), and also increased in expression in human pancreas between fetal ages 10 and 20 weeks (Fig. 1b and Supplementary Figs 13 and 14), concordant with the appearance of endocrine cells^{20,21}.

Two parallel, independent approaches unequivocally identified mutations in the *RFX6* gene in this human syndrome: direct sequencing of the *RFX6* gene and unbiased deep sequencing of all exons within the three overlapping HBD regions.

For deep sequencing, exons were captured from DNA obtained from proband 2 using a tiled oligonucleotide array²² covering 1,309

Table 1 | Regions of homozygosity-by-descent common to probands 1 and 2

Chromosome	Start SNP	Start position	End SNP	End position	Size
2p15-16	rs6754038	57,761,104	rs1426699	64,418,856	6,657,752
6q21-22	rs6913656	112,589,928	rs10499129	124,975,906	12,385,978
6q23	rs7744295	131,173,627	rs17065195	136,360,914	5,187,287

of 1,322 exons mapping within the HBD regions²³. Amplification and sequencing²⁴ of the captured fragments generated 40,379 sequences of at least 100 bp that aligned within the target regions. Median target coverage depth was 9.2, with 80% of targets having a depth of at least 4. Given that we were searching for a homozygous mutation, this was sufficient for unequivocal detection of exonic variants. Altogether, 30 novel sequence variants were detected (Supplementary Table 7): 15 in introns, 3 in both introns and untranslated regions (UTRs), 9 in UTRs, and only three in coding sequences, two synonymous. The only non-synonymous variant was 217 Ser→Pro in *RFX6*, identifying this gene as the most likely candidate.

In parallel, direct sequencing was performed on the 19 exons and the splicing junctions of *RFX6* in all probands. Missense, splicing or frameshift mutations in *RFX6* were found in five of the six available probands (Fig. 5e) with an interesting genotype–phenotype correlation. Probands 1, 4 and 5 all died in the first few months of life and were homozygous for, respectively, a loss of the donor splicing site in intron 2 (IVS2+2 t→c), an out-of-frame deletion in exon 7, and the missense mutation 181 Arg→Gln involving a highly conserved arginine in the DNA-binding domain^{8,25} (Supplementary Fig. 15). Proband 3, still alive at the age of 9 years and intermittently off insulin, was a compound heterozygote for donor-site loss in intron 6 (IVS6+2 t→g) and disruption of the acceptor site in intron 1 (IVS1–12 a→g). Proband 2, still alive at age 4.5 years¹⁸, had the homozygous missense mutation 217 Ser→Pro, confirming the unbiased exon sequencing described above. All mutations were inherited from carrier parents.

To determine the significance of the homozygous intron 2 donor splicing site mutation in proband 1, we amplified *RFX6* mRNA by RT–PCR of high-quality RNA from autopsy pancreas and failed to detect the properly spliced transcript, which was easily amplified from normal fetal pancreas as was the reference gene cyclophilin in the proband's RNA. We also failed to detect any RNA from exons 1+2, upstream of the splicing mutation, probably due to nonsense-mediated decay²⁶ (Supplementary Fig. 16).

We also tested the two missense mutations for their effect on DNA binding by Rfx6. We found that 181 Arg→Gln (proband 5), which alters a conserved amino acid in the DNA-binding domain, completely abrogated DNA binding, whereas 217 Ser→Pro (proband 2), which lies between the DNA-binding domain and dimerization domain of Rfx6, only modestly reduced DNA binding (Fig. 5d) and did not affect dimer formation (data not shown).

Finally, we failed to identify any mutation in *RFX6* in proband 6. In the absence of DNA from the proband, we sequenced both parents and found no point mutation of *RFX6* or *Neurog3*; and long-range PCR did not reveal any deletions of *RFX6* (Supplementary Fig. 17). In the absence of proband DNA we cannot rule out a *de novo* mutation, but this case is most likely a phenocopy. We also failed to find *RFX6* mutations in a case of the Martinez-Frias syndrome²⁷. Finally, a search of the *RFX6* linkage disequilibrium block in our genome-wide association data^{28,29}, combined with those of the WTCC³⁰, did not reveal any common variants associated with type 1 or type 2 diabetes (data not shown).

