nature genetics

Recessive mutations in a distal *PTF1A* enhancer cause isolated pancreatic agenesis

Michael N Weedon[1,12], Inês Cebola[2–4,12], Ann-Marie Patch[1,12], Sarah E Flanagan[1], Elisa De Franco[1], Richard Caswell[1], Santiago A Rodríguez-Seguí[2,3], Charles Shaw-Smith[1], Candy H-H Cho[5], Hana Lango Allen[1], Jayne A L Houghton[1], Christian L Roth[6], Rongrong Chen[7], Khalid Hussain[8,9], Phil Marsh[10], Ludovic Vallier[5], Anna Murray[1], International Pancreatic Agenesis Consortium[11], Sian Ellard[1,13], Jorge Ferrer[2–4,13] & Andrew T Hattersley[1,13]

The contribution of *cis*-regulatory mutations to human disease remains poorly understood. Whole-genome sequencing can identify all noncoding variants, yet the discrimination of causal regulatory mutations represents a formidable challenge. We used epigenomic annotation in human embryonic stem cell (hESC)-derived pancreatic progenitor cells to guide the interpretation of whole-genome sequences from individuals with isolated pancreatic agenesis. This analysis uncovered six different recessive mutations in a previously uncharacterized ~400-bp sequence located 25 kb downstream of *PTF1A* (encoding pancreas-specific transcription factor 1a) in ten families with pancreatic agenesis. We show that this region acts as a developmental enhancer of *PTF1A* and that the mutations abolish enhancer activity. These mutations are the most common cause of isolated pancreatic agenesis. Integrating genome sequencing and epigenomic annotation in a disease-relevant cell type can thus uncover new noncoding elements underlying human development and disease.

Most individuals with syndromic pancreatic agenesis have heterozygous dominant mutations in *GATA6* (refs. 1,2). Extrapancreatic features in these individuals include cardiac malformations, biliary tract defects, and gut and other endocrine abnormalities. Four families have been reported with syndromic pancreatic agenesis, with severe neurological features and cerebellar agenesis caused by recessive coding mutations in *PTF1A*[3–5]. Most cases of isolated, non-syndromic pancreatic agenesis remain unexplained, with the only cause described being recessive coding mutations in *PDX1* that were reported in two families[6,7]. We previously noted that individuals with unexplained pancreatic agenesis were often born to consanguineous parents and rarely had extrapancreatic features[1]. These observations suggested an autosomal recessive defect underlying isolated pancreatic agenesis.

To identify recessive mutations causing isolated pancreatic agenesis, we used linkage and whole-genome sequencing analyses. Initially, we performed homozygosity mapping in six affected subjects and one unaffected subject from three unrelated consanguineous families (**Supplementary Fig. 1**). This analysis highlighted a single shared locus on chromosome 10 that included *PTF1A*, but mutations in the coding and promoter sequences of *PTF1A* and in the coding sequences of 24 other genes in the region were excluded by Sanger sequencing (**Supplementary Fig. 1** and **Supplementary Table 1**). We next performed whole-genome sequencing of probands from the two families with multiple affected individuals. We first looked for homozygous coding mutations in the exomes of the two individuals for whom whole-genome sequencing was performed. Each genome contained ~3.6 million variants, from which we filtered out any that were present in 81 control genomes or that were present at a frequency of >1% in 1000 Genomes Project data[8]. This filtering left 2,868 and 3,188 rare or newly identified homozygous single-nucleotide variants (SNVs) and indels per subject. Of these, 8 and 19 per subject were annotated as missense, nonsense, frameshift or essential splice site (**Supplementary Table 2**). However, these coding variants either did not cosegregate with disease or were not considered plausible candidates for having a role in pancreas development (**Supplementary Table 2**).

We next searched for noncoding disease-causing mutations among the remaining candidate homozygous variants. We reasoned that any causal variant should disrupt a noncoding genomic element that is active in cells that are relevant to this disease. As isolated pancreatic agenesis must be the result of a defect in early pancreas development, we determined whether any of the rare or newly identified

[1]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK. [2]Genomic Regulation of Pancreatic Beta-Cells Laboratory, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. [3]Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas, Barcelona, Spain. [4]Department of Medicine, Imperial College, London, UK. [5]Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute, Anne McLaren Laboratory for Regenerative Medicine, Cambridge, UK. [6]Seattle Children's Hospital Research Institute, Seattle, Washington, USA. [7]School of Biomedical Science, Waterloo Campus, King's College London, London, UK. [8]London Centre for Paediatric Endocrinology and Metabolism, in partnership with the Great Ormond Street Hospital for Children National Health Service Trust, London, UK. [9]Institute of Child Health, University College London, London, UK. [10]Diabetes Research Group, Diabetes and Nutritional Sciences Division, School of Medicine, King's College London, London, UK. [11]A full list of members and affiliations appears in the **Supplementary Note**. [12]These authors contributed equally to this work. [13]These authors jointly directed this work. Correspondence should be addressed to A.T.H. (a.t.hattersley@exeter.ac.uk) or J.F. (j.ferrer@imperial.ac.uk).

