Multidimensional chromatin profiling of zebrafish pancreas to uncover and investigate disease-relevant enhancers

Renata Bordeira-Carriço1,2,9, Joana Teixeira1,2,3,9, Marta Duque1,2,3,9, Mafalda Galhardo1,2,4, Diogo Ribeiro1,2, Rafael D. Acemel5, Panos. N. Firbas5, Juan J. Tena5, Ana Eufrásio1,2,3, Joana Marques1,2, Fábio J. Ferreira1,2,6, Telmo Freitas1,2, Fátima Carneiro7,8, José Luís Goméz-Skarmeta5 & José Bessa1,2

The pancreas is a central organ for human diseases. Most alleles uncovered by genome-wide association studies of pancreatic dysfunction traits overlap with non-coding sequences of DNA. Many contain epigenetic marks of cis-regulatory elements active in pancreatic cells, suggesting that alterations in these sequences contribute to pancreatic diseases. Animal models greatly help to understand the role of non-coding alterations in disease. However, interspecies identification of equivalent cis-regulatory elements faces fundamental challenges, including lack of sequence conservation. Here we combine epigenetic assays with reporter assays in zebrafish and human pancreatic cells to identify interspecies functionally equivalent cis-regulatory elements, regardless of sequence conservation. Among other potential disease-relevant enhancers, we identify a zebrafish ptf1a distal-enhancer whose deletion causes pancreatic agenesis, a phenotype previously found to be induced by mutations in a distal-enhancer of PTF1A in humans, further supporting the causality of this condition in vivo. This approach helps to uncover interspecies functionally equivalent cis-regulatory elements and their potential role in human disease.

https://doi.org/10.1038/s41467-022-29551-7

1 Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto 4200-135, Portugal. 2 Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto 4200-135, Portugal. 3 Doctoral program in Molecular and Cell Biology (MCIbiology), IBAS—Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal. 4 Instituto de Ciências, Tecnologias e Agroambiente (CIBIO), Universidade do Porto, 4051-401 Porto, Portugal. 5 Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide, Sevilla 41013, Spain. 6 Doctoral program in Areas of Basic and Applied Biology (GABBBA), Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto 4050-313, Portugal. 7 Departamento de Patologia, Faculdade de Medicina da Universidade do Porto (FMUP), Porto 4200-319, Portugal. 8 Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto 4200-465, Portugal. 9 These authors contributed equally: Renata Bordeira-Carriço, Joana Teixeira, Marta Duque. ✉ email: jose.bessa@ibmc.up.pt
The mechanisms that tightly control transcription are essential for organ function. The transcriptional regulation of genes is controlled by non-coding cis-regulatory elements (CREs) spread over large genomic distances. Genome-Wide Association Studies (GWAS) have identified many non-coding disease-associated alleles that have a hereditary component and overlap with CREs epigenetic signatures, suggesting that the disruption of CREs may be one of the genetic bases of human disease. This is the case of some pancreatic diseases, such as pancreatic cancer and diabetes, that have a heavy societal burden, with incidence and death rates increasing worldwide. Many previous studies demonstrated an enrichment of diabetes-associated variants in adult human islet enhancers, corroborating the hypothesis of pancreatic diseases being caused by alterations in CREs. Likewise, experimental in vivo and in vitro enhancer reporter assays also showed that specific islet enhancer variants correlate with altered regulatory functions. Studies of the role of CREs' mutations in the development of pancreatic diseases using in vivo models would provide invaluable insight given the complex regulatory networks involved. However, evidence from in vivo models of the role of CREs' mutations in the development of pancreatic diseases are still scarce.

The zebrafish is a vertebrate model suitable for genetic manipulation, with a pancreas that shares many similarities with the human pancreas, including similar transcription factors (TFs) and genetic networks of pancreatic development and function. Thus, the zebrafish is a suitable in vivo model to validate causal regulatory variants. Yet, the identification of interspecies functionally equivalent CREs faces unsolved fundamental challenges, such as low conservation of interspecies non-coding sequences, and for the minority of CREs whose sequence is conserved, their fast-evolving functionality. Indeed, although sequence conservation of non-coding sequences has successfully been used to find enhancers, many with interspecies orthologous identities, it has also been demonstrated to be insufficient for identifying all enhancers within a genome and between species. To bypass these limitations, this work profiled the chromatin state of zebrafish pancreas cells and chromatin interaction points. We were able to accurately identify zebrafish pancreatic enhancers and, by comparisons with similar human datasets, we predicted functionally equivalent pancreatic enhancers. These findings revealed a previously unidentified human enhancer in the landscape of the tumour suppressor ARID1A, with a potential role in the susceptibility to pancreatic cancer. Additionally, we explored the regulatory landscape of PTF1A, known to contain a human distal enhancer whose deletion leads to pancreatic agenesis/hypoplasia and found a zebrafish distal ptf1a enhancer that contains similar regulatory information to its human counterpart. We further demonstrated its functional equivalency by showing that its ablation induces pancreatic agenesis, explained by a reduction in the pancreatic progenitor domain early in development. Taken together, the multidimensional chromatin profiling used here allowed the establishment of previously unknown functional connections between human and zebrafish enhancers. These bridges between different species are invaluable for the prediction of new disease-relevant enhancers and the study of their role in human disease.

## Results

### Zebrafish putative pancreatic enhancers share developmental roles.

When comparing the basic structure of the human and zebrafish adult pancreas we observed that the organ structure is analogous between the two species (Fig. 1a). We further extended this comparison to the cellular composition of the main cell types of the pancreas between zebrafish, mouse and human, and found that the predominance of the major cellular types is maintained in these three vertebrates (Supplementary Fig. 1). Because of these extended similarities between the zebrafish and mammal pancreas, the zebrafish has been used as a model to study pancreatic diseases. Furthermore, these similarities hint at the existence of shared genetic networks that operate, likely through equivalent sets of CREs, in these three species. Thus, we explored the chromatin state and chromatin interaction points of zebrafish whole pancreas, to gather information about endocrine and exocrine cells, and compared it to human datasets. To identify CREs active in the zebrafish adult pancreas, we performed ChIP-seq for H3K27ac, a key histone modification associated with active enhancers, and ATAC-seq, an assay that identifies regions of open chromatin (Fig. 1b). We also performed HiChIP against H3K4me3 to detect active promoters interacting with the uncovered enhancers (Fig. 1b).