Discussion

In the pancreas, Rfx6 acts downstream of the pro-endocrine factor *Neurog3* (Fig. 5c), and mutation of the two genes give similar but distinct phenotypes. Like *RFX6* mutations, mutation of *Neurog3* in humans also causes intractable diarrhoea and diabetes³¹, but *Neurog3* mutation does not cause the small bowel atresia or biliary abnormalities seen with *RFX6* mutation, which is not surprising since *Neurog3* is not expressed in the early gut endoderm as is Rfx6^{4,32,33}. Despite severe intestinal malabsorption, pancreatic exocrine function is intact in both syndromes, and the only histological abnormality in the gut in the mutant *Neurog3* syndrome is loss of the intestinal endocrine cells³⁴. Comparison of the gut endocrine cells affected by these two related

syndromes may provide new insight into mechanisms that regulate nutrient absorption in the small bowel.

In addition, *RFX6* mutations cause diabetes at birth, whereas the reported patients with homozygous mutations in *NEUROG3* did not develop diabetes until several years later³¹, despite evidence that Neurog3 is absolutely required for the generation of islet cells and production of insulin in mice². Incomplete loss of function in the reported human *NEUROG3* mutations could explain the continued insulin production in these patients³⁵. Alternatively, recent evidence that *Neurog3*^{-/-} mice still generate a small number of islet cells³⁶ indicates some redundancy of Neurog3's pro-endocrine function, possibly owing to the presence of related bHLH proteins in the pancreas^{4,7}. In contrast, our data do not indicate any redundancy of Rfx6 function in endocrine differentiation and demonstrate remarkable conservation of the genetic control of islet development, despite some discrepancy between mouse and human *Neurog3* mutants.

It should be noted that our patients with *RFX6* mutations have some similarities with other reported cases, especially those with the Martinez-Frias syndrome (MFS)²⁷, under which OMIM currently lists our cases. However, since the original MFS patients also had oesophageal atresia and hypospadias, and did not have diabetes, and we did not find any *RFX6* mutation in the parents of one case of MFS, our cases are distinct from MFS. We propose to name their condition, now of defined molecular aetiology, the Mitchell–Riley syndrome after the two clinicians who first described it¹⁷.

In summary, we have identified a novel factor in endoderm and islet development, Rfx6, that is required for the differentiation of four of the five islet cell types and for the production of insulin in both humans and mice. In the hierarchy of islet developmental factors, it lies downstream of Neurog3 and upstream of many of the other islet transcription factors (Fig. 4d). A full understanding of the role of Rfx6 will help to clarify how islets and β -cells are generated, what goes wrong in this process in diabetes, and how to generate new β -cells for patients with diabetes.

METHODS SUMMARY

All studies involving mice were approved by the UCSF Institutional Animal Care and Use Committee. Timed matings were carried out with embryonic day 0.5 being set as midday of the day of discovery of a vaginal plug. The *Rfx6*-targeting allele (Supplementary Fig. 8) was generated by recombining in a modified bacterial artificial chromosome followed by recombination into plasmid DNA by gap-repair³⁷. This construct was used by the UCSF DERC Transgenic Core Laboratory to target the *Rfx6* allele in 129 (E14) mouse embryonic stem cells, which were injected into mouse blastocysts to generate chimaeric *Rfx6*^{+;eGFP} mice. Chimeras and subsequent generations were crossed to C57BL/6 mice. Mouse tissue processing, β -galactosidase detection, immunofluorescence staining³⁸, automated cell counting³⁹, transfections of mPAC cells, EMSA⁴⁰ and mRNA quantification with low density TaqMan arrays⁶ were performed as described previously. All antisera used for immunofluorescence studies are listed in Supplementary Table 8. Sequences of oligonucleotides used for RT–PCR, ESC colony screening, mouse genotyping and EMSA are available on request.