Received 12 July; accepted 16 October; published online 10 November 2013; doi:10.1038/ng.2826

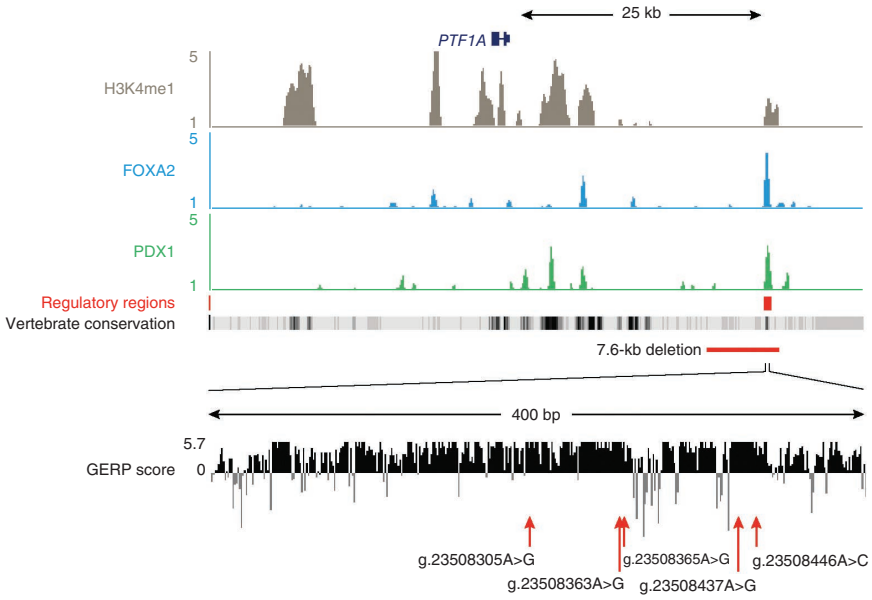

Figure 1 Epigenome annotation of variants from genome sequencing identifies a shared variant in a putative enhancer element. A variant identified by whole-genome sequencing and five additional variants identified in subjects with pancreatic agenesis map to a 25-kb region downstream of *PTF1A*, which contains a single candidate pancreatic progenitor–specific enhancer in a highly conserved 400-bp element. The top panel shows chromatin immunoprecipitation and sequencing (ChIP-seq) density plots for the enhancer mark H3K4me1, and the second and third panels show occupancy for FOXA2 and PDX1, respectively. A broad panel of embryonic and adult human tissues do not show active chromatin marks in this region (**Supplementary Fig. 5**). Vertebrate conservation and mammalian conservation tracks (with conservation measured by GERP score) show the high conservation of this element. The red line depicts the approximate location of a 7.6-kb deletion in this region, and red arrows indicate the positions of point mutations on chromosome 10.

homozygous variants in these subjects mapped to active regulatory regions in pancreatic endoderm cells derived from hESCs (**Fig. 1** and **Supplementary Figs. 2** and **3**). We thus defined 6,109 putative transcriptional enhancers in embryonic pancreatic progenitors that were enriched in monomethylation of histone H3 at lysine 4 (H3K4me1), a post-translational histone modification that is associated with enhancer regions, and were also bound by two or more pancreatic developmental transcription factors that are known to be essential for early pancreas development. Seven homozygous variants from each subject occurred in one of these annotated noncoding regions. However, only 1 of the 6,109 regulatory regions contained a variant in both sequenced individuals, and this was the same variant in the two unrelated subjects (**Supplementary Fig. 2**). This variant on chromosome 10, g.23508437A>G, was located ~25 kb downstream of *PTF1A*, in the region previously identified by homozygosity mapping (**Fig. 1**). The new variant occurred in a short (~400-bp) evolutionarily conserved region that showed enrichment for enhancer marks (H3K4me1 and acetylation of histone H3 at lysine 27 (H3K27ac)) and was bound by the transcription factors FOXA2 and PDX1 in hESC-derived pancreatic progenitor cells (**Fig. 1** and **Supplementary Fig. 4**). Remarkably, this region lacked active chromatin features in 68 embryonic and adult cell types from the Epigenome Roadmap project and in 125 cell types from the Encyclopedia of DNA Elements (ENCODE) Project (which includes an adult pancreatic exocrine cell line), indicating that it is specifically active in pancreatic embryonic progenitors (**Supplementary Fig. 5**). A combination of whole-genome sequencing and annotation of *cis*-regulatory elements therefore identified a recessive mutation that mapped to a putative stage- and lineage-restricted transcriptional enhancer.

We sequenced this putative pancreatic developmental enhancer in 19 additional probands with pancreatic agenesis of unknown etiology (9 with extrapancreatic features and 10 isolated cases) and identified recessive mutations in 7 of the 10 individuals with non-syndromic pancreatic agenesis. We also identified a homozygous mutation in one subject with pancreatic agenesis and intrahepatic cholestatic failure (**Fig. 2** and **Supplementary Table 3**). Of the ten probands with mutations in this element, six had the same chromosome 10 mutation, g.23508437A>G, as part of a shared extended haplotype (minimal shared haplotype of 1.2 Mb; **Supplementary Fig. 6**). Three of the remaining probands had different base-substitution mutations on chromosome 10: a homozygous g.23508363A>G mutation, a homozygous g.23508305A>G mutation and compound heterozygous g.[23508365A>G];[23508446A>C] mutations (**Fig. 2**). In the tenth family, we identified a 7.6-kb deletion by long-range PCR, and sequence analysis showed that the deleted region (chr. 10: 23,502,416–23,510,031) included the entire putative enhancer (**Supplementary Fig. 7**).