We identified 14,753 putative active enhancers, mostly in intergenic regions (57.8%), and 23,298 putative active promoters corresponding to 9,848 genes (Fig. 1c; Supplementary Dataset 1a–c). To identify a subset of pancreatic enhancers with higher tissue-specificity, we compared the H3K27ac data from adult zebrafish pancreas to whole zebrafish embryos at four developmental stages, Dome, 80% epiboly, 24 h post-fertilisation (hpf) and 48 hpf, since these comprise differentiated and non-differentiated cells from many different tissues. We found that 7,115 putative enhancers (48.2%) are active only in the differentiated adult pancreas (PsE; Fig. 1c; Supplementary Dataset 1a–c) while the remaining 7,638 (51.8%) are also broadly active in developing embryos (DeVE), suggesting that their activity is not restricted to the pancreas. DeVE presented 4 clusters (C1–4) with different H3K27ac abundance profiles during the different developmental stages (Fig. 1d; Supplementary Fig. 2a; Supplementary Dataset 1a–l), suggesting that, apart from their activity in the adult pancreas, these enhancers might function in other cell types. C1 and C4 show similar levels of H3K27ac in all developmental stages, compatible with a putative ubiquitous enhancer activity, while C2 and C3 show different levels of H3K27ac during development, which may reflect a dynamic state of repression (C2) and activation (C3) of enhancers, or alternatively, differences in the abundance of cells where these enhancers are active during development.

### Functional similarities between human and zebrafish pancreatic enhancers.

Pancreatic enhancers are expected to activate the expression of genes in the pancreas. To test if the predicted enhancers correlate with the expression of target genes in the pancreas, we identified the nearest genes to each putative pancreatic enhancer and observed that genes nearby PsE are enriched for exocrine pancreas expression (p < 4.27E–9; Supplementary Fig. 2b; Supplementary Dataset 2a–c), detected by in situ hybridisation. These results contrast with the ones obtained for DeVE, for which nearby genes are enriched for expression in several other tissues, including epidermis and endothelial cells (Supplementary Fig. 2; Supplementary Dataset 2d–f), suggesting a higher tissue-specificity of PsE. Additionally, the presence of endothelial expression also in genes associated to the PsE group suggests the detection of endothelial enhancers, likely derived from the vasculature present in the zebrafish adult pancreas (Supplementary Dataset 2d–f).

To improve the enhancer to gene association, we used H3K4me3 HiChIP to detect chromatin interactions between active promoters and putative enhancers in the zebrafish adult pancreas (Fig. 1b; Supplementary Dataset 3a) and used RNA-seq to evaluate transcription (Fig. 1b). We found that, compared to all genes, PsE-associated genes have a higher average expression in multiple pancreatic cell types (Fig. 2a, Supplementary Dataset 2a; Supplementary Dataset 3a).
Dataset 3b). As expected, these expression results contrast with the lower average expression levels of the PsE-associated genes compared to all genes in a distantly related control tissue such as the muscle (Fig. 2a, Supplementary Dataset 3b). Similar results were obtained when analysing genes associated to the other identified clusters of pancreatic enhancers, specifically, DevE, C1-C4 and the total dataset of pancreatic enhancers altogether (PsEs+DevE; Supplementary Fig. 2c, d, Supplementary Dataset 3c–g), which had higher expression levels for at least one pancreatic adult tissue and lower expression levels in the muscle (control tissue), when compared to all transcribed genes. Next, we performed a similar analysis by calculating the ratio of the average expression level of genes associated to C1-4 and PsE putative enhancers (HC) divided by the average expression of all genes (AllG), using the previously published transcriptome of whole zebrafish embryos from 18 developmental stages. We found that the genes associated to C1-4 and PsE have a HC/AllG ratio ≥ 1 (Fig. 2b; Supplementary Fig. 2e) and that the HC/AllG ratio of the DevE associated genes is higher than the one of PsE-associated genes, for most of the analysed developmental time points (Fig. 2b). These results suggest that DevE enhancers likely control gene expression during development in embryonic stages of the zebrafish. This hypothesis is further supported by the observed variation of the HC/AllG ratio during development that
The zebrafish pancreas, from histology to chromatin state. a Comparison of the basic structure of the human and zebrafish adult pancreas. Above: Dissected adult male Tg(insulin:GFP, elastase:mCherry) zebrafish; insulin and elastase promoters drive GFP expression in beta-cells (green) and mCherry in acinar cells (red), respectively. IN, intestine; LRL, Liver right lobe; LT, left testis; PI, principal islet; SI, secondary islets; SB, swim bladder. Below: Histology of the pancreas; transverse sections with hematoxylin/eosin staining showing islets of Langerhans (black dashed lines) surrounded by exocrine tissue in zebrafish and human pancreas. Magnification: ×40 and scale bar: 1 mm. b Genomic landscape of gata6 in the zebrafish adult pancreas showing the H3K27ac ChIP-seq profile (black) and ATAC-seq peaks (blue) from whole pancreas, RNA-seq from exocrine pancreas (green) and a heat map for chromatin interactions with gata6 promoter detected by HiChIP for H3K4me3 from whole pancreas (below). A putative enhancer sequence that interacts with the gata6 promoter is highlighted by the light blue box. c Bar plot (left panel) showing the number of genes with active promoters (defined by H3K4me3 signal, gray bar) and putative active enhancers in adult zebrafish pancreas (defined by H3K27ac mark, green bar), and their distribution throughout the regions of the genome (right panel). d Above: Venn diagram showing the overlap of putative active enhancers in adult zebrafish pancreas and stages of zebrafish embryonic development. Putative active enhancers exclusive to the adult pancreas form the pancreas-specific enhancers (PsE) group, while the shared enhancers belong to the developmental shared enhancers (DevE) group (Supplementary Dataset 1e, f). Below: Heat maps showing clusters of H3K27ac mark for PsE and DevE enhancers during embryonic development [dome, 80% epiboly (80%e), 24 hpf, 48 hpf] and in adult pancreas. A window of 10 kb around the reference coordinates for each sequence was used and the density files were subjected to k-means clustering, obtaining four different clusters in DevE: C1, Cluster 1; C2, Cluster 2; C3, Cluster 3; and C4, Cluster 4. For c, d, source data are provided as a Source Data file.

