The Company of Biologists

RESEARCH ARTICLE

Mitchell-Riley syndrome iPSCs exhibit reduced pancreatic endoderm differentiation due to a mutation in *RFX6*

Jamie Trott[1,‡], Yunus Alpagu[1,2,‡], Ee Kim Tan[1,3], Mohammad Shboul[1,4], Yousif Dawood[5,6], Michael Elsy[7], Heike Wollmann[8], Vincent Tano[3], Carine Bonnard[1], Shermaine Eng[1], Gunaseelan Narayanan[1], Seetanshu Junnarkar[1], Stephen Wearne[1], James Strutt[1], Aakash Kumar[1,3], Lucian B. Tomaz[3], Pierre-Alexis Goy[8], Slim Mzoughi[8], Rachel Jennings[7,9], Jaco Hagoort[5], Ascia Eskin[10,11], Hane Lee[10,11], Stanley F. Nelson[10,11,12], Fawaz Al-Kazaleh[13], Mohammad El-Khateeb[14], Rajaa Fathallah[14], Harsha Shah[15], Jonathan Goeke[16], Sarah R. Langley[3], Ernesto Guccione[8], Neil Hanley[7,9], Bernadette S. De Bakker[5], Bruno Reversade[1,8,17,18] and N. Ray Dunn[1,2,3,*]

ABSTRACT

Mitchell-Riley syndrome (MRS) is caused by recessive mutations in the regulatory factor X6 gene (*RFX6*) and is characterised by pancreatic hypoplasia and neonatal diabetes. To determine why individuals with MRS specifically lack pancreatic endocrine cells, we micro-CT imaged a 12-week-old foetus homozygous for the nonsense mutation *RFX6* c.1129C>T, which revealed loss of the pancreas body and tail. From this foetus, we derived iPSCs and show that differentiation of these cells *in vitro* proceeds normally until generation of pancreatic endoderm, which is significantly reduced. We additionally generated an *RFX6*[HA] reporter allele by gene targeting in wild-type H9 cells to precisely define RFX6 expression and in parallel performed *in situ* hybridisation for *RFX6* in the dorsal pancreatic bud of a Carnegie stage 14 human embryo. Both *in vitro* and *in vivo*, we find that RFX6 specifically labels a subset of PDX1-expressing pancreatic endoderm. In summary, *RFX6* is essential for efficient differentiation of pancreatic endoderm, and its absence in individuals with MRS specifically impairs formation of endocrine cells of the pancreas head and tail.

INTRODUCTION

Pancreas development begins with the emergence of the dorsal and ventral pancreatic buds from the posterior foregut (PFG) (Bastidas-Ponce et al., 2017). Progenitor cells in these buds proliferate extensively before giving rise to the endocrine, duct and acinar lineages. The exocrine pancreas is composed of acinar cells that produce digestive enzymes and duct cells that carry them to the digestive tract. Pancreatic endocrine cells reside in the islets of Langerhans and each secretes a different hormone involved in regulating blood glycaemia: β-cells (insulin), α-cells (glucagon), δ-cells (somatostatin), ε-cells (ghrelin) and PP cells (pancreatic polypeptide). Failure of the β-cells to secrete sufficient insulin results in persistent hyperglycaemia and diabetes. Type 1 and 2 diabetes account for the vast majority of cases and are the result of a complex interplay between numerous genetic and environmental factors. In contrast, maturity onset diabetes of the young (MODY) and neonatal diabetes are extremely rare (1-2% of individuals with diabetes) and are usually caused by mutations in single genes, typically encoding factors important for the development or function of β-cells (reviewed by Schwitzgebel, 2014). As well as aiding diagnosis and prognosis, studies of causative genes can shed light on the roles of these factors during normal development, as well as in more common forms of diabetes.

Mitchell-Riley syndrome (MRS) (https://www.omim.org/entry/615710) is caused by autosomal recessive mutations in the winged-helix transcription factor regulatory factor X6 gene (*RFX6*) (Smith et al., 2010). MRS is characterised by intrauterine growth retardation, atresia of the duodenum, atresia or hypoplasia of the gall bladder, annular or hypoplastic pancreas, and permanent neonatal diabetes (Mitchell et al., 2004; Smith et al., 2010; Spiegel et al., 2011). In severe cases, these symptoms may be accompanied by chronic diarrhoea, biliary defects, including cholestasis, and varying degrees of pancreatic exocrine deficiency, with death typically occurring within the first few years of life. The spectrum of congenital abnormalities observed in MRS suggests that either *RFX6* is important for patterning of the PFG, from which the

[1]Institute of Medical Biology, Agency for Science Technology and Research (A*STAR), 8A Biomedical Grove, #06-06 Immunos, 138648, Singapore. [2]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551. [3]Lee Kong Chian School of Medicine, Nanyang Technological University, Clinical Sciences Building, 11 Mandalay Road, 308232, Singapore. [4]Department of Medical Laboratory Sciences, Jordan University of Science and Technology, Irbid 2210, Jordan. [5]Department of Medical Biology, Section Clinical Anatomy and Embryology, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. [6]Department of Obstetrics and Gynaecology, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. [7]Faculty of Biology, Medicine & Health, University of Manchester, Oxford Road, Manchester M13 9PT, UK. [8]Institute of Molecular and Cellular Biology, Agency for Science Technology and Research (A*STAR), 61 Biopolis Drive, 138673, Singapore. [9]Endocrinology Department, Manchester University NHS Foundation Trust, Grafton Street, Manchester M13 9WU, UK. [10]Department of Human Genetics, David Geffen School of Medicine at UCLA, 695 Charles E. Young Drive South, Box 708822, Los Angeles, CA 90095-7088, USA. [11]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA. [12]Department of Pediatrics, UCLA Mattel Children's Hospital, Los Angeles, CA 90095, USA. [13]Department of Obstetrics and Gynecology, University of Jordan, Amman 19241, Jordan. [14]National Center for Diabetes, Endocrinology and Genetics, Amman 19241, Jordan. [15]Department of Obstetrics and Gynaecology, Queen Charlotte's & Chelsea Hospital, Imperial College London, Du Cane Road, London W12 0HS, UK. [16]Genome Institute of Singapore, Agency for Science Technology and Research (A*STAR), 60 Biopolis Street, 138672, Singapore. [17]Department of Paediatrics, National University of Singapore, Yong Loo Lin School of Medicine, 1E Kent Ridge Road, NUHS Tower Block, Level 12, 119228, Singapore. [18]Koç University School of Medicine, Medical Genetics Department, Istanbul 34450, Turkey.
‡These authors contributed equally to this work
*Author for correspondence (ray.dunn@ntu.edu.sg)

M.E., 0000-0001-5764-8930; H.W., 0000-0002-7666-8432; A.E., 0000-0002-3975-3734; B.R., 0000-0002-4070-7997; N.R.D., 0000-0002-6259-2975

DEVELOPMENT

1

affected tissues are derived, or during subsequent development and specialisation of these tissues. In the case of the pancreas, individuals with MRS do develop this organ but it contains varying quantities of functional exocrine tissue and very few chromogranin A-positive (CHGA+) endocrine cells (Mitchell et al., 2004). Together, these observations suggest that *RFX6* is required during development of the pancreas itself, and not during the initial patterning of the PFG.

Genetic studies in the mouse point to a similar but different role for the *Rfx6* gene in pancreas development. Like human patients, *Rfx6*-null mice show obstruction of the small intestine, variable degrees of pancreatic hypoplasia and premature death (Smith et al., 2010). Significantly, however, germline deletion of *Rfx6* or conditional inactivation of *Rfx6* in neurogenin 3-positive (Ngn3+) endocrine precursor cells does not affect the generation of Chga+ endocrine cells, although they do not contain insulin, glucagon or somatostatin (Smith et al., 2010; Piccand et al., 2014). Thus, *Rfx6* is not required for the formation of endocrine cells in the mouse, but only for their correct differentiation and consequently downstream function in maintaining homeostasis. This discrepancy highlights the need to specifically develop models to study RFX6 function in the human pancreas.

To achieve this goal, we derived induced pluripotent stem cells (iPSCs) from two individuals with MRS and their unaffected father.

Micro-CT imaging of one of these individuals confirmed the spectrum of congenital defects typical of MRS and exome sequencing identified a homozygous nonsense mutation in *RFX6*. MRS patient-derived iPSCs form a normal PFG, but exhibit reduced efficiency of differentiation into pancreatic endoderm. Aberrant differentiation is accompanied by expression of genes associated with mesoderm differentiation, indicating that RFX6 is crucial for maintaining the transcriptional program that specifies pancreatic endoderm. Finally, we demonstrate for the first time that *RFX6* is expressed in the nascent pancreatic bud of Carnegie stage 14 (CS14) human embryos.