Testing of parents and siblings demonstrated cosegregation of the mutations with diabetes and exocrine insufficiency (**Fig. 2** and **Supplementary Table 3**). None of the mutations were present in 1,092 individuals from the 1000 Genomes Project[8] or in dbSNP137, and Sanger sequencing of 299 controls did not detect any of these variants. The deletion

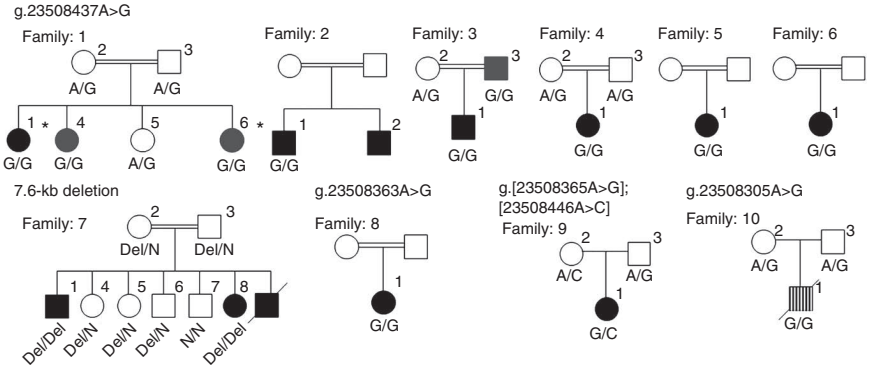

Figure 2 Families with mutations in the *PTF1A* enhancer. Filled black symbols represent individuals with isolated pancreatic agenesis. The striped symbol represents an individual who had pancreatic agenesis with intrahepatic cholestatic failure from which he died. Filled dark gray symbols represent individuals with exocrine insufficiency and young-onset diabetes (age at diagnosis of <22 years). Whole genome–sequenced individuals are indicated by asterisks. DNA from the parents in families 2, 5, 6 and 8 was not available. Del, deletion; N, no deletion.