partially reflects the variation of H3K27ac signal observed in the enhancers of the C1-4 clusters (Fig. 1d, Fig. 2b and Supplementary Fig. 2e). For instance, the C2 group that shows an increased presence of H3K27ac signal at Dome and 80% epiboly developmental time-points (Fig. 1d), also shows an increased HC/AllG ratio in the earliest developmental time points (BDO:blastula to G75: 75%epiboly; Fig. 2b and Supplementary Fig. 2e). These results suggest that C1-4 enhancers control gene expression in the adult differentiated pancreas, in addition to other cell types during development. Overall, these results increase the robustness of the pancreatic enhancers predictions, since it is possible to correlate with the transcription of the respective putative target genes.

To determine if the detected H3K27ac signal is a good predictor of active pancreatic enhancers, we performed in vivo enhancer reporter assays for 17 regions within the regulatory landscapes of known pancreatic genes. We selected sequences with detectable, but variable, H3K27ac signal overlapping with open chromatin, detected by ATAC-seq55. Of the 10 sequences with the highest H3K27ac values (-log10(p-value) from 36.5 to 92.1), 6 were validated in vivo as pancreatic enhancers (60%; Fig. 2c, d, Supplementary Fig. 3a and Supplementary Dataset 4a). Conversely, of the remaining 7 sequences with the lowest H3K27ac values (-log10(p-value) from 18.5 to 28.4), only 1 showed strong and reproducible evidence of pancreatic enhancer activity (14%, Supplementary Fig.3a–c and Supplementary Dataset 4a). Previous studies described similar percentages of validated enhancers from H3K27ac positive sequences52–54. These results validate the robustness of pancreatic enhancers prediction based on chromatin state and further suggest that the abundance of H3K27ac mark in genomic locations might improve such predictions.

We observed that out of 14753 putative zebrafish pancreatic enhancers, only 12.49% (n = 1842) could be directly aligned to the human genome55 (Fig. 3a and Supplementary Dataset 3i–l). A similar proportion was found in the group of developmental enhancers (11.36%; 7326 out of 64,498; Fig. 3a). Using the corresponding human sequences from the pancreas and developmental enhancers groups, we found that they share similar PhastCons conservation scores (Fig. 3b; Supplementary Fig. 3d and Supplementary Dataset 3m–p). Next, we wanted to determine if the zebrafish putative pancreatic enhancers that align to the human genome also overlap with H3K27ac signal from human pancreas. Only a minority of interspecies aligned sequences shared H3K27ac signal (total pancreas data set: 227 out of 1842; PsE: 115 out of 1052; DevE: 112 out of 790). The human sequences, that shared H3K27ac signal with zebrafish, did not show a higher average conservation score than the aligned sequences that showed H3K27ac signal in zebrafish alone (Fig. 3b and Supplementary Fig. 3e; Average sequence conservation score for H3K27ac non-shared vs shared signal, Pancreas: 0.40±0.36, PsE:0.42±0.41, DevE:0.36±0.34). Notwithstanding the low absolute numbers of aligned sequences that share H3K27ac signal in human and zebrafish pancreas, these sequences represent a clear enrichment compared to the overlap obtained by randomized set of sequences in the human genome (3.21 times higher for pancreas, 2.79 times higher for PsE, 3.76 times higher for DevE and 1.76 times higher for embryo, Fig. 3c; Supplementary Dataset 3g). Overall, these results suggest that pancreatic enhancer function is not a strong condition to impose sequence conservation.

Following these data, we assessed whether functionally equivalent pancreatic CREs exist between human and zebrafish, despite an overall lack of sequence conservation. To explore this possibility, we investigated if the genes interacting with each cluster of zebrafish enhancers were enriched for homologs of human genes associated with pancreatic diseases, which would suggest the existence of functionally equivalent pancreatic CREs with potential biomedical relevance. Such enrichment was observed for the clusters of late development and adult pancreas (PsE, C3 and C4; Fig. 3d; Supplementary Dataset 3r, s). Human gene-disease associations were retrieved from DisGeNET56 and we observed that 306 out of 836 zebrafish genes (36.6%) homologous to human pancreas disease-associated genes also interact with zebrafish pancreatic enhancers.

Enhancers can exist in their typical form, as short and restricted regions of DNA, or they can be present as large regions of hyperactive chromatin referred to as super enhancers13,57,58. Several computational approaches have been applied to identify super enhancers in vertebrate genomes, including in human and zebrafish59. We searched for super enhancers active in the pancreas of human and zebrafish (Supplementary Dataset 1m, n; 275 in zebrafish and 875 in human), to understand if pancreatic super enhancers control the same genes in both species, further suggesting an equivalency in function. Gene ontology for putative target genes showed a similar enrichment for transcriptional regulation in both species and several of these genes corresponded to the same orthologues (32 out of the 271 zebrafish genes; Supplementary Fig. 3f–g), some with important pancreatic functions, such as INSR, a critical regulator of glucose homoeostasis60 and GATA6, which plays a crucial role in pancreas development and β-cell function61 (Supplementary Fig. 3h). We further inquired if human and zebrafish enhancers might operate similarly, using equivalent TFs. To test this, we performed a motif enrichment search for TF binding sites (TFBS)
in regions of open chromatin identified by ATAC-seq\textsuperscript{45}, within the 14753 pancreatic enhancers, and found several TFBS for known pancreatic TFs (ZP; Fig. 3f, Supplementary Fig. 4a, and Supplementary Dataset 3t, u). We also performed a similar analysis using available human whole pancreas datasets (HP\textsuperscript{62}; Datasets summarised in Supplementary Dataset 4g). To compare the extent of overlap of enriched motifs in human and zebrafish pancreatic enhancers with motifs enriched in other pancreas unrelated enhancers, we have performed a similar motif enrichment search for datasets of zebrafish embryos (D80, dome and 80\% epiboly; 24 HPF, 24 hpf) and human heart ventricle (\textsuperscript{7}62; Datasets summarised in Supplementary Dataset 4g). We selected the top 140 enriched motifs from each dataset and observed that the majority of the common motifs were found in zebrafish (ZP) and human (HP) pancreas datasets (ZP,HP:98; ZP,HP:80:63; HP,HP:80:61) (Fig. 3g, Supplementary Fig. 4b), while comparisons with the human ventricle (V) showed that ZP,HP was the second largest group following HP, V (Supplementary Fig. 4c).