RESULTS
A consanguineous family with MRS patients
We identified a consanguineous Syrian family in which three newborn children exhibited symptoms characteristic of MRS (Fig. 1A). The eldest two female siblings (III:1 and III:2) presented with neonatal diabetes, anaemia and intrauterine growth restriction, weighing 1.15 and 1.3 kg at birth (below the 2nd centile), respectively, and died within a few days of birth. Their stomachs, but not bowel, were distended by gas, consistent with congenital duodenal atresia. The third daughter (III:3) was born with the same symptoms as the eldest two and also died a few days after birth. In this case, further analysis failed to detect abnormalities of the brain, skull, face, heart, lungs or spine, and revealed a normal
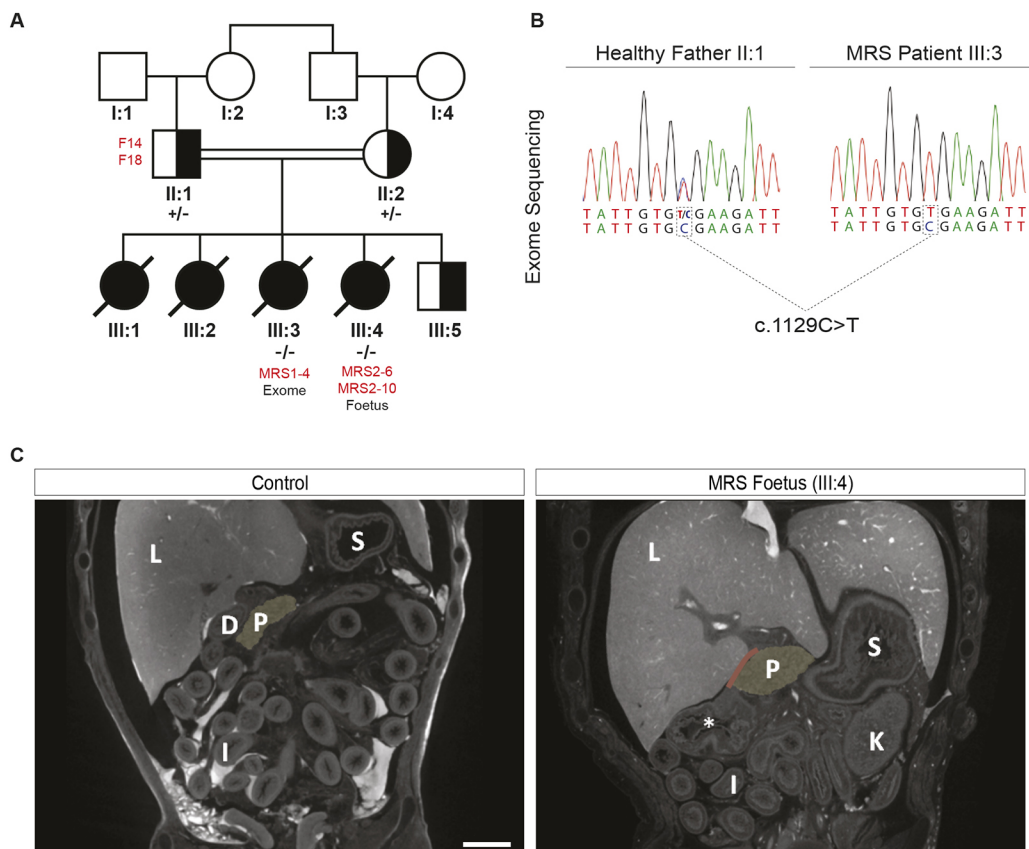
Fig. 1. A Syrian family with Mitchell-Riley syndrome (MRS) and a nonsense mutation in *RFX6*. (A) Pedigree of an inbred family from Syria. Black symbols indicate individuals with MRS. Crossed symbols indicate deceased individuals. Exome sequencing of individual III:3 revealed a homozygous nonsense mutation (−/−) in *RFX6* (c.1129C>T). The genotypes of sequenced individuals are shown below each symbol. Black and white symbols indicate a heterozygous carrier (+/−). Red text indicates the names of iPSC clones derived from the father and two affected individuals. Following termination of pregnancy III:4, the foetus was fixed and imaged. (B) Sanger sequencing for the nonsense mutation *RFX6* c.1129C>T. Healthy father II:1 is heterozygous C/T, while patient III:3 is homozygous T/T. (C) Coronal micro-CT images of a healthy control (left panel) and foetus III:4 (right panel). In foetus III:4, the pancreas head (highlighted in yellow) appears fully developed but the body and tail are absent. The asterisk highlights a distended distal duodenum/proximal jejunum. Intestine (I), pancreas (P), stomach (S), kidney (K), liver (L) and duodenum (D). See Movie 1 for the complete scan of foetus III:4. Scale bar: 2 cm.

karyotype with no evidence of structural abnormalities. Exome sequencing revealed a homozygous, nonsense mutation in *RFX6* (*RFX6* c.1129C>T) (Fig. 1B and described in detail below). During the fourth pregnancy (III:4), chorionic villus sampling determined that the foetus was homozygous for the same *RFX6* mutation as patient III:3. Based on this information, the pregnancy was electively terminated and the foetus preserved at the 12th week. The fifth and final pregnancy resulted in the birth of a healthy male child (III:5), who was found to be heterozygous for the *RFX6* c.1129C>T mutation. Skin fibroblasts were obtained from patients III:3 and III:4, and their healthy, heterozygous father (II:1), and reprogrammed into iPSC. All five iPSC lines were found to be karyotypically normal (Fig. S1A and data not shown).

To better understand at which stage pancreas development fails in individuals with MRS, we carefully chronicled at high spatial resolution the development of the pancreas in healthy human embryos. The pancreas is a composite organ derived from two primordia that arise from the ventral and dorsal sides of the distal foregut, commonly referred to as the ventral and dorsal pancreatic buds. These buds appear histologically as early as CS13, corresponding to 28-32 days post-conception (Fig. S2). Thereafter, the ventral pancreatic bud undergoes complex morphogenesis, moving towards the dorsal bud, with the two fusing at CS17/18 (42-48 days post-conception) to form the final template for the adult pancreas. The pancreatic head is formed from the ventral bud, and the body and tail mainly from the dorsal bud. Although similar

morphological changes have been shown in mouse, this is the first time movement of the buds and the exact timing of these events have been pinpointed in human embryos in an interactive three-dimensional fashion (see 3D interactive PDF in the supplementary information).

To visualise congenital defects due to loss of *RFX6*, we micro-CT scanned the 12-week foetus III:4 and a healthy control of similar age (Fig. 1C). The $RFX6^{-/-}$ foetus lacks the pancreas body and tail, with the maximal pancreas length roughly half that of the comparably staged wild-type foetus (3.81 mm $RFX6^{-/-}$ versus 7.29 mm wild type), lacks the gall bladder and additionally displays duodenal atresia (Fig. S1B). All detected abnormalities are in line with the spectrum of congenital defects typical of MRS (Smith et al., 2010; Spiegel et al., 2011 and https://www.omim.org/entry/615710). No abnormalities were observed in other abdominal organs (liver, spleen, kidneys, etc.) (Movie 1). We therefore propose that a defect specific to the dorsal pancreatic bud underlies the absence of the pancreas body and tail.

RFX6 c.1129C>T encodes a truncated protein p.Arg377Ter

In order to model human pancreatic development in individuals with MRS, we used a well-established, robust 12-day directed protocol adapted from Rezania et al. (2014) (schematised in Fig. 2A) to differentiate the control H9 human embryonic stem cell (hESC) line and iPSCs derived from patient III:4 and her carrier father into pancreatic progenitors. We have previously demonstrated
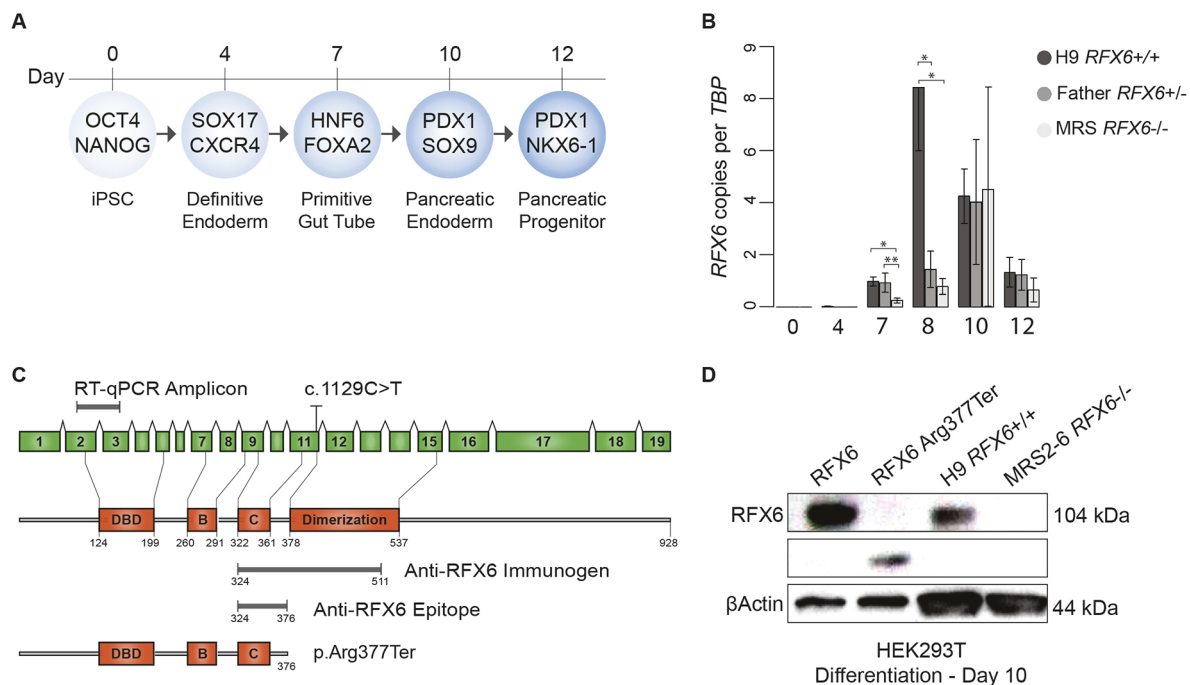
Fig. 2. Molecular characterisation of the *RFX6* c.1129C>T allele. (A) Schematic of the protocol used throughout this study to differentiate human pluripotent stem cells into pancreatic progenitor cells. Markers used to assess differentiation efficiency at key stages are indicated. (B) *RFX6* expression levels determined by RT-qPCR. Numbers on *x*-axis show days at which measurements were made. The number of copies per μl in each sample was determined by reference to a standard curve and plotted relative to the number of copies of the housekeeping gene *TBP*. Data are mean±s.e.m. of 3-6 independent differentiation experiments. Data for two father (F14 and F18) and two MRS (MRS2-6 and 2-10) iPSC lines, as well as wild-type H9 cells are included. Student's *t*-test (two-sided) was used to calculate *P*-values for differences in the mean of each group. *P<0.05, **P<0.01. (C) The human *RFX6* gene and the protein it encodes. The RFX6 protein contains a DNA-binding domain (DBD) and three additional domains involved in dimerisation. The *RFX6* c.1129C>T transcript is predicted to produce a truncated protein (p.Arg377Ter). The amplified PCR product quantified in B, the immunogen sequence used to generate the RFX6 antibody used in D and the range over which the epitope is found are indicated in the diagram. (D) Detection of full-length and p.Arg377Ter RFX6 by western blotting. Samples were harvested from HEK293T cells transfected with expression constructs encoding full-length (104 kDa) and p.Arg377Ter (44 kDa) RFX6 proteins (15 μg protein per lane), and from H9 and MRS2-6 day 10 differentiation cultures (50 μg protein per lane). Failure to detect p.Arg377Ter RFX6 in MRS2-6 cells indicates that the protein is not produced or is rapidly degraded.

that this platform yields near-homogenous definitive endoderm by day 4 (SOX17+;CXCR4+), followed by PGT (HNF6+;FOXA2+) on day 7, pancreatic endoderm (PDX1+;SOX9+) on day 10 and pancreatic progenitors (PDX1+;NKX6-1+) on day 12, even when applied to genetically diverse human pluripotent stem cells (hPSC) (Trott et al., 2017). To investigate the consequences of the *RFX6* c.1129C>T mutation at the mRNA level, we first documented *RFX6* expression over the course of 12 days of *in vitro* differentiation. Quantitative RT-PCR reveals that *RFX6* expression peaks on day 8 of differentiation in wild-type H9 cells (Fig. 2B) – a result consistent with the finding that *RFX6* is expressed in the dorsal pancreatic bud at CS12-14, coinciding with the emergence of pancreatic endoderm (Jennings et al., 2017). In contrast, iPSCs from both patient III:4 and her father express *RFX6,* but at levels significantly lower than in H9 cells on day 8 (Fig. 2B).