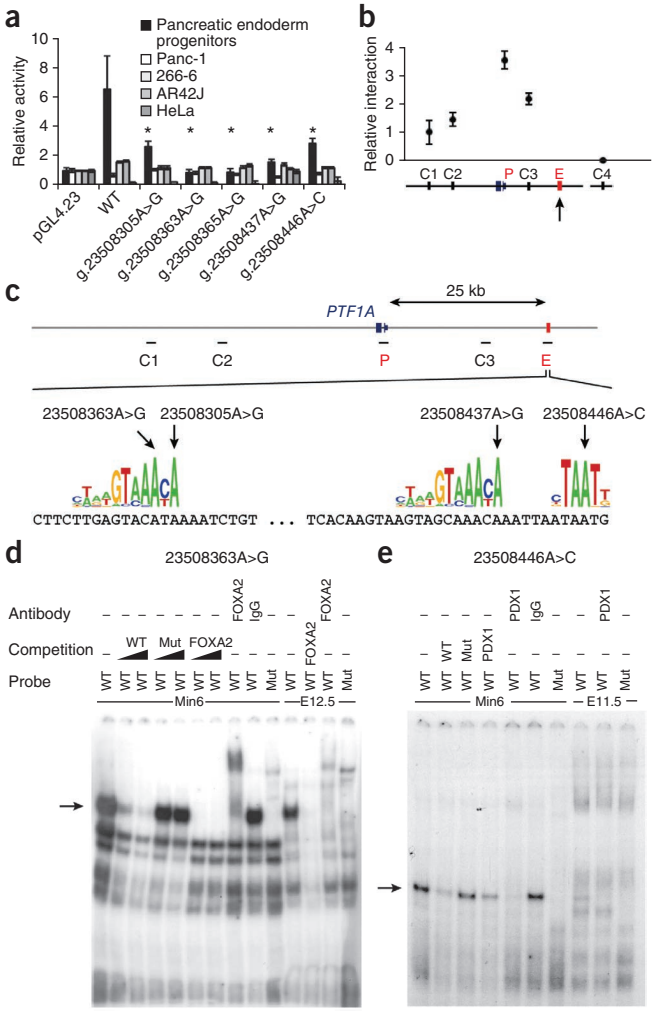

Figure 3 Pancreas agenesis–associated mutations disrupt the function of a transcriptional enhancer that is specifically active in pancreatic progenitors. (**a**) Reporter assays for the new enhancer in hESC-derived pancreatic endoderm progenitors, several adult pancreatic exocrine transformed cells (Panc-1, 266-6 and AR42J) and HeLa cells. Transcriptional enhancer activity was only observed in hESC-derived pancreatic progenitor cells, and this activity was disrupted by all five mutations. Asterisks indicate Student's *t* test $P < 0.05$ for comparisons of mutant and wild-type enhancers in progenitors. Error bars, s.d. from three independent experiments. (**b**) 3C analysis shows that the newly identified enhancer interacts directly with the *PTF1A* promoter in hESC-derived pancreatic progenitor cells. The viewpoint at the *PTF1A* enhancer (E) is indicated by an arrow, and the approximate regions that were tested for interaction with the viewpoint (C1, C2, C3 and P for the *PTF1A* promoter) are shown in **c**. C4 represents a control region in an unrelated locus. Error bars, s.d. from three independent experiments. (**c**) Pancreas agenesis–associated mutations target critical residues in predicted binding sites for the pancreatic regulators FOXA2 and PDX1. (**d**) Electrophoretic mobility shift assay showing high-affinity, sequence-specific interaction of FOXA2 with a double-stranded oligonucleotide containing the wild-type (WT) sequence but not the g.23508363A>G mutation (Mut). The retardation signal was suppressed by unlabeled consensus high-affinity binding site for FOXA2 and was supershifted with antibodies recognizing FOXA2. Black triangles represent competition gradients of 30- and 100-fold excess cold probe. Binding activity is shown for MIN6 cells and dissected pancreatic buds from embryonic day (E) 12.5 mouse embryos. Probe refers to the radioactively labeled probe used in binding assays. The arrow indicates the FOXA2 retardation complex. (**e**) Electrophoretic mobility shift assay showing sequence-specific interaction of PDX1 with oligonucleotide containing the wild-type (WT) sequence but not the g.23508446A>C mutation (Mut). The retardation signal was suppressed by unlabeled consensus high-affinity binding site for PDX1 and was supershifted with antibodies recognizing PDX1. Competition was performed with 100-fold excess cold probe. Binding activity is shown for MIN6 cells and pancreatic buds from E11.5 mouse embryos. The arrow indicates the PDX1 retardation complex.

was not observed in the Database of Genomic Variants[9]. There is very little diversity in humans within this element; the only three variants reported in dbSNP137 or the 1000 Genomes Project are rare (allele frequency of <0.2%). These results provide overwhelming genetic evidence that we have identified the mutations causing non-syndromic pancreatic agenesis in a noncoding genomic region that is likely to be a transcriptional enhancer during pancreas development.

We next tested whether this previously uncharacterized noncoding element acts as a developmental enhancer of *PTF1A*. We linked the wild-type sequence to a minimal promoter and performed luciferase assays in human pancreatic progenitor cells, demonstrating lineage-specific enhancer activity (**Fig. 3a**). The enhancer was not active in adult exocrine pancreatic cell lines, consistent with its having a stage-specific regulatory function (**Fig. 3a**). To assess whether this enhancer truly targets *PTF1A*, we performed chromatin conformation capture (3C) experiments. This analysis demonstrated that the enhancer region establishes direct interactions with the *PTF1A* promoter in human pancreatic progenitor cells (**Fig. 3b**,**c**).

We next demonstrated that the five base-substitution mutations prevent enhancer activity by abolishing transcription factor binding. We noted that three of the mutations disrupted binding sites for FOXA2, and a fourth disrupted a binding site for PDX1 (**Fig. 3c**). FOXA2 and PDX1 are essential transcription factors for pancreatic development[6,10]. Electrophoretic mobility shift assays confirmed that these four mutations abolished binding of FOXA2 or PDX1, as predicted, whereas the remaining point mutation disrupted the affinity of an uncharacterized sequence-specific DNA-binding protein present in mouse pancreatic progenitors (**Fig. 3d**,**e** and **Supplementary Fig. 8**). Notably, all five mutations disrupted the enhancer activity of this region in hESC-derived human pancreatic progenitors (**Fig. 3a**). Collectively, these findings show that multiple mutations causing isolated pancreas agenesis disrupt the function of a previously unrecognized enhancer that targets *PTF1A* in human embryonic pancreatic progenitor cells.

The contribution of noncoding variants to human disease remains poorly understood. There are examples of mutations in distal regulatory elements causing monogenic disease[11–14], but the number of such mutations is small compared to that for coding mutations. Although whole-genome sequencing technologies allow the identification of all genetic variation, the discrimination of functional noncoding causal variants among the millions of noncoding variants present in each individual remains a formidable challenge. The ENCODE Project has recently uncovered functional elements throughout the noncoding genome, leading to expectations that annotation of these elements can be integrated with genome sequencing to discover causal noncoding mutations[15]. Our study now provides an example that validates this expectation and shows that recessively inherited distal *cis*-regulatory mutations in a newly identified developmental enhancer are the most common cause of a rare mendelian disease. The fact that the mutated regulatory element was exclusive to embryonic pancreatic progenitors out of a broad panel of adult and embryonic tissues highlights the importance of analyzing disease-relevant genomic annotations. Our results support efforts to identify mutations in regulatory elements in monogenic disorders by integrating genome sequencing

data with functional annotation from projects such as ENCODE[15] and the Epigenome Roadmap[16]. These findings may also be relevant for future efforts to discover causal alleles in common non-mendelian diseases, where many susceptibility variants appear to lie outside coding regions[17].