Several TFs, such as Ptf1a, Pdx1, Pax6 and Sox9, are known to be important for pancreas function or development in several vertebrate species, including human and zebrafish\textsuperscript{2,63–65}. As shown above, human and zebrafish pancreatic enhancers are enriched for many shared TFBS, therefore it is reasonable to expect that many of these TFBS are from TFs known to have an important pancreatic function. To test this hypothesis, we have selected 25 TFs known to be required for pancreas function and development and calculated the distribution of the respective TFBS motifs within the previously identified enriched motifs.
described in Supplementary Dataset 3t. We found that the majority of the TFBS motifs from the pancreatic TFs were within the ZP,HP overlapping datasets, regardless of the compared groups (Supplementary Fig. 4d–f). These results suggest that the same set of TFs operates in zebrafish and human pancreatic enhancers. Overall, these results argue in favour of interspecies functional equivalency of enhancers.

**Landscape of arid1a reveals potential pancreatic cancer associated enhancer.** To better address the hypothesis of interspecies functional equivalency of enhancers, we focused on the regulatory landscape of a gene that is potentially linked to human pancreatic diseases. We selected arid1ab, the orthologue of human ARID1A, a tumour-suppressor gene associated with cancer in several different cell types, including pancreatic ductal adenocarcinoma. ARID1A plays a key role in the regulation of DNA damage repair, by promoting an efficient processing of double-strand breaks into single-strand ends, being required to sustain DNA damage signalling and repair, hence suppressing tumorigenesis.

We identified several putative enhancers (zA.E1-4, Fig. 4a), that we tested in vivo using enhancer reporter assays (Supplementary...
Fig. 3 The zebrafish and human pancreas share cis-regulatory similarities. a Percentage of predicted zebrafish pancreatic enhancer sequences aligned to the human genome. Sequences are grouped in different clusters: “Pancreas” that includes PsE and DevE; “PsE”; “DevE”; “Embryo” that include putative enhancers active only during embryonic development. b PhastCons scores (99 vertebrate genomes against hg38) for human sequences converted from zebrafish putative enhancers. Grey dots label conserved sequences that do not overlap with H3K27ac mark in human pancreas (Pancreas-1801, PsE-1017, DevE-784 and Embryo-6792). Blue dots label conserved sequences that also show H3K27ac signal in human pancreas (ENCODE data; Pancreas-227, PsE-112, DevE-115). Green diamonds: average (grey dots: 0.40, 0.42, 0.36, 0.39; blue dots: 0.36, 0.41, 0.34, respectively for Pancreas, PsE, DevE and Embryo). Red line: median (grey dots: 0.10, 0.17, 0.05, 0.08; blue dots: 0.06, 0.09, and 0.03, respectively for pancreas, PsE, DevE and Embryo). The embryo dataset is composed by different developmental stages (Dome, 80% Epiboly, 24 hpf and 48 hpf). c Ratio between the number of human sequences conserved with the zebrafish putative active enhancers (Pancreas-3.2, PsE-2.79, DevE-3.76 or Embryo-1.76) overlapping H3K27ac signal in human pancreas (ENCODE data) over the average of a 10^5 random shuffling of human sequences overlapping with H3K27ac signal in human pancreas (Supplementary Dataset 3q; empirical p-value < 1E−5). d Heatmap showing -\log_10(p-values) from hypergeometric enrichment test for pancreatic disease association on the genes linked by HiChIP to each enhancer cluster. Represented values meet the criteria: q-value ≤ 0.05 and fold enrichment ≥ 1.5. e Genomic landscape of the human INS gene (top) and zebrafish arid1ab ortholog (bottom), showing H3K27ac signal and predicted super-enhancers (blue). f Relevant pancreatic transcription factors whose binding motifs are enriched in zebrafish pancreas H3K27ac ChIP-seq data. g Venn diagram of the top 140 enriched TFBS motifs in H3K27ac positive sequences in three different datasets: zebrafish pancreas (ZP), human pancreas (HP) and dome+80%epiboly embryos (DBO). Number of motifs shared between pairs of groups (arrows). p-values are described (p: hypergeometric enrichment test). The enrichment of the observed vs expected is represented (E). p-values ≤ 0.05 were considered significant. For a–d, g, source data provided in Source Data file.

Dataset 4a). Of these, z.A.E2 and z.A.E4 were validated as pancreatic enhancers. z.A.E4 was the most robust pancreatic enhancer of this set (Fig. 4a and Supplementary Dataset 4a), driving expression in endocrine, acinar and duct cells of the zebrafish pancreas (Fig. 4b and Supplementary Fig. 5a) and interacting with the promoter of aridlab (Fig. 4a and Supplementary Fig. 5b). Additionally, we detected a human/zebrafish syntenic block containing the zebrafish z.A.E4 enhancer and a human pancreatic CRE (h.A.E4) (Fig. 4a). In vivo enhancer assays for h.A.E4 demonstrated its ability to drive expression in endocrine cells of the zebrafish pancreas, and in vitro in a human pancreatic duct cell line (hTERT-HPNE), suggesting a functional equivalency to the zebrafish z.A.E4 enhancer (Fig. 4b, c and Supplementary Fig. 5a). To study the influence of this human enhancer on ARID1A expression, we deleted the h.A.E4 enhancer in the hTERT-HPNE cell line, relevant for the pancreatic tumour suppressor role of ARID1A, through CRISPR-Cas9 system (Fig. 4d and Supplementary Fig. 5c–e), using a deletion in an unrelated genomic region16 as a control. We observed lower levels of ARID1A upon deletion of h.A.E4 compared to the control (Fig. 4e, f and Supplementary Fig. 5e), suggesting that the loss of this enhancer may interfere with the DNA-damage response, with possible implications in the increased risk for pancreatic cancer68,69.