The RFX6 protein consists of four functional domains: a DNA-binding domain plus three domains that mediate dimerisation (Fig. 2C) (Aftab et al., 2008). The *RFX6* c.1129C>T mutation is predicted to be a nonsense mutation that produces a truncated protein (p.Arg377Ter) lacking the core C-terminal dimerisation domain. To determine whether the *RFX6* c.1129C>T allele generates a partial protein product, we measured the levels of wild-type and p.Arg377Ter RFX6 by western blotting (Fig. 2D). We failed to detect any RFX6 protein species in MRS2-6 on day 10 of differentiation, which provides strong evidence that c.1129C>T is a bona fide *RFX6* loss-of-function allele. Hereafter, individuals homozygous for the c.1129C>T mutation are referred to as *RFX6*−/− and carriers as *RFX6*+/−.

RFX6 is required for efficient differentiation of pancreatic endoderm

To determine whether the MRS phenotype is reproduced *in vitro*, we compared gene expression between cells derived from healthy subjects and individuals with MRS. As anticipated from the clinical presentation of MRS, patient-derived iPSCs were able to generate definitive endoderm and PGT normally (Fig. 3A,B). However, beginning on day 8, loss of *RFX6* resulted in significantly reduced expression of *PDX1* and *SOX9*, the co-expression of which identifies the multipotent, proliferative progenitors of all three pancreatic lineages (reviewed in Kawaguchi, 2013). Indeed, colocalisation of PDX1 and SOX9 is near ubiquitous in H9 cells on day 10 of differentiation, but in *RFX6*−/− iPSCs, large areas of PDX1−;SOX9− cells are observed, suggesting an overt loss of pancreatic identity (Fig. 3C).

Co-expression of the transcription factors PDX1 and NKX6-1 defines a population of multipotent pancreatic progenitors poised to differentiate into endocrine cells (Russ et al., 2015). We therefore sought to quantify the effects of *RFX6* loss on these pancreatic progenitors. The differentiation protocol that we employ typically generates >80% PDX1+;NKX6-1+ pancreatic progenitors (Figs. 3D,E and Trott et al., 2017). Loss of a single copy of *RFX6* substantially reduces the proportion of NKX6-1+ cells, leaving large numbers of singly PDX1-positive cells, whereas loss of both copies further reduces the proportion of both PDX1+ and NKX6-1+ cells (Fig. 3D,E). This suggests that *RFX6* is not absolutely required to generate PDX1+ pancreatic endoderm or endocrine-competent PDX1+;NKX6-1+ progenitor cells, but that it controls the efficiency of differentiation in a dose-sensitive manner. These observations are consistent with the MRS phenotype in which loss of *RFX6* results in the absence of CHGA+ endocrine cells and pancreatic hypoplasia, but exocrine tissue still forms.

RFX6 maintains the transcriptional program that defines pancreatic endoderm

To elaborate the requirement for RFX6 to form pancreatic endoderm, we compared the transcriptomes of cells from patient III:4 and her heterozygous father on day 8 of differentiation, which coincides with peak *RFX6* expression and the onset of changes in gene expression (Figs 2B and 3A). Principal component analysis revealed that changes in gene expression were largely due to differences between the affected and unaffected individuals, rather than between iPSC lines or biological replicates (Fig. 4A). Lack of *RFX6* leads to downregulation of 88 genes and upregulation of 309 genes (Fig. 4B; Tables S1 and S2). Analysis of gene ontology terms associated with genes whose expression depends on RFX6 identified a single enriched term, 'Pancreas development' (Fig. 4C). Conversely, genes upregulated in the absence of RFX6 were associated with development of mesoderm derivatives: specifically, heart, skeleton and muscle (Fig. 4C). Interestingly, these genes were often involved in morphogenesis or formation of the extracellular matrix. These observations suggest that RFX6 has dual functions: to initiate expression of genes required for pancreatic development and to maintain pancreatic identity by blocking activation of genes that direct formation of other lineages.

To characterise the effects of RFX6 function on the pancreatic transcriptional program, we analysed the expression of genes with known roles in mammalian pancreas development (Fig. 4D). Several genes known to be expressed in pancreatic endoderm were strongly downregulated in MRS cells (e.g. *PDX1*, *SOX9*, *PROX1*, *ONECUT1* and *ONECUT2*), alongside the endocrine markers *NEUROG3* and *INSM1*. Conversely, genes with roles in both definitive and pancreatic endoderm development (*GATA4/6*, *FOXA2* and *HNF1A*, etc.) tended to be unaffected by loss of *RFX6*. Furthermore, although *SOX17* expression is normally downregulated following PGT differentiation, *SOX17* levels remained elevated at day 8 in MRS cells, suggesting a delay in further differentiation in the absence of RFX6. We next examined the expression of genes that mark the CS12-14 dorsal pancreatic bud, but not the nearby hepatic cords (Fig. 4E) (Jennings et al., 2017). Of these genes, 15 are significantly downregulated in MRS patient cells: *PDX1*, *SOX9*, *NETO1*, *ONECUT1*, *NPY1R*, *FAM84A* (*LRATD1*), *PRDM16*, *RAP1GAP2*, *VWA5B2*, *RP11-351J23.1*, *GOLGA6L9*, *PAQR5*, *NRG3*, *MST1L* and *TLE2*. When taken together, these observations suggest that RFX6 is specifically required for PFG cells to differentiate into pancreatic;endoderm and identifies several candidates that influence pancreatic endoderm differentiation downstream of RFX6.

Using CRISPR/Cas9 gene editing, we introduced a triplicated HA epitope into the 3′ end of the *RFX6* gene in H9 cells (designated *RFX6*HA/HA) (Fig. 5A, Fig. S3A-D). Pluripotent *RFX6*HA/HA cells were then differentiated into day 8 pancreatic progenitors, and endogenous RFX6-HA detected by immunofluorescence (Fig. S3E). Importantly, we observed no effect of the HA tag on the production of pancreatic progenitors, indicating that the addition of the HA tag does not interfere with RFX6 function (Fig. 5B). To eliminate background signal caused by non-specific antibody binding, a control experiment using wild-type H9 hESC was performed in parallel.

To identify direct transcriptional targets of RFX6, we performed chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq). We identified 1350 regions bound by RFX6 that were not enriched in control experiments using wild-type H9 cells (Fig. 5C; Tables S3 and S4). These sites were typically located more than 1 kb away from the transcription start sites of the
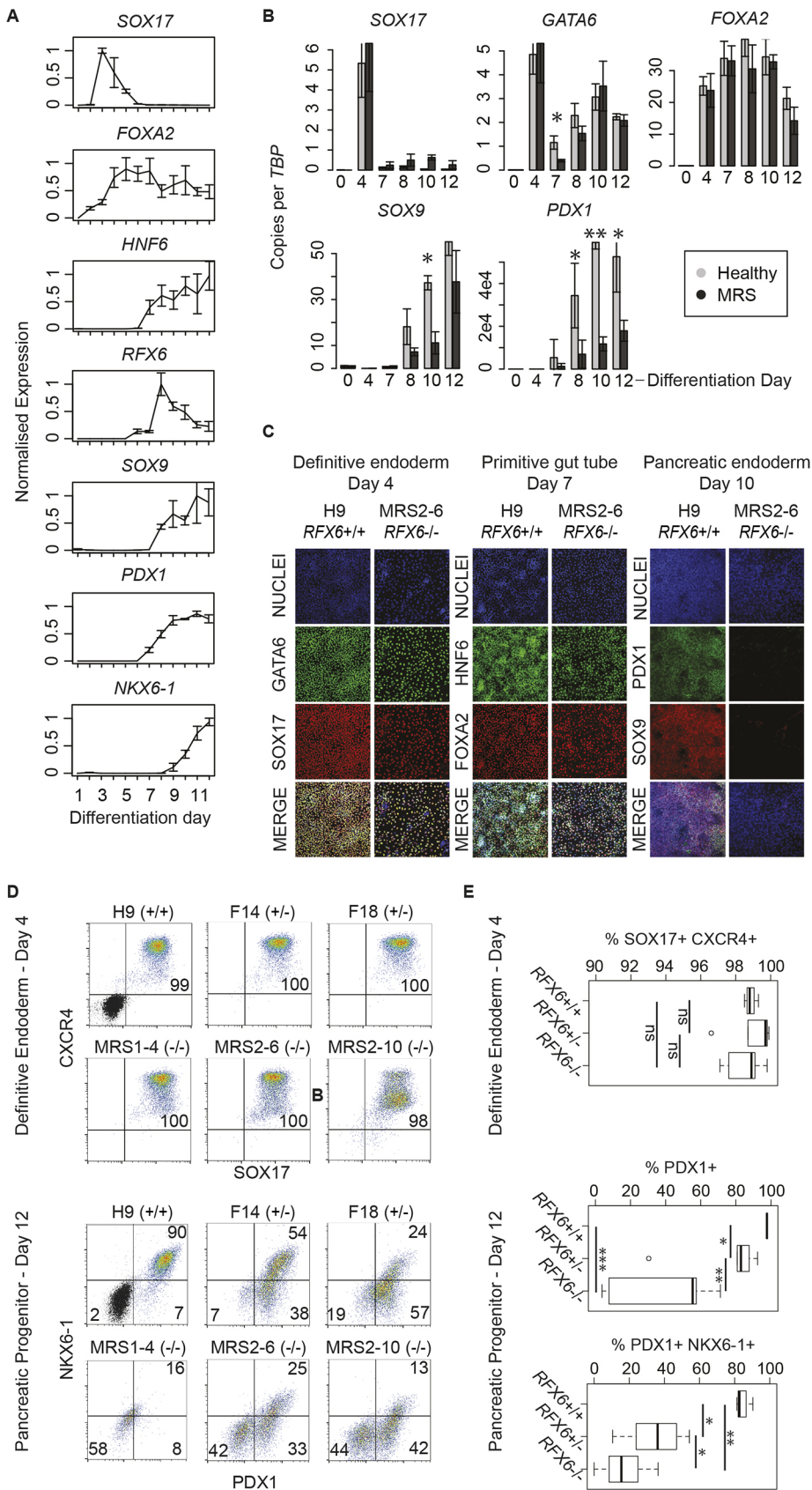
Fig. 3. Loss of *RFX6* reduces the efficiency of pancreatic differentiation. (A) Expression of *RFX6* and markers of definitive endoderm (*SOX17*), primitive gut tube (*FOXA2* and *HNF6*), pancreatic endoderm (*PDX1* and *SOX9*) and pancreatic progenitors (*NKX6-1*) measured daily during differentiation of H9 cells. The number of copies per μl in each sample was determined by reference to a standard curve and plotted relative to the number of copies of the housekeeping gene *TBP*. Data are mean±s.e.m. of three independent differentiation experiments. (B) Comparison of gene expression during differentiation of pluripotent cells from healthy subjects (H9 hESC and F14 and 18 iPSC) and MRS patients (MRS1-4, 2-6 and 2-10 iPSC). Three independent differentiation experiments were conducted using each cell line and the mean number of copies of each gene per copy of *TBP* at each time point was calculated as in A. Data represent the mean of these values for the three different cell lines representing healthy or MRS patients±s.e.m. Student's *t*-test (one-sided) was used to calculate *P*-values for differences in the mean between groups. *P<0.05, **P<0.01. (C) Immunofluorescence staining for the indicated transcription factors on days 4, 7 and 10 of pancreatic differentiation. (D) Flow cytometric detection of markers of definitive endoderm (SOX17 and CXCR4) and pancreatic progenitors (PDX1 and NKX6-1) after 4 and 12 days of differentiation, respectively. Cells stained with isotype control antibodies used to gate positive cells are overlaid in black. Numbers represent the percentage of cells in each quadrant. The data are taken from a single representative experiment. (E) Boxplots showing percentages of cells staining positive for markers of definitive endoderm (SOX17 and CXCR4) and pancreatic progenitors (PDX1 and NKX6-1), as determined by flow cytometry. Differentiation data were grouped according to genotype as follows: *RFX6*[+/+] (H9), *RFX6*[+/−] (F14, F18) and *RFX6*[−/−] (MRS1-4, MRS2-6 and MRS2-10). Three independent experiments were conducted for each cell line. The median percentage for each group is shown by a horizontal bar. Boxes and whiskers indicate the lower and upper quartiles, and minimum and maximum values (excluding outliers), respectively. Unfilled circles indicate outliers. Student's *t*-test (one-sided) was used to calculate *P*-values for differences in the mean of each group. *P<0.05, **P<0.01, ***P<0.001.