In summary, we have demonstrated that mutation of a new distal developmental enhancer of *PTF1A* is a common cause of isolated pancreatic agenesis in humans, and we demonstrate the potential of integrating genome sequencing with epigenomics to identify mutations in new regulatory elements that cause disease.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All ChIP sequencing data from this study have been deposited in ExpressArray under accession E-MTAB-1990.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank M. Day, A. Damhuis and J. Garcia-Hurtado for technical assistance and R. Tearle (Complete Genomics), J. Tena and J.L. Skarmeta (Centro Andaluz de Biología del Desarrollo) for advice. J.F., S.E. and A.T.H. are supported by Wellcome Trust Senior Investigator awards. M.N.W. is supported by the Wellcome Trust as part of WT Biomedical Informatics Hub funding. E.D.F. is funded by the BOLD grant (European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement FP7-PEOPLE-ITN-2008 (Marie Curie Initial Training Networks, Biology of Liver and Pancreatic Development and Disease)). This work was supported by the National Institute for Health Research Exeter Clinical Research Facility through funding for S.E. and A.T.H. and general infrastructure and by the Ministerio de Economía y Competitividad (SAF2011-27086, PLE2009-0162 to J.F.). The views expressed here are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health, UK.

AUTHOR CONTRIBUTIONS

M.N.W., S.E., J.F. and A.T.H. designed the study. M.N.W., A.-M.P., J.A.L.H. and H.L.A. performed bioinformatic analyses. I.C., S.A.R.-S., C.H.-H.C., A.M., L.V. and J.F. performed functional studies. A.-M.P., J.A.L.H., E.D.F., R. Caswell, S.E.F. and S.E. performed Sanger sequencing or deletion analysis and interpreted the results. S.E.F., C.S.-S., K.H., C.L.R., R. Chen, P.M. and A.T.H. analyzed the clinical data. M.N.W., I.C., A.-M.P., S.E., J.F. and A.T.H. prepared the draft manuscript. All authors contributed to discussion of the results and to manuscript preparation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Lango Allen, H. *et al. GATA6* haploinsufficiency causes pancreatic agenesis in humans. *Nat. Genet.* **44**, 20–22 (2012).
2. De Franco, E. *et al. GATA6* mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes* **62**, 993–997 (2013).
3. Sellick, G.S. *et al.* Mutations in *PTF1A* cause pancreatic and cerebellar agenesis. *Nat. Genet.* **36**, 1301–1305 (2004).
4. Tutak, E. *et al.* A Turkish newborn infant with cerebellar agenesis/neonatal diabetes mellitus and *PTF1A* mutation. *Genet. Couns.* **20**, 147–152 (2009).
5. Al-Shammari, M., Al-Husain, M., Al-Kharfy, T. & Alkuraya, F.S. A novel *PTF1A* mutation in a patient with severe pancreatic and cerebellar involvement. *Clin. Genet.* **80**, 196–198 (2011).
6. Stoffers, D.A., Zinkin, N.T., Stanojevic, V., Clarke, W.L. & Habener, J.F. Pancreatic agenesis attributable to a single nucleotide deletion in the human *IPF1* gene coding sequence. *Nat. Genet.* **15**, 106–110 (1997).
7. Schwitzgebel, V.M. *et al.* Agenesis of human pancreas due to decreased half-life of insulin promoter factor 1. *J. Clin. Endocrinol. Metab.* **88**, 4398–4406 (2003).
8. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
9. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
10. Gao, N. *et al.* Dynamic regulation of *Pdx1* enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev.* **22**, 3435–3448 (2008).
11. Cooper, D.N. *et al.* Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum. Mutat.* **31**, 631–655 (2010).
12. Smemo, S. *et al.* Regulatory variation in a *TBX5* enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21**, 3255–3263 (2012).
13. Spielmann, M. *et al.* Homeotic arm-to-leg transformation associated with genomic rearrangements at the *PITX1* locus. *Am. J. Hum. Genet.* **91**, 629–635 (2012).
14. Sankaran, V.G. *et al.* A functional element necessary for fetal hemoglobin silencing. *N. Engl. J. Med.* **365**, 807–814 (2011).
15. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
16. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
17. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).

ONLINE METHODS

Subjects. Pancreatic agenesis was defined as (i) pancreatic β cell failure indicated by neonatal diabetes requiring insulin treatment and (ii) exocrine pancreatic insufficiency requiring enzyme replacement therapy, as previously described[1]. Isolated disease was defined as pancreatic agenesis with normal development and no neurological or other major clinical features. Clinical details of the subjects are provided in **Supplementary Table 3**. Subjects with pancreatic agenesis were recruited by their clinicians for molecular genetic analysis in the Exeter Molecular Genetics Laboratory. The study was conducted in accordance with the Declaration of Helsinki, and all subjects or their parents gave informed consent for genetic testing.

Whole-genome sequencing. Whole-genome sequencing of probands from families 1 and 2 was performed at Complete Genomics. The method has been described previously[18]. Complete Genomics software version 1.8.0.30 was used to align reads to the hg19 genome and call SNVs and indels. In total, 222 and 190 Gb of sequence were mapped with average coverage of 73× and 63×, respectively. Ninety-five percent of hg19 bases had sufficient coverage to be fully called. 3,197,771 and 3,182,809 SNVs were called per sample with transition/transversion ratios of 2.15 and 2.14 and novel (compared to dbSNP131) SNP rates of 4.7 and 4.8%, respectively. 445,141 and 440,357 indels were called per sample with dbSNP131 novelty rates of 22.4 and 22.8%.