A ptf1a enhancer explains pancreatic agenesis causal variant in vivo. To further evaluate the interspecies functional equivalency of enhancers and their role in human pancreatic diseases, we focused on the human PTF1A locus, known to be controlled by a distal downstream enhancer whose deletion causes pancreatic agenesis35 (Fig. 5a; h.P.E3). Concomitantly, we detected a zebrafish distal ptf1a enhancer, downstream of ptf1a (z.P.E3), as well as two previously identified proximal enhancers (z.P.E1 and z.P.E270). z.P.E3 interacts with the promoter of ptf1a, observed by Hi-ChIP and 4C-seq (Fig. 5a and Supplementary Fig. 5b), and could correspond to the functional equivalent enhancer whose deletion causes pancreatic agenesis in humans (h.P.E3), although its sequence partially aligns with a more distal human sequence likely inactive in human pancreatic cells (Supplementary Fig. 6). In vivo enhancer assays for z.P.E3 and h.P.E3 showed strong and robust expression in progenitor cells (Fig. 5b), that is in agreement with the described activity of h.P.E3 in vitro as a human developmental enhancer35. These results suggest that the human and zebrafish enhancers share some regulatory information. This is further supported by binding sites for FOXA2 and PDX1 in the human h.P.E3, also predicted to bind to the zebrafish z.P.E3 (Supplementary Fig. 7a, b71). To further evaluate the role of z.P.E3, we generated genomic deletions in the z.P.E3 sequence (Fig. 5c–g, Supplementary Fig. 8 and Fig. 9). Deletion1, a 632 bp deletion that includes the predicted Foxa2 and Pdx1 binding sites and the majority of transposase-accessible chromatin within z.P.E3 (Supplementary Fig. 9a), results in a decrease of the pancreatic progenitor domain area in homozygous mutants (Fig. 5c, d, f), as well as a reduction in the expression levels of ptf1a (Supplementary Fig. 9b). Furthermore, after pancreatic differentiation, the Deletion1 mutants displayed pancreatic hypoplasia (Fig. 5e, g; Supplementary Fig. 9c–e), and we observed the same phenotype for multiple independent deletions of z.P.E3 generated in somatic cells (Supplementary Fig. 8). In contrast, no phenotypes were observed for a 517 bp deletion within the z.P.E3 enhancer, adjacent to Deletion1, which excludes the majority of accessible chromatin and predicted TF binding sites (Deletion2; Supplementary Fig. 9a, d, e), suggesting that the functional core of z.P.E3 coincides with the regions of available chromatin that overlap with the predicted binding of Foxa2 and Pdx1. In agreement with the observed phenotypes, pancreatic hypoplasia is compatible with the described loss-of-function of ptf1a in zebrafish70 and the loss of h.P.E3 function in humans35. In light of these results, we suggest that pancreatic hypoplasia is the consequence of the reduction in the pancreatic progenitor domain caused by decreased levels of ptf1a due to the loss of an important pancreatic progenitor enhancer.

Later on, after pancreatic differentiation, z.P.E3 and h.P.E3 enhancers acquire distinct activity patterns. The zebrafish z.P.E3 enhancer is able to drive a consistent expression in differentiated pancreatic cells from late embryos up to adults (Supplementary Fig. 10), including acinar and duct cells, while the human h.P.E3 enhancer shows almost a total lack of activity in differentiated acinar and duct cells, as previously observed in vitro35 driving expression only in very few cells (Supplementary Fig. 10). Overall, these results suggest that zebrafish and humans share a functionally equivalent distal enhancer of PTF1A during development, whose loss-of-function results in a reduction of the pancreatic progenitor domain, elucidating, in vivo, the causal link between the disruption of this enhancer in humans and pancreatic agenesis.