nearest gene, in introns or intergenic regions (Fig. 5D,E), suggesting RFX6 does not usually act by binding to gene promoters. Transcription factor target motif analysis identified an RFX6 target binding sequence that is similar to, yet distinct from, sequences bound by RFX1 and RFX3 (Fig. 5F). In order to determine which differentially expressed genes are direct targets of RFX6, we
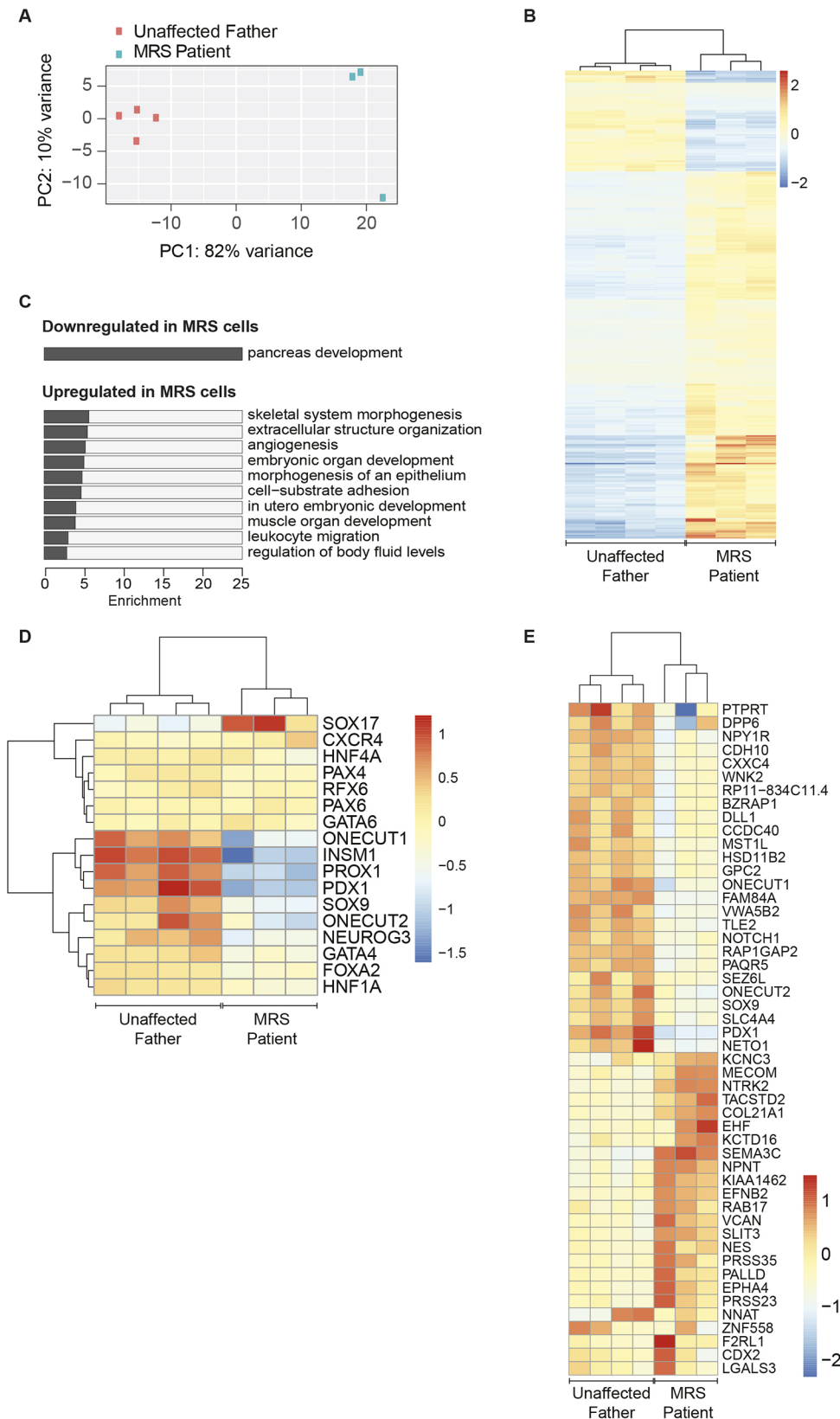
Fig. 4. Comparative transcriptomics between pancreatic endoderm differentiated from MRS patient and unaffected father iPSCs. (A) Four independent experiments using two different father iPSC lines were compared with three independent experiments using two different MRS patient lines on day 8 of pancreatic differentiation. Principal component analysis shows that pancreatic endoderm derived from healthy father iPSCs are clearly separated from MRS patient cells along axis PC1, which accounts for 82% of the variance between samples. Data points represent independent differentiation experiments. (B) Differentially expressed genes between healthy father and MRS patient cells (log2 fold-change >1 and adjusted *P* value <0.01). (C) Biological process gene ontology (GO) terms associated with genes plotted in B. GO terms shown are those associated with more than five genes with adjusted *P*<0.05, after removal of redundant terms. (D) Expression of signature pancreatic marker genes between the father and the MRS patient. (E) Expression of the top 50 most variably expressed genes in the dorsal pancreas of pooled CS12-14 human embryos, but not in the nearby hepatic cords (Jennings et al., 2017), between the father and the MRS patient.

identified those that are closest to an RFX6-binding site. Of 53 putative direct targets of RFX6, in the absence of *RFX6*, 10 genes are downregulated (*HHEX*, *CDHR3*, *ATOH1*, *RBFOX1*, *CPM*, *CASC18*, *BCHE*, *RALYL*, *RAB3C* and *KCNJ3*), while 43 are upregulated (Table S5). Interestingly, RFX6 is present at distal 3′ enhancers of *HHEX* and *ISL1* (Fig. S4), with the murine *Hhex* and *Isl1* homologs well known for their crucial roles in early pancreatic morphogenesis in mice (Ahlgren et al., 1997; Bort et al., 2004).
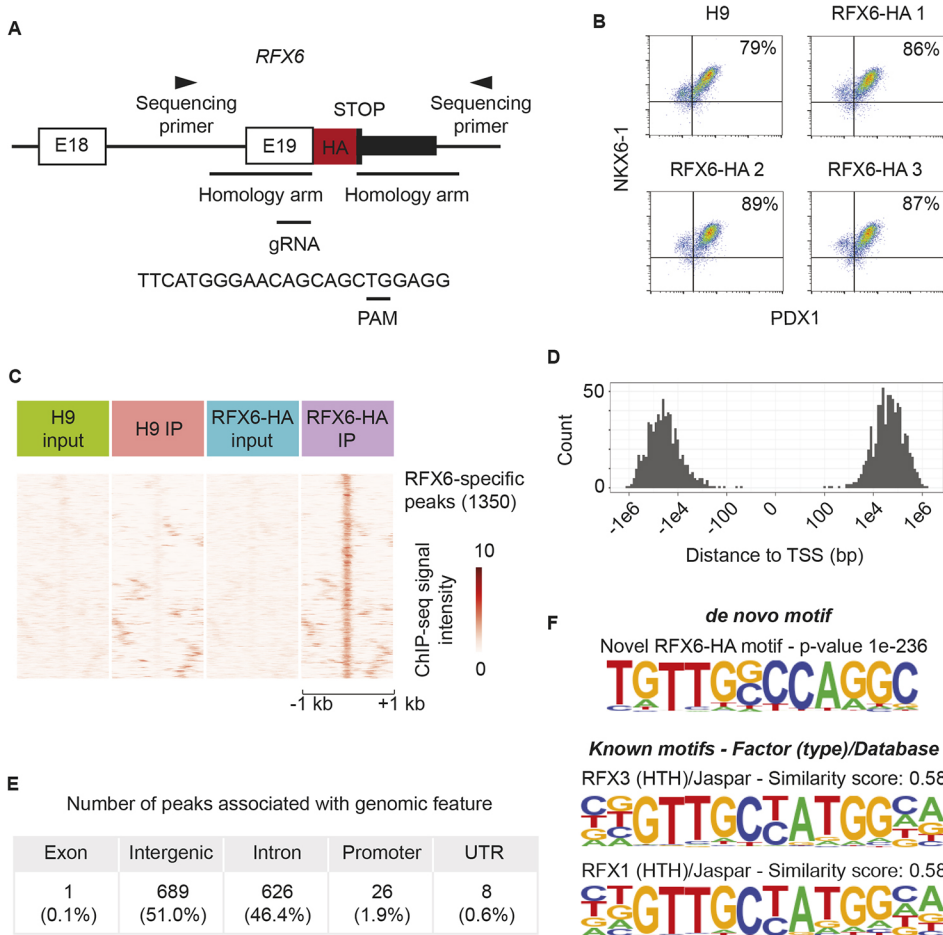
Fig. 5. Genome-wide chromatin-binding analysis of endogenous HA-tagged RFX6 protein by ChIP-seq. (A) Schematic of the strategy used to introduce a triplicated HA epitope into exon 19 of the *RFX6* locus using CRISPR-Cas9 gene editing. H9 hESC clones were recovered in which both alleles underwent homologous recombination, yielding *RFX6$^{HA/HA}$* hESC. (B) Intracellular flow cytometric analysis for PDX1 and NKX6-1 in H9 cells and three *RFX6$^{HA/HA}$* subclones on day 12 of differentiation. (C) ChIP signal heatmap of DNA-binding enrichment in RFX6-HA-specific peaks. (D) Distances between RFX6-HA-specific peaks to the nearest transcriptional start site (TSS). (E) Number of RFX6-HA-specific peaks associated with annotated genomic features in the HOMER database. (F) Over-represented *de novo* motif identified in RFX6-HA-specific peaks with matches to known RFX family target motifs in the JASPAR database (RFX1 and RFX3).