For filtering SNVs and indels, we used 69 publically available whole genomes provided by Complete Genomics[18] and 12 additional whole genomes that had also been sequenced by Complete Genomics for a non-overlapping disease. We also filtered out variants present at minor allele frequency of >1% in the 1000 Genomes Project[8].

Differentiation of pancreatic endoderm from human ESCs. Human ESCs (H9, WiCell) were imported under the guidelines of the UK Stem Cell Bank Steering Committee (authorization SCSC10-44). Cells were maintained and differentiated into artificial pancreatic progenitors using a protocol that was previously fully described[19]. These artificial pancreatic progenitors express a constellation of pancreatic endoderm markers including PDX1, HLXB9, NKX6.1, SOX9, HNF6 and PTF1A[19]. In brief, definitive endoderm was induced by growing hESCs in CDM-PVA supplemented with Activin-A (100 ng/ml), bFGF (20 ng/ml), BMP4 (10 ng/ml) and LY294002 (10 μM) (AFBLy). The CDM-PVA AFBLy cocktail was replenished daily, and daily changes of medium were made during the entire differentiation protocol. After the definitive endoderm stage (days 1–3), cells were cultured in Advanced DMEM (Invitrogen) supplemented with SB-431542 (10 μM; Tocris), FGF10 (50 ng/ml; AutogenBioclear), all-trans retinoic acid (2 μM; Sigma) and Noggin (150 ng/ml; R&D Systems) for 3 d (days 4–6). For the next stage (days 7–10), cells were cultured in Advanced DMEM supplemented with human FGF10 (50 ng/ml; AutogenBioclear), all-trans retinoic acid (2 μM), KAAD-cyclopamine (0.25 μM; Toronto Research Chemicals) and Noggin (150 ng/ml) for 3 d. For the last stage (days 10–12), cells were cultured in Advanced DMEM supplemented with human FGF10 (50 ng/ml; R&D Systems) for 3 d. For maturation of pancreatic progenitors (day 15 and day 18 artificial pMPCs), cells were grown in Advanced DMEM supplemented with 1% B27 and DAPT (1 mM) for 3 d and were grown for 3 d more in Advanced DMEM supplemented with 1% B27.

ChIP-seq maps of pancreatic progenitor regulatory elements. Chromatin immunoprecipitation for H3K4me1 (Abcam, ab-8895; $n = 2$), FOXA2 (Santa Cruz Biotechnology, sc-6554; $n = 2$), GATA6 (Santa Cruz Biotechnology, sc-9055X; $n = 1$), HNF1β (Santa Cruz Biotechnology, sc-22840-X; $n = 1$), ONECUT1 (Santa Cruz Biotechnology, sc-13050; $n = 1$) and PDX1 (BCBC, AB2027; $n = 1$) was performed essentially as described[20], using ~10 million artificially derived pancreatic progenitors for each experiment. These transcription factors were chosen because they are known to be essential regulators of early pancreas development[1,6,10,21–25] and because of the availability of antibodies that recognize human epitopes in chromatin immunoprecipitation experiments. Sequencing of immunoprecipitated chromatin and input DNA was performed on an Illumina HiSeq 2000 platform. Transcription factor enrichment sites were detected with MACS v1.4.0beta[26], and H3K4me1-enriched regions were defined with SICER v1.03 (ref. 27). We identified genomic regions that showed H3K4me1 enrichment in duplicate samples and then defined H3K4me1-enriched regions bound by at least two transcription factors that were not located within 1 kb of the transcriptional start sites of RefSeq genes. We then defined the limits of the remaining regions as the outer limits of the transcription factor binding sites that clustered in each H3K4me1-enriched region. Using these definitions, we identified 6,109 putative enhancer regions.

H3K27ac chromatin immunoprecipitation. Chromatin immunoprecipitation for H3K27ac (Abcam, ab-4729: $n = 2$) was performed as previously described[28], using ~10 million artificially derived pancreatic progenitors. Fold enrichment was calculated using the *NANOG* transcriptional start site as the negative control. Oligonucleotides used in this analysis are listed in **Supplementary Table 4**.

Homozygosity mapping. Genome-wide SNP genotyping was performed using the Affymetrix Mapping 10K Xba SNP genotyping chip by Medical Solutions (formerly GeneService) with an average call rate of >96%. Runs of homozygous SNP calls that exceeded 3 cM from at least 20 consecutive probes were identified in 6 affected probands and 1 unaffected sibling from 3 families. Common genomic regions of homozygosity were sought across the affected individuals, excluding any shared with the unaffected sibling (**Supplementary Fig. 1**). The coding exons of all 25 RefSeq genes contained within the single shared region of homozygosity on chromosome 10 and the promoter and upstream conserved region of *PTF1A* were sequenced using capillary sequencing on the Applied Biosystems 3730xl DNA Analyzer (Life Technologies). Primers were designed to cover positions –50 to +10 of each exon, in overlapping fragments if required (primer sequences are available upon request). There was no evidence to support any causative variants within these genes.