The human subjects protocol was approved by the IRB of the Montreal Children's Hospital and written informed consent was obtained from all participating families. Clinical findings and case-report references are summarized in the web supplement. For homozygosity mapping, results from the Illumina Hap 550 (proband 1, call rate 0.99) or 1M (proband 2, call rate 0.978) microarrays⁴¹, were used to scan autosomes in 300-kb windows and HBD was defined as absence of any heterozygous single nucleotide polymorphism (SNP) in the proband and presence of at least one in either parent or the unaffected sibling. Long-oligo Nimblegen capture arrays²² included around 100 bp extension of intronic sequence coverage from the boundary of the exons. The eluted enriched regions from proband 1 were run on the Roche 454 FLX Genome Sequencer²⁴. The 19 exons of *RFX6* were PCR-amplified manually for Sanger sequencing in the remaining probands. For the long range PCR, we used kit K0182 from Fermentas for fragments ranging from 6 to 17 kb (Supplementary Fig. 17).

Received 30 June; accepted 7 December 2009.

Published online XX 2009.

1. Murtaugh, L. C. Pancreas and beta-cell development: from the actual to the possible. *Development* **134**, 427–438 (2007).

2. Gradwohl, G., Dierich, A., LeMeur, M. & Guillemot, F. *neurogenin3* is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl Acad. Sci. USA* **97**, 1607–1611 (2000).
3. Apelqvist, A. *et al.* Notch signalling controls pancreatic cell differentiation. *Nature* **400**, 877–881 (1999).
4. Schwitzgebel, V. M. *et al.* Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* **127**, 3533–3542 (2000).
5. McCarthy, M. I. & Hattersley, A. T. Learning from molecular genetics: novel insights arising from the definition of genes for monogenic and type 2 diabetes. *Diabetes* **57**, 2889–2898 (2008).
6. Miyatsuka, T., Li, Z. & German, M. S. Chronology of islet differentiation revealed by temporal cell labeling. *Diabetes* **58**, 1863–1868 (2009).
7. Gasa, R. *et al.* Induction of pancreatic islet cell differentiation by the neurogenin-neuroD cascade. *Differentiation* **76**, 381–391 (2008).
8. Aftab, S., Semenec, L., Chu, J. S. & Chen, N. Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evol. Biol.* **8**, 226 (2008).
9. Emery, P., Durand, B., Mach, B. & Reith, W. RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res.* **24**, 803–807 (1996).
10. Lee, C. S., Perreault, N., Brestelli, J. E. & Kaestner, K. H. *Neurogenin 3* is essential for the proper specification of gastric enteroendocrine cells and the maintenance of gastric epithelial cell identity. *Genes Dev.* **16**, 1488–1497 (2002).
11. Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
12. Ait-Lounis, A. *et al.* Novel function of the ciliogenic transcription factor RFX3 in development of the endocrine pancreas. *Diabetes* **56**, 950–959 (2007).
13. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
14. Emery, P. *et al.* A consensus motif in the RFX DNA binding domain and binding domain mutants with altered specificity. *Mol. Cell. Biol.* **16**, 4486–4494 (1996).
15. Cano, D. A., Sekine, S. & Hebrok, M. Primary cilia deletion in pancreatic epithelial cells results in cyst formation and pancreatitis. *Gastroenterology* **131**, 1856–1869 (2006).
16. Cano, D. A., Murcia, N. S., Pazour, G. J. & Hebrok, M. Orpk mouse model of polycystic kidney disease reveals essential role of primary cilia in pancreatic tissue organization. *Development* **131**, 3457–3467 (2004).
17. Mitchell, J. *et al.* Neonatal diabetes, with hypoplastic pancreas, intestinal atresia and gall bladder hypoplasia: search for the aetiology of a new autosomal recessive syndrome. *Diabetologia* **47**, 2160–2167 (2004).