## Pancreatic endoderm cells lacking *RFX6* arise during normal development

iPSCs from individuals with MRS are capable of forming both pancreatic endoderm and PDX1$^+$;NKX6-1$^+$ pancreatic progenitors, although loss of RFX6 substantially reduces the efficiency of differentiation (Fig. 3B-E). This suggests that *RFX6* is not absolutely required for pancreatic endoderm differentiation, but affects the efficiency of the process. However, it is unclear whether pancreatic endoderm cells lacking *RFX6* expression are observed during normal development, or whether all PDX1+ pancreatic endoderm cells express RFX6. Given the lack of reliable antibodies that recognise human RFX6, we performed immunofluorescence staining using the *RFX6$^{HA/HA}$* H9 hESC line at different stages of pancreatic differentiation (Fig. 6A and Fig. S5A). Differentiation cultures composed largely of PDX1$^+$ pancreatic endoderm contain a mixture of RFX6-HA$^+$ and RFX6-HA$^-$ cells. Specifically, 30% of PDX1$^+$ pancreatic endoderm cells lack RFX6-HA protein (Fig. 6B). However, PDX1$^-$;RFX6-HA$^+$ cells are seldom observed.

To determine whether RFX6$^-$ pancreatic endoderm also arises *in vivo*, we performed high resolution *in situ* hybridisation for *RFX6* on human embryos. Between CS12 and CS19, *PDX1* is expressed ubiquitously in the multipotent progenitor cells of the pancreatic bud (Fig. S5B-D and Jennings et al., 2013). As observed *in vitro*, *RFX6* is expressed in the majority of *PDX1$^+$* cells but *RFX6$^-$* cells are present throughout the nascent pancreatic bud at CS14 and later in the fused bud at CS17 (Fig. 6C). A similar pattern is observed at 8 weeks post-conception (8 wpc), around the time the endocrine lineage becomes segregated from ductal cells (Fig. 6C; Fig. S5)

(Jennings et al., 2015). These observations are consistent with studies in mice, in which PDX1$^+$RFX6$^-$ cells are observed throughout the embryonic day 10 dorsal and ventral pancreatic buds (Smith et al., 2010; Soyer et al., 2010). When taken together, these observations suggest that human pancreatic endoderm consists of PDX1$^+$ cells with and without RFX6 expression.

DISCUSSION

Although the mouse and human pancreas function similarly, with isolated human islets able to impart normoglycemia in diabetic mice, there are prominent anatomical, transcriptional and physiological differences between the two species. The developmental origins of these differences have been difficult to interrogate due in part to the inaccessibility of human foetal material. Our present study begins to fill this knowledge gap. Using micro-CT imaging, we chronicled human pancreatic development between CS13 and CS23, corresponding to roughly 30 days of gestation, with eight intact human embryonic specimens. From these images, sequential three-dimensional reconstructions were assembled that provide high-resolution temporal and spatial insight into the complex growth and morphogenesis of the dorsal and ventral pancreatic buds that coalesce to form the pancreatic anlagen (Fig. S1 and see 3D interactive PDF in the supplementary information). This timeline further allowed us to better understand the embryological consequences of homozygous *RFX6* loss. At 12 weeks, the *RFX6$^{-/-}$* foetus (III:4) lacks the pancreas body and tail – a phenotype that presages the overt pancreatic hypoplasia and neonatal diabetes observed in patients III:1, III:2 and III:3 (Fig. 1A;
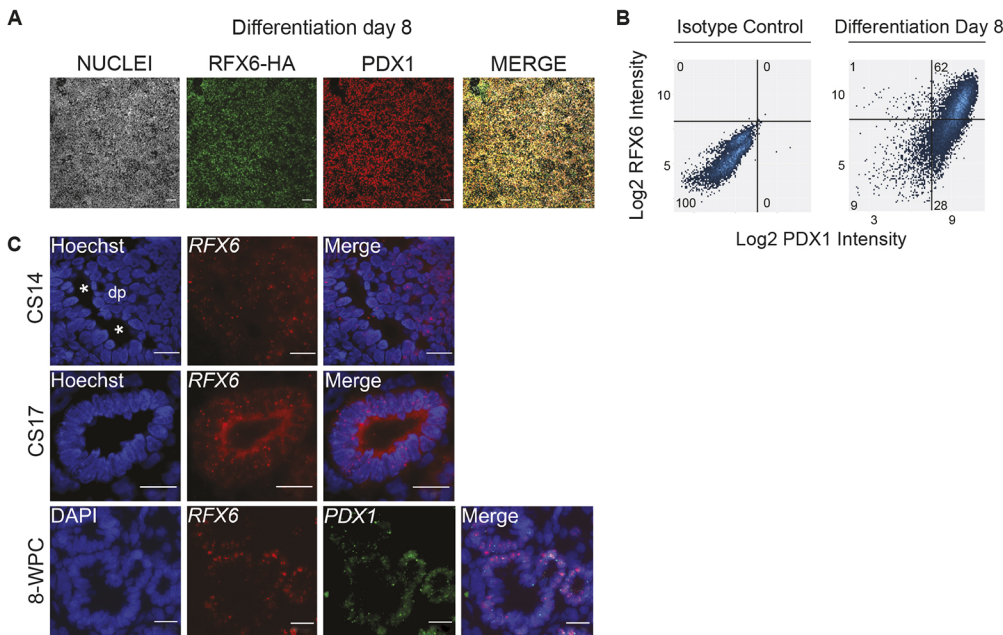
Fig. 6. *RFX6* expression in differentiation cultures and human embryos. (A) Immunofluorescence staining in *RFX6^HA/HA* cells for HA and PDX1 on day 8 of differentiation. (B) Quantification of immunofluorescence staining in A at single cell resolution. Cells were segmented and staining intensities extracted with FIJI. The plotted data are aggregated from three independent experiments. (C) Sagittal sections through the dorsal pancreatic bud of a human embryo at CS14; sagittal sections through the pancreas of a human embryo at CS17 and foetus at 8 wpc, after RNAScope *in situ* hybridisation staining for *RFX6* (red) and *PDX1* (green). dp, dorsal pancreatic bud; asterisk represents microlumen. Scale bars: 20 μm in C; 100 μm in A.

Fig. S1B and Movie 1). In contrast, Smith et al. (2010) reported that the major defect in *Rfx6*-deficient mice – i.e. the presence of Chga$^+$ endocrine cells that lack hormones – is confined to the islets, with only incompletely penetrant, but poorly described hypoplasia (Smith et al., 2010). This apparent species-specific difference in the role of *Rfx6/RFX6* during mammalian pancreatic development motivated our derivation of iPSC from patients III:3 and III:4 in an effort to model the embryological and molecular origins of MRS *in vitro*.

We deployed a robust pancreatic differentiation protocol based on the work of Rezania et al. (2014) to identify the specific stage at which *RFX6* loss results in the prominent pancreatic hypoplasia observed in the MRS pedigree (Fig. 1A) (Rezania et al., 2014). Molecular analyses at the mRNA and protein level demonstrate that the formation of both the definitive endoderm as well as early PFG proceeds normally. However, upon the provision of signals that impart pancreatic identity (e.g. retinoic acid), *RFX6*-deficient cells fail to robustly activate the cardinal pancreatic progenitor marker *PDX1*. This result was also independently observed by Zhu et al. in H1 cells in which *RFX6* loss-of-function alleles were engineered using CRISPR/Cas9 gene editing of exon 3 (Zhu et al., 2016). Considering that *PDX1* and *RFX6* are activated almost synchronously, the simplest explanation for this result is that RFX6 directly regulates *PDX1* expression. However, our RFX6-HA ChIP-seq did not identify PDX1 as a potential RFX6 target (Fig. 5.; Table S5). A brief report by Cheng et al. (2019) who performed RFX6 ChIP-seq on the adult mouse pancreas did show slight enrichment within one region of the *PDX1* promoter (Cheng et al., 2019). *In silico* analysis with MatInspector predicts three potential X-box motifs with modest matrix similarity scores within the *PDX1* gene. Further experiments are therefore necessary to establish whether PDX1 and RFX6 form a feedback loop during early human pancreatic development.