Discussion
Cis-regulatory mutations and sequence variations are associated with pancreatic cancer and diabetes2–6. However, the in vivo implications of these genetic changes are still unknown. Here, we explore the chromatin state of the zebrafish pancreas to uncover pancreatic enhancers and establish comparisons with humans,
so that we can predict and model human pancreas disease-associated enhancers. We found that, although most of the zebrafish pancreatic enhancers do not share significant sequence identity with human pancreatic enhancers, they share many TFBS and their target genes are enriched for human pancreas diseases. These results suggest the existence of functionally equivalent enhancers in zebrafish and humans, as proposed for other tissues and species\textsuperscript{72,73}. Indeed, recent studies looking into highly divergent species as human and sponges have located similarly functional enhancers within microsyntenic regions that, although do not share significant sequence identity, clearly recapitulate similar expression patterns in enhancer reporter assays, arguing in favour of functional equivalency\textsuperscript{74}. This is likely the consequence of enhancers being fast-evolving sequences operating with a high degree of sequence flexibility\textsuperscript{75}. Several mechanisms that may operate together during evolution can illustrate the potential for sequence flexibility of enhancers while retaining a consistent TFBS code. Among them, nucleotide alterations within
Fig. 4 The zebrafish and human arid1ab/ARID1A regulatory landscapes contain an equivalent pancreatic enhancer. a Genomic landscape of the zebrafish arid1ab gene, showing profiles for H3K27ac ChIP-seq (black), ATAC-seq (blue) and 4C with viewpoint in the arid1ab promoter (magenta) in adult zebrafish pancreas (top); zoom-in in arid1ab regulatory landscape (middle). Human ARID1A genomic landscape (bottom) with H3K27ac enriched intervals from human pancreatic cell lines (HPCL, black bars, top-to-bottom: PT-45-P1, CFPAC-1 and HPAF-II), H3K27ac profile from human pancreas (WPT, black) and from non-pancreatic human cell lines (NPHCL, GM12878, H1-hESC, HSMIM, HUVSEC, K562, NHEK and NHLF, Data from ENCODE). Human/zebrafish sequence conservation (dark green). Tested putative enhancers are highlighted in grey (zA.E1 and zA.E3; no enhancer activity) and green (zA.E2, zA.E4 and hA.E4; enhancer activity). Zebrafish/human syntenic box (red box). b Transient in vivo enhancer reporter assays of zA.E4 and hA.E4 showing the percentage of zebrafish embryos with GFP expression in endocrine, acinar and duct cells (two-sided chi-square test with Yates correction; \( p < 0.05 \); Endocrine cells: zA.E4, \( p = 0.0001 \); hA.E4, \( p = 0.0294 \); Acinar cells: zA.E4, \( p = 0.0391 \); hA.E4, \( p = 0.1167 \); Duct cells: zA.E4, \( p = 0.00001 \); hA.E4, \( p = 0.9731 \)). Number of analysed embryos (n). Negative control (NC). c Luciferase enhancer reporter assays performed in human hTERT-HPNE cells for hA.E4, showing luc2/Nlu ratios, relative to the negative control (two-sided t-test; \( ** * p < 0.0001 \); hA.E4 p-value = 0.0001; PC p-value < 0.0001). Data from three biological replicates (grey dots, mean+SD (error bar). Negative control (NC). Positive control (PC). d Strategy for CRISPR-Cas9 deletions in the hA.E4 locus, indicating sgRNA target sites. e Representative images of transfected hTERT-HPNE human cells expressing pairs of sgRNAs and Cas9 (arrows). In control, sgRNAs target a H3K27ac depleted region, while sgRNAs in sgPair1 and sgPair2 target the hA.E4 locus. Left column show anti-ARID1A (grey) and right column GFP (green), mCherry (red) and DAPI (blue; nuclei). Representative images from three biological replicates. Scale bar: 40 μm. f Normalized ARID1A levels from immunocytochemistry images. Two-sided t-test depicted for \( p < 0.05(\ast) \), \( p < 0.0001(****) \) and not significant (ns; \( p \)-values of: Control vs sgPair1 = 0.0208, Control vs sgPair2 = 0.0044, sgPair1 vs sgPair2 = 0.6227). A black line represents the mean of values. Data from three biological replicates. Data included in Source Data file for b, c, f.
Enhancers can be highly tissue specific, while others can be active in multiple tissues, as observed by the identification of PsE and DevE. The former showed H3K27ac profiles more restricted to the zebrafish adult pancreas, while the latter had broad profiles throughout development, suggesting their activity to be present in multiple tissues. The zP.E3 enhancer is not detected in the embryonic H3K27ac dataset, likely because its activity is highly restricted to pancreatic progenitor cells during development, resulting in its inclusion in the PsE group. A detailed analysis of the activity of this enhancer, from the larval stage to adulthood, shows it to be almost exclusively active in exocrine pancreatic cells (Supplementary Fig. 10e), illustrating the expected tissue specificity of PsE enhancers.
In this work, we identified pancreatic CREs in zebrafish, a model organism that is amenable to genetic manipulation and phenotyping. By establishing a correlation between human and zebrafish pancreatic CREs, functional testing of CREs can be performed in vivo, helping to clarify the role of CREs in pancreatic function and disease. In summary, the integration of human cis-regulatory elements involved in disease. We show that transcriptional cis-regulation of the human and zebrafish adult pancreas have a high degree of similarity, allowing the functional exploration of cis-regulatory sequences in zebrafish, with the potential of translation to human pancreatic diseases.

**Methods**

**Experimental procedures**

Zebrafish stocks, husbandry, breeding and embryo rearing. Adult zebrafish AB/TU WT strains where obtained from the Gomez-Skarmeta's laboratory in Seville (CABD). WT, transgenic and mutant lines were maintained at 26–28 °C under a 10 h dark/14 h light cycle in a recirculating housing system according to standard protocols.

Embryos were grown at 28 °C in E3 medium (5 mM NaCl (#S/3161/60, Fisher Chemical), 0.17 mM KCl (#2667.298, WVR), 0.33 mM CaCl2·2H2O (#C3881, Sigma-Aldrich), 0.33 mM MgSO4·7H2O (#61340, Sigma-Aldrich) and 0.01% methylene blue (#66120, Sigma-Aldrich), pH 7.2) or E3 supplemented with 0.01% PTU (1–phenyl-2-thiourea, #P7629, Sigma-Aldrich) for the in vivo enhancer assays, embryos were anesthetized by adding tricaine (MS222; ethyl-3-aminobenzoate methanesulfonate, #E10521-10G, Sigma-Aldrich) to the medium and selected by the internal positive control of transcription. For the establishment of transgenic and mutant zebrafish lines, embryos were microinjected, selected, bleached and grown until adulthood. Adult F0s were outcrossed with WT adults and the offspring screened for the internal control of transgenesis and the pattern of expression of the regulatory element, or for the respective mutations, by genotyping. In vivo reporter lines, Tg(smcherry) and Tg(stsmcherry), were used to label the exocrine and endocrine domain, respectively. The i59 animal facility and this project were licensed by Dirección General de Alimentación y Veterinaria (DGA V) and all the protocols used for the experiments were approved by the i59 Animal Welfare and Ethics Review Body.

**Cell culture.** hTERT-HPNE (ATCC CRL-4023) cells were cultured in a 5% CO2 humidified chamber at 37 °C in DMEM (1x, 4.5 g/L D-glucose with pyruvate; #D6429, Gibco, ThermoFisher Scientific), supplemented with 10% fetal bovine serum (#BSCS0615, biotecnomica), 10 ng/mL human recombinant EGF (#11343406, PD), 750 mg/mL puromycin (#P8833-25MG, Sigma-Aldrich) in TC Dish 100 (SARSTEDT). When cells reached 90% of confluence, they were split using TrypLE Express (#12604-021, Gibco, ThermoFisher Scientific; ~0.5 mL per 10 cm²).