18. Chappell, L. *et al.* A further example of a distinctive autosomal recessive syndrome comprising neonatal diabetes mellitus, intestinal atresias and gall bladder agenesis. *Am. J. Med. Genet. A* **146A**, 1713–1717 (2008).
19. Liu, X., Yu, X., Zack, D. J., Zhu, H. & Qian, J. TIGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* **9**, 271 (2008).
20. Stefan, Y., Grasso, S., Perrelet, A. & Orci, L. A quantitative immunofluorescent study of the endocrine cell populations in the developing human pancreas. *Diabetes* **32**, 293–301 (1983).
21. Lyttle, B. M. *et al.* Transcription factor expression in the developing human fetal endocrine pancreas. *Diabetologia* **51**, 1169–1180 (2008).
22. Hodges, E. *et al.* Genome-wide *in situ* exon capture for selective resequencing. *Nature Genet.* **39**, 1522–1527 (2007).
23. Karolchik, D. *et al.* The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* **36**, D773–D779 (2008).
24. Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nature Methods* **4**, 903–905 (2007).
25. Gajiwala, K. S. *et al.* Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**, 916–921 (2000).
26. Chang, Y. F., Imam, J. S. & Wilkinson, M. F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **76**, 51–74 (2007).
27. Gentile, M. & Fiorente, P. Esophageal, duodenal, rectoanal and biliary atresia, intestinal malrotation, malformed/hypoplastic pancreas, and hypospadias: further evidence of a new distinct syndrome. *Am. J. Med. Genet.* **87**, 82–83 (1999).
28. Hakonarson, H. *et al.* A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes* **57**, 1143–1146 (2008).
29. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
30. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
31. Wang, J. *et al.* Mutant neurogenin-3 in congenital malabsorptive diarrhea. *N. Engl. J. Med.* **355**, 270–280 (2006).
32. Jensen, J. *et al.* Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. *Diabetes* **49**, 163–176 (2000).
33. Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–2457 (2002).
34. Cortina, G. *et al.* Enteroendocrine cell dysgenesis and malabsorption, a histopathologic and immunohistochemical characterization. *Hum. Pathol.* **38**, 570–580 (2007).
35. Jensen, J. N. *et al.* Mutant neurogenin-3 in congenital malabsorptive diarrhea. *N. Engl. J. Med.* **356**, 1781–1782; author reply 1782 (2007).



36. Wang, S. *et al.* Myt1 and Ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation. *Dev. Biol.* **317**, 531–540 (2008).
37. Copeland, N. G., Jenkins, N. A. & Court, D. L. Recombineering: a powerful new tool for mouse functional genomics. *Nature Rev. Genet.* **2**, 769–779 (2001).
38. Nekrep, N., Wang, J., Miyatsuka, T. & German, M. S. Signals from the neural crest regulate beta-cell mass in the pancreas. *Development* **135**, 2151–2160 (2008).
39. Lynn, F. C. *et al.* MicroRNA expression is required for pancreatic islet cell genesis in the mouse. *Diabetes* **56**, 2938–2945 (2007).
40. Lynn, F. C. *et al.* Sox9 coordinates a transcriptional network in pancreatic progenitor cells. *Proc. Natl Acad. Sci. USA* **104**, 10500–10505 (2007).
41. Steemers, F. J. & Gunderson, K. L. Whole genome genotyping technologies on the BeadArray platform. *Biotechnol. J.* **2**, 41–49 (2007).
42. David, E., Garcia, A. D. & Hearing, P. Interaction of EF-C/RFX-1 with the inverted repeat of viral enhancer regions is required for transactivation. *J. Biol. Chem.* **270**, 8353–8360 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all the patients and their families for the participation in this study and P. Riley for allowing us access to her clinical data. We thank G. Grodsky, G. Bell, W. Rutter, R. Gasca and members of the German laboratory for discussions, F. Schaufle and the DERC Microscopy Core Laboratory, C. Mrejen and the UCSF DERC Genomics Core, N. Killeen and the UCSF DERC Transgenic Core Laboratory for help with the generation of the *Rfx6*-targeted mice,