Much of our understanding of the gene regulatory networks governing human pancreatic development comes from studies in the mouse. For example, it was established more than 25 years ago that *Pdx1* is absolutely required for the formation of the murine pancreas, as Pdx1-deficient neonates entirely lack all three pancreatic lineages and succumb to extreme hyperglycaemia within

a few days of birth (Ahlgren et al., 1996; Jonsson et al., 1994; Offield et al., 1996). Rare individuals with homozygous or compound heterozygous *PDX1* mutations were subsequently identified who also displayed pancreatic agenesis (Stoffers et al., 1997; https://www.omim.org/entry/600733). The MRS pancreatic phenotype is less severe than loss of *PDX1*, hypoplasia versus agenesis, which suggests a simple epistatic relationship in which *PDX1* is upstream of *RFX6*. In support of this hypothesis, we and others previously identified PDX1 binding sites in the human *RFX6* promoter using ChIP-seq (Teo et al., 2015; Wang et al., 2018), and *PDX1* and *RFX6* display similar expression kinetics between days 7 and 8 of directed differentiation into the pancreatic lineage (Fig. 3A). Interestingly, however, not every PDX1$^+$ cell co-expresses RFX6 both *in vitro* and *in vivo* (Fig. 6A-C). How this segregation is achieved developmentally and *in vitro* is at present unknown, but this observation suggests that PDX1$^+$;RFX6$^+$ and PDX1$^+$;RFX6$^-$ cells give rise to distinct lineages (or sub-populations) in the adult pancreas. Loss of *RFX6 in vitro* results in an approximately fourfold reduction in PDX1$^+$;NKX6-1$^+$ pancreatic progenitors (Fig. 3D,E), a result consistent with the rare CHGA$^+$ cells observed post-mortem in individuals with MRS (Mitchell et al., 2004). It is, however, important to emphasise that, although reduced, acinar and ductal tissue is present in individuals with MRS (Mitchell et al., 2004). This raises the possibility that PDX1$^+$;RFX6$^-$ cells principally contribute to the non-endocrine component of the adult pancreas. In addition, our micro-CT imaging of the 12-week-old *RFX6^{-/-}* foetus reveals the presence of the pancreas head, which is derived from the ventral pancreatic bud, but the absence of the body and tail, which is derived from the dorsal pancreatic bud (Fig. 1C). These pathological observations raise an alternative possibility whereby PDX1$^+$;RFX6$^-$ cells and their descendants contribute predominantly to the pancreas head, while PDX1$^+$RFX6$^+$ cells form the body and tail. These hypotheses could be tested either by the generation of sophisticated reporter alleles in pluripotent cells (Bao et al., 2019) coupled with extended *in vitro* differentiation into regionally patterned pancreatic organoids or by conventional lineage tracing in mice because, as in human, not all Pdx1$^+$ cells within the embryonic day 10 pancreatic bud are Rfx6$^+$ (Smith et al., 2010; Soyer et al., 2010). Last, the number of PDX1$^+$;

NKX6-1$^+$ pancreatic progenitors is sensitive to *RFX6* gene dose *in vitro* (Fig. 3D,E). Thus, one simple explanation for the recent report of MODY in three generations of a Japanese family harbouring the same *RFX6* allele described here (c.1129C>T) is a reduction in β-cell mass (Akiba et al., 2019).

Although studies of the RFX family of transcription factors during embryonic development are limited, RFX6 was previously shown *in vitro* to bind X-box sequences to activate expression of the insulin gene (*INS*) in the human β-cell line Endo-C-βH2, as well as the glucokinase (Gck) and ATP-binding cassette subfamily C member 8 (*Abcc8*, also known as *Sur1*) genes in the Min6 mouse insulinoma cell line (Chandra et al., 2014; Piccand et al., 2014). However, RFX6 is unable to activate transcription directly, as it lacks an activation domain (Aftab et al., 2008), but has been shown to dimerise with RFX2 and RFX3, which do contain such domains (Rual et al., 2005; Smith et al., 2010). Rfx3 also binds the *Gck* promoter in Min6 cells (Ait-Lounis et al., 2010). These *in vitro* findings suggest that Rfx3 and Rfx6 commonly co-regulate target genes. Additional support for this relationship comes from genetic studies in mouse: the phenotype of *Rfx3* null mutant mice mirrors *Rfx6* loss, although less severe, as both mutants produce endocrine cells that fail to express hormones (Ait-Lounis et al., 2010, 2007; Piccand et al., 2014; Smith et al., 2010). Significantly, *RFX2* and *RFX3* are co-expressed in the developing human pancreas (Cebola et al., 2015; Jennings et al., 2017) and during pancreatic differentiation *in vitro* (Fig. S6). PDX1 ChIP-seq studies also confirm that *RFX3*, like *RFX6*, is a PDX1 target gene during pancreatic differentiation *in vitro* (Wang et al., 2018). From these observations, one simple model emerges whereby specification of pancreatic identity in the developing human embryo – i.e. the onset of *PDX1* transcription – is tightly followed by the activation of *RFX2*, *RFX3* and *RFX6*, which form transcriptional complexes that activate (or repress) target genes that orchestrate the development of the pancreas body and tail.

We identified two such targets bioinformatically – the transcription factor genes *ISL1* and *HHEX* – by combining comparative transcriptomics and ChIP-seq (Fig. 5 and Fig. S4). In the mouse, homozygous inactivation of *Isl1* results in dorsal pancreatic bud agenesis (Ahlgren et al., 1997), whereas loss of *Hhex* leads to a complete failure of ventral pancreatic specification (Bort et al., 2004). Both mutations are lethal during mid-gestation, with pancreatic phenotypes much more severe than loss of *Rfx6* (Piccand et al., 2014; Smith et al., 2010). These genetic studies therefore suggest that Rfx6 is not absolutely required for the early transcriptional activation of *Isl1* and *Hhex* in the mouse, even though these three genes show overlapping expression patterns during early pancreatic development. This observation perhaps reflects the temporally distinct requirements between mouse *Rfx6* and human *RFX6* for pancreatic development emphasised by this study and previous work describing significant transcriptional differences between the mouse and human pancreas (Jennings et al., 2017). Whether human RFX6 directly regulates *ISL1* and *HHEX* is also an unresolved issue, which can be addressed with *in vitro* reporter assays. To our knowledge, human pluripotent cell lines lacking either *ISL1* or *HHEX* have not been generated. It would therefore be interesting to derive such lines and subject them to pancreatic differentiation and to assess their phenotypes in comparison with loss of RFX6. Comparative RNA-seq and ChIP-seq studies like those described here at key stages of *in vitro* differentiation for ISL1 and HHEX would additionally contribute to the construction of the regulatory 'wiring diagram' underlying early human pancreatic development.

Last, RFX6 is expressed in the insulin-secreting β-cells of the adult human islet (Chandra et al., 2014). Interestingly, recent genome-wide association studies for type 2 diabetes identified loci that are specifically enriched in RFX-binding motifs (Varshney et al., 2017). Our *RFX6^{HA}* reporter allele coupled with extended differentiation protocols that efficiently produce 'β-like' cells *in vitro* (e.g. Nair et al., 2019) will provide a platform to formally identify those disease loci bound by RFX6. In the future, we intend to exploit *RFX6$^{-/-}$* iPSC and the RFX6-HA-tagged line to further elaborate the function of RFX6, both at the level of the transcriptome and the proteome, in other endodermal lineages, including the lung and intestine, where its role is poorly described.

MATERIALS AND METHODS

Human subjects

Saliva samples were collected from parents and siblings (II:1, II:2 and III:5), and skin tissues from one affected daughter (III:3) and one affected foetus (III:4). Chorionic villus sampling was performed for prenatal diagnosis when the mother was pregnant with III:4. A skin biopsy was also collected from the healthy father (II:1). Genomic DNA (gDNA) was extracted using DNeasy Blood sample and Tissue Kits (Qiagen). All human biological materials were obtained after the parents gave their informed written consent (obtained and witnessed by co-authors Drs Fawaz Al-Kazaleh and Mohammad Shboul, respectively) and the local ethics commission gave its approval. The derivation of iPSC from the pedigree in Fig. 1A and micro-CT imaging of the CS14 foetus (III:4) are described in detail in the supplementary Materials and Methods.

Whole-exome sequencing

For proband (III:3) whole-exome sequencing, 1 μg of purified genomic DNA was subjected to exome capture using Illumina TruSeq Exome Enrichment Kit (Illumina). Illumina HiSeq2500 high output mode was used for sequencing as 100 bp paired-end runs at the UCLA Clinical Genomics Centre. Sequence reads were aligned to the human reference genome (Human GRCh37/hg19 build) using Novoalign (v2.07, http://www. novocraft.com/main/index.php). PCR duplicates were identified by Picard (v1.42, http://picard.sourceforge.net/) and Genome Analysis Toolkit (v1.1, http://www.broadinstitute.org/gatk/) (McKenna et al., 2010) was used to re-align indels, recalibrate the quality scores, and call, filter, recalibrate and evaluate the variants. SNVs and INDELs across the sequenced protein-coding regions and flanking junctions were annotated using Variant Annotator X (VAX), a customised Ensembl Variant Effect Predictor (Yourshaw et al., 2015). Each variant was annotated with information including gene names and accession numbers, reference variant, variant consequences, protein positions and amino-acid changes, conservation scores, population minor allele frequencies and expression pattern, as well as PolyPhen2 (Adzhubei et al., 2010), SIFT (Kumar et al., 2009) and Condel (González-Pérez and López-Bigas, 2011) predictions. Almost 20,000 variants were identified across the RefSeq protein-coding exons and flanking introns (±2 bp). Of these, 37 homozygous and 119 compound heterozygous variants were protein-changing variants with population minor allele frequencies of less than 1%. Out of the 37 homozygous variants, only three passed quality filters, were included in regions of homozygosity, were not seen in population and were predicted to be deleterious, including one stop-gained mutation in the *RFX6* gene.

Cell culture and differentiation

Human pluripotent stem cells were seeded onto Matrigel (Corning, 354277)-coated plates and cultured in mTeSR 1 medium (STEMCELL Technologies, 85850). Cells were passaged at a 1:20-1:25 split ratio every 4 days. For passaging, Gentle Cell Dissociation Reagent (STEMCELL Technologies, 07174) or ReLeSR (STEMCELL Technologies, 05872) was used. For differentiation into pancreatic progenitors, we employed the STEMdiff Pancreatic Progenitor kit (STEMCELL Technologies, 05120) with the following modifications: (1) cells were initially seeded onto 12-well plates (Corning, 353043) at a density of 10^6 cells/well; (2) stage 1 was

extended to 3 days by repeating the final day's treatment; and (3) stage 3 was shortened to 3 days. All tissue culture was carried out in 5% CO$_2$ at 37°C.