**HiC-seq.** Whole pancreas was dissected from 25 adult zebrafish (~50 × 10⁶ cells; both genders and with 12–24 months), kept on ice in PBS (137 mM NaCl (#S/3161/60, Fisher Chemical), 2.7 mM KCl (#2667.298, WVR), 10 mM NaHPO4 (#1.06342.0250, Merck), and 1.8 mM KH2PO4 (#1.06585.1000, Merck)) with 1x Complete Protease Inhibitor (#1169749801, Roche) and 1x Complete Proteinase Inhibitor (1169749801, Roche). Chromatin was sheared with a Bioruptor Plus (Diagenode) with the following cycling conditions: 10 min high–30 s on, 30 s off; 10 min on; 10 min high–30 s, 30 s off. The sonicated chromatin had a size in the range of 10 to 1000 bp and was incubated overnight at 4 °C with the anti-HJK3ac636 antibody (1:2, #ab4729, Abcam). Samples were incubated for 1 h at 4 °C with Dynabeads Protein G for Immunoprecipitation (#10035D, Invitrogen, ThermoFisher Scientific). Final DNA was purified with MinElute (#28004, Qiagen) and sequenced on Illumina HiSeq 2000 platform.

**HiChIP-seq.** HiChIP-seq was performed as previously described, with minor alterations. Whole pancreas was dissected from 24 to 9 days post-fertilization (dpf) Embryos were grown at 48 hpf. Unpaired student’s t-test (two-tailed), p-values < 0.05 were considered significant (*p < 0.01, **p = 0.0002). Representative confocal images (maximum intensity projections) of the pancreatic progenitor domain (yellow dashed line) of zP.E3/wt (n = 6) and zP.E3/–/– siblings embryos (n = 5) at 48 hpf. Nuclei are stained with DAPI. Scale bar: 25 µm. Epifluorescence live images of representative phenotypes quantified in e59. Scale bar: 250 µm. elastase, st somatostatin. For d, e, source data are provided as a Source Data file.

**ARTICLE**

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-29551-7 | www.nature.com/naturecommunications

**4C-seq.** 4C-seq was performed as previously described, with minor alterations. Whole pancreas was dissected from 6 to 12 adult zebrafish (7–15 × 10⁶ cells; both genders and with 12–24 months), kept on ice in PBS (137 mM NaCl (#S/3161/60, Fisher Chemical), 2.7 mM KCl (#2667.298, WVR), 10 mM NaHPO4 (#1.06342.0250, Merck), and 1.8 mM KH2PO4 (#1.06585.1000, Merck) with 1x Complete Protease Inhibitor (1169749801, Roche), fixed in 2% formaldehyde (#F8355-500ML, Sigma-Aldrich) for 10 min, and stored at ~80 °C. Cell lysis was performed on ice, with a 15 mL Tenbroek Homogenizer, not exceeding 10 min. Lysates were performed with 60 U T4 DNA Ligase (#HE0012, ThermoFisher Scientific). The restriction enzymes used were DpnI (#05435M, NEB) and Csp6I (#ER0211, ThermoFisher Scientific) for the first and second cuts, respectively. Chromatin was purified by Amicon Ultra 15 Centrifugal Filter Device (#UFC901024, Milipore). 4 C libraries were prepared for Illumina sequencing by the Expand Long Template Polymerase (#1175960001, Roche) with primers targeting the TSSs of each gene and including Illumina adapters (Supplementary Dataset 4c). Final PCR products were purified with the High Pure PCR Product Purification Kit (#1179682001, Roche) and Ampure XP PCR purification kit (#B37419AB, Agencourt Ampure XP).

**HiChIP-seq.** HiChIP-seq was performed as previously described, with minor alterations. Whole pancreas, from both genders and with 12–24 months, was dissected, fixed in 1% formaldehyde (#F8355-500ML, Sigma-Aldrich) and cells lysed as described for 4C-seq. Immediately after lysis, samples were washed with HiChIP Wash Buffer (Tris-HCl pH 8.50 mL (Tris Base #MB11601, NZYTech), 1x Complete Proteinase Inhibitor (#1169749801, Roche)). Chromatin was sheared with a Bioruptor Plus (Diagenode) with the following cycling conditions: 10 min high–30 s on, 30 s off; 15 min on ice, to obtain a size in the range of 100–500 bp. Samples were incubated with anti-HJK3ac636 antibody (1:5, #AB8580, Abcam) and Dynabeads Protein G for Immunoprecipitation (#10035D, Invitrogen, ThermoFisher Scientific) and purified with DNA Clean and Concentrator columns (#D4004, Zymo Research). Up to 150 ng of the DNA was then biotinylated with Streptavidin–C1 beads (#65001, ThermoFisher Scientific). Tagmentation was performed using Nextera DNA Library Preparation Kit (#FC-121-1030, Illumina). Libraries were amplified using NEBNext High-Fidelity 2X PCR Master Mix (#M0541S, NEB).
with primers Ad1, Ad2.3 and Ad2.434. The final product was purified with DNA Clean and Concentrator kit (#D4004, Zymo Research).

**Generation of plasmids for enhancer assays.** Putative enhancer sequences were selected based on the overlap between H3K27Ac ChIP-seq and ATAC-seq signal in non-coding regions within the landscape of each pancreas-relevant gene. Sequences were PCR amplified from zebrafish genomic DNA using the primers in Supplementary Dataset 4b (designed to span the ChIP-seq and ATAC-seq signals) (Sigma-Aldrich), with the proof-reading iMaxTM II DNA polymerase (#25261, INRION Biotechnology) following the manufacturer’s instructions for a standard 20 μl PCR reaction. PCR products were visualized by electrophoresis on an 1% agarose gel, the bands excised, purified with NZYGelpure kit (#MB011, NZYTech) and cloned into the entry vector pCR®8/GW/TOPO (#25020 Invitrogen, ThermoFisher Scientific) according to manufacturer’s instructions. The vectors were then recombined into the destination vectors Z4890, for transient enhancer assays, and ZED91, for stable transgenic lines, using Gateway® LR Clonase II Enzyme mix (#11791020, Invitrogen, ThermoFisher Scientific), following manufacturer’s instructions.