Conservation analysis. Eighty percent of bases on chromosome 10 between positions 23,508,149 and 23,508,510 were classified as being part of a vertebrate conserved element by PhastCons[29] (logarithm of odds (LOD) > 17). Multiz alignment results obtained from the UCSC Genome Browser[30] showed that there is conservation of the entire element from human to chicken and that over half the element is conserved to *Xenopus tropicalis*. All mutated bases are highly conserved, with all having GERP[31] scores of 5.65.

Sanger sequencing of the *PTF1A* element. We amplified the conserved ~400-bp element using the primers listed in **Supplementary Table 4**. PCR products were sequenced on an ABI3730 capillary machine (Applied Biosystems) and analyzed using Mutation Surveyor v3.98 (SoftGenetics).

Shared haplotype analysis and testing for cryptic relatedness. We first tested for a shared haplotype in the individuals in whom whole-genome sequencing was undertaken. We only used SNPs for which alleles were fully called in both samples. There were 1,234 consecutive SNP calls on chromosome 10 between positions 21,314,935 and 24,693,292 that were identical between the 2 samples, with the exception of 4 discrepancies (which is within the expected genotyping error rate). For shared haplotype analyses of additional families with the g.23508437A>G mutation we also included three subjects genotyped on the Affymetrix Genome-Wide Human SNP 5.0 or 6.0 array and extracted genotypes from these SNPs for the two individuals who underwent genome sequencing. Sample 6-1 was not included in this analysis because of a lack of dense genotyping data. Any SNP that was not called for at least one sample was excluded from the analysis. One discrepancy per 50 SNPs was tolerated to allow for genotyping error. A graphical representation of the shared haplotype is shown in **Supplementary Figure 6**.

To test whether the shared haplotype could be explained by cryptic relatedness between families, we used KING[32] to estimate the relatedness between probands from each of the families with the g.23508437A>G mutation. We only used SNPs that were present on both the Affymetrix Genome-Wide Human SNP 5.0 and 6.0 arrays. All pairs of probands had a kinship coefficient of <0.022, consistent with them being 'unrelated' (ref. 32).

Deletion analysis. The genomic region chr. 10: 23,501,386–23,512,912 was amplified in subjects 7-4 and 7-8 by long-range PCR using the SequalPrep Long PCR kit (Life Technologies). PCR products were sheared by sonication (Diagenode Bioruptor), and fragments in the size range of 200–300 bp

were isolated for library preparation with NextFlex adaptors with a 6-base index sequence tag. Individual libraries were enriched by six cycles of PCR amplification and were then pooled in equimolar quantities for 100-bp paired-end sequencing on an Illumina HiSeq 2000 sequencer. We used the Burrows-Wheeler Aligner (BWA v0.6.2)[33] to align sequence reads to the hg19 reference genome and then visualized the breakpoints using the Integrative Genomics Viewer[34], which demonstrated that the deletion breakpoints occurred at positions 23,502,416 and 23,510,031 on chromosome 10 (**Supplementary Fig. 7**). The deletion mutation was investigated in all available members of family 7 using a junction fragment PCR assay (primer sequences available upon request).

Sanger sequencing of the *PTF1A* enhancer in control samples. We sequenced the putative *PTF1A* enhancer element in 150 healthy controls of European descent from the Exeter Family Study of Childhood Health[35] and in 149 individuals from Turkey using Sanger sequencing (**Supplementary Table 4**). No variants were identified.

Transcription factor binding motif analysis. Motif discovery over the point mutation sites comparing wild-type and mutation-containing sequences was performed using HOMER[36].

Electrophoretic mobility shift assays. Mouse pancreatic buds were dissected from E11.5 and E12.5 C57BL/6J mouse embryos as described[37]. The study was approved by the Animal Experimentation Ethics Committee of the University of Barcelona. Nuclear extracts were purified as described[38]. Binding of nuclear extracts from embryonic pancreas and MIN6 β cells (kindly provided by J.I. Miyazaki, Tokyo, Japan) to [^{32}P]-labeled oligonucleotides that included either wild-type sequence or the mutation-containing sequence was performed as described previously[39]. The oligonucleotide sequences used are listed in **Supplementary Table 4**. Assay specificity was assessed by preincubation of nuclear lysates with 30- and 100-fold excess of unlabeled wild-type, mutant or consensus double-stranded oligonucleotide. Supershifts were performed using 2 µl of goat polyclonal serum to FOXA2 (sc-6554) or PDX1 (sc-14662) or control IgG (sc-2028) (all from Santa Cruz Biotechnology).