Western blotting

Cells were lysed in RIPA buffer (ThermoFisher Scientific, 89901) supplemented with protease inhibitors (Calbiochem). Protein concentrations of lysates were measured by Bradford assay and equal quantities of protein were loaded on precast 7% or 10% SDS-polyacrylamide gels (Bio-Rad). Proteins were transferred to PVDF membranes using the BioCraft (BE-300) transfer system. Membranes were blocked for 1 h at room temperature in blocking buffer [5% non-fat milk in 1× Tris-buffered saline (TBS)+0.1% Tween-20] then incubated with primary antibody in dilution buffer (1% non-fat milk in TBS+0.1% Tween-20) overnight at 4°C. Membranes were washed three times with wash buffer (TBS +0.3 Tween-20) at room temperature for 15 min each. Membranes were then incubated with secondary antibody at room temperature for 1 h. Primary antibody against RFX6 (R&D, AF7780, 1/500) was recognised by the secondary antibody goat anti-sheep IgG HRP conjugate (R&D, HAF016, 1/3000). Western blots were detected using SuperSignal West Pico or Dura Chemiluminescent Substrate (ThermoFisher Scientific) and imaged on X-ray film (GE Healthcare) by using X-Ray processor (Carestream-200).

Immunofluorescence staining

Immunofluorescence was carried out essentially as previously described by Trott et al. (2017). Briefly, adherent cells were washed twice with PBS, fixed in 4% paraformaldehyde for 20 min at room temperature, washed three times with PBS+0.1% BSA, and then incubated with blocking buffer (PBS+20% normal donkey serum+0.1% BSA+0.3% Triton X-100) for 1 h at room temperature. Samples were then incubated overnight at 4°C with primary antibodies diluted in blocking buffer. After washing three times with PBS+0.1% BSA for 15 min, samples were incubated at room temperature for 1 h with secondary antibodies diluted 1:500 in blocking buffer. All subsequent steps were carried out in the dark. After washing three times with PBS+0.1% BSA for 15 min, samples were incubated at room temperature for 15 min with 2 µg/ml Hoechst-33342 (Thermo Fisher Scientific, 62249) diluted in PBS. Finally, samples were washed twice with PBS for 15 min and imaged. Primary antibodies were recognised by Alexa-fluorophore-conjugated secondary antibodies raised in donkey. Primary and secondary antibodies used for immunofluorescence staining are listed in Table S6. Images were acquired using an Olympus FV1000 inverted confocal microscope. The protocol we used to segment cells and extract staining intensities using FIJI is described at https://www.youtube.com/watch?v=82N-eIPqnwM (Fig. 6B). The FACS-style gating is based on staining with secondary antibodies alone. Three fields of view were selected from three independent differentiation experiments.

Flow cytometry

Single cells were obtained from human pluripotent stem cells or their differentiated derivatives at selected time points following Accutase treatment (Thermo Fisher Scientific, 14190), washed once with PBS+1% serum, fixed in 4% paraformaldehyde for 10 min at room temperature, and washed once with wash/permeabilisation buffer (WPB) (Becton Dickinson, 554723). Up to 10^6 cells were incubated with primary or isotype control antibodies diluted in 250 µl WPB. For unconjugated antibodies, cells were washed once with WPB then incubated for 15 min with secondary antibodies diluted in WPB. If staining for a second antigen, cells were washed once with WPB then subjected to the aforementioned incubation step(s). After washing once with WPB, cells were resuspended in PBS+1% serum and analysed using a BD FACSCalibur flow cytometer. All steps were carried out at room temperature. Cells were pelleted by centrifugation at 3500 *g* for 5 min in a microcentrifuge. Antibodies and their respective dilutions are listed in Table S7.

qRT-PCR

Cells were grown in 12-well plates for total RNA isolation. Three wells were harvested per sample to obtain technical triplicates. The RNeasy Mini Kit

(Qiagen, 74104) was used in conjunction with the QIAcube (Qiagen, 9001292) for total RNA extraction. Cell culture media was aspirated and the cells were washed once with D-PBS. After removal of D-PBS, cells were lysed directly in the 12-well plates by adding 350 µl of RLT buffer. Cell lysates were transferred to 2 ml tubes and were either frozen at −80°C or used immediately with the QIAcube for RNA extraction. Each sample was treated with RNase-Free DNase (Qiagen, 7924) to avoid DNA contamination. RNA was eluted in a volume of 30 µl and frozen at −80°C or immediately taken to the next step of first-strand cDNA synthesis. If RNA samples were frozen, they were thawed on ice to prevent degradation. 400 ng of RNA was adjusted to a volume of 20 µl with nuclease free water and then reverse transcribed to generate cDNA using a high-capacity reverse transcription kit and random hexamer primers (Applied Biosystems, 4368814). The reaction tubes were subjected to PCR (10 min at 25°C for the primer annealing step, 120 min at 37°C for the extension step, and finally 5 s at 85°C for the inactivation of the enzyme). The resulting cDNA was diluted to a final volume of 200 µl with nuclease-free water prior to use for qRT-PCR. Quantitative RT-PCR was carried out using SYBR Select Master mix (Applied Biosystems, 4472908). Primers are listed in Table S8.

RNA-sequencing

RNA was isolated from samples harvested from day 8 differentiation cultures using the RNeasy mini kit (Qiagen, 74104). All RNA samples had an RNA integrity number greater than eight. The mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. RNA-seq libraries were generated using the NEBNext Ultra RNA Library Prep Kit, (NEB, E7530L). Messenger RNA was enriched by hybridisation to oligo-dT. RNA-seq libraries (200-300 bp insert cDNA library) were sequenced on a NovoSeq 6000 generating paired-end 150 bp reads. Table S1 contains metadata for all seven samples. Raw RNA-seq read files are have been deposited in ArrayExpress under accession number E-MTAB-9243.

Raw reads were filtered to remove reads containing: (1) adaptors, (2) greater than 10% undetermined bases, (3) more than 50% of bases with a Q score (Quality value) less than six. Reads were mapped using STAR (v2.5) and an index based on the soft masked primary assembly of reference genome GRCh37 and corresponding gene annotation file. HTSeq (v0.6.1) was used to count the number of reads mapped to each gene (gene counts). Differential gene expression analysis was conducted using DESeq2 (v1.22.2) and untransformed data. PCA plots (DESeq2, plotPCA) and heatmaps (pheatmap) were produced using Rlog normalised data (DESeq2). Genes expressed in the dorsal pancreas of CS12-14 human embryos were identified using a published dataset (Jennings et al., 2017). Dorsal pancreas genes are defined as those expressed at more than twice the levels of the hepatic cords (LogFC>1, FDR<0.05).

Chromatin immunoprecipitation sequencing

The derivation and clonal expansion of the triplicated HA RFX6 reporter line generated by CRISPR/Cas9 gene editing coupled with homologous recombination (Fig. 4 and Fig. S3) are described in detail in the supplementary Materials and Methods. All gRNA sequences as well as the T7E1 assay, sequencing and genotyping primers are listed in Table S9. Day 8 *RFX6^{HA}* differentiation cultures were harvested and processed for chromatin immunoprecipitation (ChIP) as described in the manufacturer's instructions for the SimpleChIP Plus Enzymatic Chromatin IP Kit (Magnetic Beads) (Cell Signaling Technology, 9005). Chromatin was immunoprecipitated using rabbit anti-HA immunoglobin (Cell Signaling Technology, 3724). Library preparation was performed using the NEBNext UltraTM II DNA Library Prep Kit for Illumina (E7645). Sequencing was carried out using SE76 on an Illumina NextSeq. Sequencing quality was assessed using FastQC (v0.11.4). Raw sequencing files are available for download at ArrayExpress under accession number E-MTAB-9335.

Reads were aligned to the Ensembl hg19 genome assembly (GRCh37 release 99) using STAR (2.7.1a) with specific options for genomic DNA sequencing ('–alignIntronMax 1 –alignEndsType EndToEnd'). Uniquely mapping aligned reads from individual sequencing lanes were merged for each sample and duplicate reads were marked and counted using SAMBAMBA (v0.6.6). To assess ChIP sequencing enrichment strength,

CHANCE analysis (v2.0) was performed on each ChIP sample with the corresponding input sample as the input control. MACS2 (v2.2.6) peak calling was performed on each ChIP sample with the corresponding input sample as the background. Peaks were filtered for ENCODE blacklisted genomic regions using bedtools intersect (v2.28.0) to remove peaks overlapping with high complexity and artefact regions (https://github.com/Boyle-Lab/Blacklist/blob/master/lists/hg19-blacklist.v2.bed.gz). After filtering, called peaks in H9 ChIP and RFX6-HA ChIP samples were merged and used for differential enrichment analysis. To identify RFX6-HA-specific peaks, differential enrichment analysis was performed on the merged peaks with RFX6-HA ChIP as the target and H9 ChIP as the background using HOMER getDifferentialPeaks (v4.9) to identify differential peaks with fold change ≥4.0 and P≤0.0001 ('-F 4.0 -P 0.0001'). Genomic feature association, gene association and ChIP signal analyses were performed on the RFX6-HA-specific peaks using HOMER annotatePeaks.pl (v4.9). *De novo* motif discovery analysis was performed using HOMER findMotifsGenome (v4.9) fasta-based analysis using the RFX6-HA-specific peaks as the target and the H9-specific peaks as the background. Discovered motifs were matched to known transcription factor target motifs in the JASPAR database. Overlap between RFX6-HA-specific peaks and DNaseI hypersensitivity was performed using bedtools intersect with the ENCODE DNaseI Hypersensitivity Site Master List (125 cell types) (tableName: wgEncodeAwgDnaseMasterSites).

Analysis of human foetal tissue

The collection, use and storage of human embryonic and foetal tissue was carried out with ethical approval from the North West Research Ethics Committee, under the codes of practice issued by the Human Tissue Authority and legislation of the UK Human Tissue Act 2008. Details on collection and handling are as described previously (Jennings et al., 2013). In brief, human embryos and foetal pancreas were fixed within 1 h in 4% paraformaldehyde under RNAase-free conditions, processed and embedded in paraffin wax for sectioning at 5 µm intervals. *In situ* detection of *RFX6* and *PDX1* RNA transcripts was carried out on paraffin wax-embedded tissue sections using RNAScope [Advanced Cell Diagnostics (ACD)]. Sections were pre-treated using an extended protease treatment and hybridised under conditions as described (RNAScope Sample Preparation and Pretreatment Guide) using automated RNAScope probes for *RFX6* and *PDX1* and standard negative *dapB* and positive *PPIB* control probes. Detection was by RNAScope LS Multiplex Reagent Kit (ACD; 322800) for the Leica Bond RX autostainer. Immunohistochemistry was performed on 5 µm sections as described previously (Jennings et al., 2013) using a guinea pig anti-PDX1 primary antibody (Abcam ab47308, 1/500). For further details, see supplementary Materials and Methods.