R. Koshy for technical assistance, Y. Zhang and S. Zhao for assistance with mouse husbandry and genotyping, the Massively Parallel Sequencing team at the McGill University and Genome Quebec Innovation Center for DNA sequencing and J. Wasserscheid for bioinformatics analyses. This work was supported by grants from the Larry L. Hillblom Foundation (S.B.S. and M.S.G.), the Juvenile Diabetes Research Foundation (S.B.S., F.C.L., T.M., R.W., C.P. and M.S.G.), the American Diabetes Association (M.S.G.), the Nora Eccles Treadwell Foundation (M.S.G.), the Canadian Institutes of Health Research (H.Q.Q.), and the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases (M.S.G.).

Author Contributions S.B.S., H.Q.Q., N.T., N.Y.K., D.W.S., F.C.L., K.D., R.W., C.P. and M.S.G. wrote the paper. K.D., C.P. and M.S.G. oversaw the studies. S.B.S., D.W.S., Y.L., J.Wa., T.M., R.W., M.E.W. and J.D.J. performed mRNA expression analyses. N.Y.K., D.W.S. and J.W. performed immunofluorescence studies. S.B.S., F.C.L. and R.S. performed *Rfx6* gene targeting studies. S.B.S. performed DNA binding and transcription studies. H.Q.Q. performed homozygosity mapping. N.T., R.G., K.D. and J.Wa. performed Nimblegen array and sequencing studies. A.-M.P., J.M., J.D., S.V.E., M.A., N.Ka., J.We., M.-E.R., M.G., I.H. and A.T.H. recruited the human subjects and provided phenotypic information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.P. (constantin.polychronakos@mcgill.ca) or M.S.G. (mgerman@diabetes.ucsf.edu).

3

Author Queries

Journal: **Nature**

Paper: **nature08748**

Title: **Rfx6 directs islet formation and insulin production in mice and humans**

Query Reference	Query
1	AUTHOR: When you receive the PDF proofs, please check that the display items are as follows (doi:10.1038/nature08748): Figs 1 (black & white); 2–5 (colour); Tables: 1; Boxes: None. Please check all figures (and tables, if any) very carefully as they have been re-labelled, re-sized and adjusted to Nature's style. Please ensure that any error bars in the figures are defined in the figure legends.
2	Nature Proofreader: Please update/confirm the tentative publication date
3	Author: Juehu Wang or Jacques Weill?

Author: Please check the wording of the following statement, which will appear online only.

[Competing interests: M.S.G. is an inventor on patents held by the University of California covering Neurog3 and its use.]

For Nature office use only:

Layout	<input type="checkbox"/>	Figures/Tables/Boxes	<input type="checkbox"/>	References	<input type="checkbox"/>
DOI	<input type="checkbox"/>	Error bars	<input type="checkbox"/>	Supp info (if applicable)	<input type="checkbox"/>
Title	<input type="checkbox"/>	Colour	<input type="checkbox"/>	Acknowledgements	<input type="checkbox"/>
Authors	<input type="checkbox"/>	Text	<input type="checkbox"/>	Author contribs (if applicable)	<input type="checkbox"/>
Addresses	<input type="checkbox"/>	Methods (if applicable)	<input type="checkbox"/>	COI	<input type="checkbox"/>
First para	<input type="checkbox"/>	Received/Accepted	<input type="checkbox"/>	Correspondence	<input type="checkbox"/>
Display items	<input type="checkbox"/>	AOP (if applicable)	<input type="checkbox"/>	Author corr	<input type="checkbox"/>