3C assays. Approximately 1×10^7 artificial pancreatic progenitors were fixed for 20 min at 4 °C in 4% paraformaldehyde, washed three times in PBS and lysed (10 mM Tris-HCl, pH 8, 10 mM NaCl, 0.3% IGEPAL CA-630 (Sigma-Aldrich, I8896) and 1× protease inhibitor cocktail (Complete, Roche)). Nuclei were digested with HindIII endonuclease (New England BioLabs). DNA was then ligated with T4 DNA ligase (Promega). Locus-specific primers were designed with Primer3 v. 0.4.0 (ref. 40) as described previously[41]. Relative enrichment of each ligation product was measured by real-time quantitative PCR. The primer specific to the *PTF1A* enhancer (3C-E) was considered fixed, and interaction with the *PTF1A* promoter was tested using primers close to either the promoter (3C-P) or adjacent control regions (3C-Crt1, 3C-Crt2 and 3C-Crt3). A primer specific to the *XBP1* promoter (3C-XBP1) was used as an unrelated locus control. Sequences for all primers are shown in **Supplementary Table 4**. Amplimers were compared with parallel amplification products from serial dilutions of control BACs that encompass the genomic region of interest (RP11-938O7) and the *XBP1* control locus (RP11-594I15), which were processed identically to the chromatin from the pancreatic progenitors.

***PTF1A* enhancer cloning and luciferase reporter assays.** Wild-type and mutant *PTF1A* enhancer sequences were PCR amplified from genomic DNA from a control individual and affected individuals carrying the mutations, respectively, with Phusion High-Fidelity DNA Polymerase (New England BioLabs) (see **Supplementary Table 4** for primer sequences) and cloned into pENTR/D-TOPO (Invitrogen). Enhancer sequences were then shuttled into a pGL4.23[*luc2*/minP] vector backbone (Promega) previously adapted for Gateway cloning pGL4.23-GW (L. Pasquali, unpublished data), using Gateway LR Clonase II Enzyme Mix (Invitrogen). Correct cloning was assessed by Sanger sequencing and restriction enzyme digestion.

DNA was prepared with the PureYield Plasmid Maxiprep System (Promega). At day 10 of differentiation, artificial pancreatic progenitors were transfected in 24-well plates with 400 ng of pGL4.23-GW-PTF1A_Enhancer vectors and 4 ng of *Renilla* normalizer control using Lipofectamine 2000 (Invitrogen) in Opti-MEM (Gibco) according to the manufacturers' instructions. Panc-1 (human pancreatic ductal; ATCC, CRL-1469), 266-6 (mouse pancreatic acinar; ATCC, CRL-2151), AR42J (rat pancreatic acinar; ATCC, CRL-1492) and HeLa (ATCC, CRM-CCL2) cells were transfected in 96-well plates using Lipofectamine 2000 in Opti-MEM at a density of 4×10^4 cells per well, according to the manufacturer's instructions for this format. Luciferase activity was measured 48 h after transfection with the Dual-Luciferase Reporter Assay System (Promega). Firefly luciferase activity was normalized to *Renilla* luciferase activity and then to the amount of pGL4.23[*luc2*/minP] vector backbone. Statistical significance was determined by comparing firefly/*Renilla* luciferase values for each mutant to those for the wild-type construct using a two-sided *t* test. All DNA preparations were transfected into cells triplicate.

18. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
19. Cho, C.H. *et al.* Inhibition of activin/nodal signalling is necessary for pancreatic differentiation of human pluripotent stem cells. *Diabetologia* **55**, 3284–3295 (2012).
20. Morán, I. *et al.* Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* **16**, 435–448 (2012).
21. Carrasco, M., Delgado, I., Soria, B., Martin, F. & Rojas, A. GATA4 and GATA6 control mouse pancreas organogenesis. *J. Clin. Invest.* **122**, 3504–3515 (2012).
22. Xuan, S. *et al.* Pancreas-specific deletion of mouse *Gata4* and *Gata6* causes pancreatic agenesis. *J. Clin. Invest.* **122**, 3516–3528 (2012).
23. Haumaitre, C. *et al.* Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. *Proc. Natl. Acad. Sci. USA* **102**, 1490–1495 (2005).
24. Jacquemin, P. *et al.* Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene *ngn3*. *Mol. Cell Biol.* **20**, 4445–4454 (2000).
25. Offield, M.F. *et al.* PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983–995 (1996).
26. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
27. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).
28. van Arensbergen, J. *et al.* Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing β-cells to adopt a neural gene activity program. *Genome Res.* **20**, 722–732 (2010).
29. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
30. Meyer, L.R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–D69 (2013).
31. Davydov, E.V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).
32. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
35. Knight, B., Shields, B.M. & Hattersley, A.T. The Exeter Family Study of Childhood Health (EFSOCH): study protocol and methodology. *Paediatr. Perinat. Epidemiol.* **20**, 172–179 (2006).
36. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
37. van Arensbergen, J. *et al.* Ring1b bookmarks genes in pancreatic embryonic progenitors for repression in adult β cells. *Genes Dev.* **27**, 52–63 (2013).
38. Maestro, M.A. *et al. Hnf6* and *Tcf2* (MODY5) are linked in a gene network operating in a precursor cell domain of the embryonic pancreas. *Hum. Mol. Genet.* **12**, 3307–3314 (2003).
39. Boj, S.F., Parrizas, M., Maestro, M.A. & Ferrer, J. A transcription factor regulatory circuit in differentiated pancreatic cells. *Proc. Natl. Acad. Sci. USA* **98**, 14481–14486 (2001).
40. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
41. Tena, J.J. *et al.* An evolutionarily conserved three-dimensional structure in the vertebrate Irx clusters facilitates enhancer sharing and coregulation. *Nat. Commun.* **2**, 310 (2011).

doi:10.1038/ng.2826