References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. and Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248-249. doi:10.1038/nmeth0410-248.

Aftab, S., Semenec, L., Chu, J. S.-C. and Chen, N. (2008). Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evol. Biol.* **8**, 226. doi:10.1186/1471-2148-8-226

Ahlgren, U., Jonsson, J. and Edlund, H. (1996). The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. *Development* **122**, 1409-1416.

Ahlgren, U., Pfaff, S. L., Jessell, T. M., Edlund, T. and Edlund, H. (1997). Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature* **385**, 257-260. doi:10.1038/385257a0

Ait-Lounis, A., Baas, D., Barras, E., Benadiba, C., Charollais, A., Nlend Nlend, R., Liègeois, D., Meda, P., Durand, B. and Reith, W. (2007). Novel function of the ciliogenic transcription factor RFX3 in development of the endocrine pancreas. *Diabetes* **56**, 950-959. doi:10.2337/db06-1187

Ait-Lounis, A., Bonal, C., Seguín-Estévez, Q., Schmid, C. D., Bucher, P., Herrera, P. L., Durand, B., Meda, P. and Reith, W. (2010). The transcription factor Rfx3 regulates beta-cell differentiation, function, and glucokinase expression. *Diabetes* **59**, 1674-1685. doi:10.2337/db09-0986

Akiba, K., Ushijima, K., Fukami, M. and Hasegawa, Y. (2020). A heterozygous protein-truncating RFX6 variant in a family with childhood-onset, pregnancy-associated and adult-onset diabetes. *Diabet. Med.* **37**, 1772-1776. doi:10.1111/dme.13970

Bao, X., Adil, M. M., Muckom, R., Zimmermann, J. A., Tran, A., Suhy, N., Xu, Y., Sampayo, R. G., Clark, D. S. and Schaffer, D. V. (2019). Gene editing to generate versatile human pluripotent stem cell reporter lines for analysis of differentiation and lineage tracing. *Stem Cells* **37**, 1556-1566. doi:10.1002/stem.3096

Bastidas-Ponce, A., Scheibner, K., Lickert, H. and Bakhti, M. (2017). Cellular and molecular mechanisms coordinating pancreas development. *Development* **144**, 2873-2888. doi:10.1242/dev.140756

Bort, R., Martinez-Barbera, J. P., Beddington, R. S. P. and Zaret, K. S. (2004). Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* **131**, 797-806. doi:10.1242/dev.00965

Cebola, I., Rodríguez-Seguí, S. A., Cho, C. H.-H., Bessa, J., Rovira, M., Luengo, M., Chhatriwala, M., Berry, A., Ponsa-Cobas, J., Maestro, M. A. et al. (2015). TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat. Cell Biol.* **17**, 615-626. doi:10.1038/ncb3160

Chandra, V., Albagli-Curiel, O., Hastoy, B., Piccand, J., Randriamampita, C., Vaillant, E., Cavé, H., Busiah, K., Froguel, P., Vaxillaire, M. et al. (2014). RFX6 regulates insulin secretion by modulating Ca2+ homeostasis in human β cells. *CellReports* **9**, 2206-2218. doi:10.1016/j.celrep.2014.11.010

Cheng, C., Lu, J., Cao, X., Yang, F.-Y., Liu, J.-Y., Song, L.-N., Shen, H., Liu, C., Zhu, X.-R., Zhou, J.-B. et al. (2019). Identification of Rfx6 target genes involved in pancreas development and insulin translation by ChIP-seq. *Biochem. Biophys. Res. Commun.* **508**, 556-562. doi:10.1016/j.bbrc.2018.11.133

Gonzalez-Perez, A. and Lopez-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* **88**, 440-449. doi:10.1016/j.ajhg.2011.03.004

Jennings, R. E., Berry, A. A., Kirkwood-Wilson, R., Roberts, N. A., Hearn, T., Salisbury, R. J., Blaylock, J., Piper Hanley, K. and Hanley, N. A. (2013). Development of the human pancreas from foregut to endocrine commitment. *Diabetes* **62**, 3514-3522. doi:10.2337/db12-1479

Jennings, R. E., Berry, A. A., Gerrard, D. T., Wearne, S. J., Strutt, J., Withey, S., Chhatriwala, M., Piper Hanley, K., Vallier, L., Bobola, N. et al. (2017). Laser capture and deep sequencing reveals the transcriptomic programmes regulating the onset of pancreas and liver differentiation in human embryos. *Stem Cell Reports* **9**, 1387-1394. doi:10.1016/j.stemcr.2017.09.018

Jennings, R. E., Berry, A. A., Strutt, J. P., Gerrard, D. T. and Hanley, N. A. (2015). Human pancreas development. *Development* **142**, 3126-3137. doi:10.1242/dev.120063

Jonsson, J., Carlsson, L., Edlund, T. and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606-609. doi:10.1038/371606a0

Kawaguchi, Y. (2013). Sox9 and programming of liver and pancreatic progenitors. *J. Clin. Invest* **123**, 1881-1886. doi:10.1172/JCI66022

Kumar, P., Henikoff, S. and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073-1081. doi:10.1038/nprot.2009.86

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S. et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303. doi:10.1101/gr.107524.110

Mitchell, J., Punthakee, Z., Lo, B., Bernard, C., Chong, K., Newman, C., Cartier, L., Desilets, V., Cutz, E., Hansen, I. L. et al. (2004). 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. doi:10.1007/s00125-004-1576-3

Nair, G. G., Liu, J. S., Russ, H. A., Tran, S., Saxton, M. S., Chen, R., Juang, C., Li, M.-L., Nguyen, V. Q., Giacometti, S. et al. (2019). Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived β cells. *Nat. Cell Biol.* **21**, 263-274. doi:10.1038/s41556-018-0271-4

Offield, M. F., Jetton, T. L., Labosky, P. A., Ray, M., Stein, R. W., Magnuson, M. A., Hogan, B. L. and Wright, C. V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983-995.

Piccand, J., Strasser, P., Hodson, D. J., Meunier, A., Ye, T., Keime, C., Birling, M.-C., Rutter, G. A. and Gradwohl, G. (2014). Rfx6 maintains the functional identity of adult pancreatic β cells. *Cell Reports* **9**, 2219-2232. doi:10.1016/j.celrep.2014.11.033

Rezania, A., Bruin, J. E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T. et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121-1133. doi:10.1038/nbt.3033

Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N. et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature Publishing Group* **437**, 1173-1178. doi:10.1038/nature04209

Russ, H. A., Parent, A. V., Ringler, J. J., Hennings, T. G., Nair, G. G., Shveygert, M., Guo, T., Puri, S., Haataja, L., Cirulli, V. et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. *EMBO J.* **34**, 1759-1772. doi:10.15252/embj.201591058

Schwitzgebel, V. M. (2014). Many faces of monogenic diabetes. *J. Diabetes Invest.* **5**, 121-133. doi:10.1111/jdi.12197

Smith, S. B., Qu, H.-Q., Taleb, N., Kishimoto, N. Y., Scheel, D. W., Lu, Y., Patch, A.-M., Grabs, R., Wang, J., Lynn, F. C. et al. (2010). Rfx6 directs islet formation and insulin production in mice and humans. *Nature Publishing Group* **463**, 775-780. doi:10.1038/nature08748

Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M. L., Mellitzer, G. et al. (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. *Development* **137**, 203-212. doi:10.1242/dev.041673

Spiegel, R., Dobbie, A., Hartman, C., de Vries, L., Ellard, S. and Shalev, S. A. (2011). Clinical characterization of a newly described neonatal diabetes syndrome caused by RFX6 mutations. *Am. J. Med. Genet. A* **155**, 2821-2825. doi:10.1002/ajmg.a.34251

Stoffers, D. A., Zinkin, N. T., Stanojevic, V., Clarke, W. L. and Habener, J. F. (1997). Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat. Genet.* **15**, 106-110. doi:10.1038/ng0197-106

Teo, A. K. K., Tsuneyoshi, N., Hoon, S., Tan, E. K., Stanton, L. W., Wright, C. V. E. and Dunn, N. R. (2015). PDX1 binds and represses hepatic genes to ensure robust pancreatic commitment in differentiating human embryonic stem cells. *Stem Cell Reports* **4**, 578-590. doi:10.1016/j.stemcr.2015.02.015

Trott, J., Tan, E. K., Ong, S., Titmarsh, D. M., Denil, S. L. I. J., Giam, M., Wong, C. K., Wang, J., Shboul, M., Eio, M. et al. (2017). Long-term culture of self-renewing pancreatic progenitors derived from human pluripotent stem cells. *Stem Cell Reports* **8**, 1675-1688. doi:10.1016/j.stemcr.2017.05.019

Varshney, A., Scott, L. J., Welch, R. P., Erdos, M. R., Chines, P. S., Narisu, N., Albanus, R. D., Orchard, P., Wolford, B. N., Kursawe, R. et al. (2017). Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci. USA* **114**, 2301-2306. doi:10.1073/pnas.1621192114

Wang, X., Sterr, M., Burtscher, I., Chen, S., Hieronimus, A., Machicao, F., Staiger, H., Häring, H.-U., Lederer, G., Meitinger, T. et al. (2018). Genome-wide analysis of PDX1 target genes in human pancreatic progenitors. *Mol. Metab.* **9**, 57-68. doi:10.1016/j.molmet.2018.01.011

Yourshaw, M., Taylor, S. P., Rao, A. R., Martin, M. G. and Nelson, S. F. (2015). Rich annotation of DNA sequencing variants by leveraging the Ensembl Variant Effect Predictor with plugins. *Brief Bioinform.* **16**, 255-264. doi:10.1093/bib/bbu008

Zhu, Z., Li, Q. V., Lee, K., Rosen, B. P., González, F., Soh, C.-L. and Huangfu, D. (2016). Genome editing of lineage determinants in human pluripotent stem cells reveals mechanisms of pancreatic development and diabetes. *Cell Stem Cell* **18**, 755-768. doi:10.1016/j.stem.2016.03.015