**Standard chemical transformation was performed with MultiShotTM FlexPlate Mach1TM TR1 (#C6861201, Invitrogen, ThermoFisher Scientific), grown O.N. at 37 °C. Vector selection was performed with 100 μg/ml Spectinomycin (#54014, Sigma-Aldrich) in the growth medium for the pCR®8/GW/TOPO vectors, or 100 μg/ml Ampicillin (#24619.1, Normom) for the Z48 and ZED vectors. Plasmids were purified with the Wizard® Plus PCR Prep Kit (Promega) and converted into linear cDNA using Sanger sequencing using the primers in Supplementary Dataset 4b. Final plasmids were purified with phenol/chloroform (#A931S100 and #C/4920/15, Fisher Chemical) and concentration was determined by NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific).

**In vitro mRNA synthesis, microinjection and transgenesis.** Z48 and ZED zebrafish lines were generated through TOL2-mediated transgenesis93. TOL2 cDNA was provided by D. Hackeng (Addgene) (#252561, INRION Biotechnology) following the manufacturer’s instructions for a standard 20 μl PCR reaction. PCR products were visualized by electrophoresis on an 1% agarose gel, the bands excised, purified with NZYGelpure kit (#MB011, NZYTech) and cloned into the entry vector pCR®8/GW/TOPO (#25020 Invitrogen, ThermoFisher Scientific) according to manufacturer’s instructions. The vectors were then recombined into the destination vectors Z4890, for transient enhancer assays, and ZED91, for stable transgenic lines, using Gateway® LR Clonase II Enzyme mix (#11791020, Invitrogen, ThermoFisher Scientific), following manufacturer’s instructions.

**Nucleic acid extraction from zebrafish and human cell lines.** Genomic DNA was extracted from whole zebrafish embryos at 24 hpf, after removal of the chorion, with a standard phenol–chloroform DNA extraction (#A931S100 and #C/4920/15, Fisher Chemical), and used as template for PCR amplification in order to genotype the tested conditions (Supplementary Dataset 4b). The DNA samples were resuspended in 20 μl of TE buffer with RNase (10 μml Tris, pH 8.0 (Tris Base #EP0152, Fisher bioreagents, HCl #205325.290, VWR); 1 mM EDTA pH 8.0 (#EN0521, VWR) and 100 μg/ml RNase (#1019142011, Sigma-Aldrich)) for 1 h at 37 °C, and stored at –20 °C.

Genomic DNA from hTERT-HPNE cells was extracted after transfection and used as template for PCR amplification in order to genotype the tested conditions (Supplementary Dataset 4b). RNA was extracted from zebrafish embryos, pancreas and muscle, with 500 μl TRIzol (#15960626, Invitrogen, ThermoScientific), following the manufacturer’s instructions. Samples were incubated 30 min at 37 °C with 1 μl DNAse I (#EN0521, ThermoFisher), 1 μl 10x reaction buffer and 0.5 μl NZY Ribonuclease Inhibitor (40U/μl; # MB084, NZYTech) at 0.05 μl final concentration. After adding 1 μl EDTA (#2031201901, VWR) 50 μM per l of estimated RNA, final volume was completed to 60 μl with H2O, phenol–chloroform (#A931S100 and #C/4920/15, Fisher Chemical) standard purification was performed and the RNA stored at 80 °C.

**Zebrafish pancreatic progenitor cells were extracted from 48 hpf embryos, immediately following euthanasia by rapid chilling, by repeated pipetting up and down in a gentle motion with 300 μl of Ginzburg fish Ringer’s solution (35 mM NaCl (#53161/60, Fisher Chemical), 1.8 mM KCl (#2676/298, VWR), 1.25 mM CaCl2 (#80393/600, Sigma-Aldrich)). ZEBrafish embryos were allowed to settle to the bottom and the suspension containing the detached pancreatic progenitor cells and yolk was collected, washed with PBS (#137 mM NaCl (53161/60, Fisher Chemical), 2.7 mM KCl (#2676/298, VWR), 10 mM NaHPO4 (#1.06342.0250, Merk), and 1.8 mM KH2PO4 (#1.06355.1000, Merk)), and RNA was extracted using Quick-RNA Micro Kit (#RI0150, Zymo Research), according to manufacturer’s instructions. For real-time qPCR, RNA samples were treated with DNasel (#EN0521, ThermoScientific) and reverse transcribed using the iScript cDNA Synthesis Kit (#7080980, Bio-Rad) according to the manufacturer’s instructions.

**Immunohistochemistry in zebrafish embryos and human cell lines.** Zebrafish embryos/larvae were euthanized by prolonged immersion in 200–300 mg/l tricaine (MS222; ethyl-3-aminobenzoate methanesulfonate, #E10521-10G, Sigma-Aldrich). After each antibody incubation, embryos were washed 6 times in PBS–1% Triton X-100 (MS222; ethyl-3-aminobenzoate methanesulfonate, #E10521-10G, Sigma-Aldrich) and 1.8 mM KH2PO4 (#1.06355.1000, Merk), and 1.8 mM KH2PO4 (#1.06355.1000, Merk) 5 min at RT. Embryos were permeabilized by incubation with 1% Triton X-100 (#X100, Sigma-Aldrich) in PBS (137 mM NaCl (53161/60, Fisher Chemical), 2.7 mM KCl (#2676/298, VWR), 10 mM NaHPO4 (#1.06342.0250, Merk), and 1.8 mM KH2PO4 (#1.06355.1000, Merk)) 5 min at RT. Embryos were blocked in 5% Glycerol/PBS (#BP229-1, Fisher bioreagents, HCL #016342.0250, Merk) and 1.8 mM KH2PO4 (#1.06355.1000, Merk) for 1 h at RT, followed by blocking with 5% bovine serum albumin (BSA; #MB0460, NZYTech) in 0.1% Triton X-100 (#X100, Sigma-Aldrich) for 1 h at RT. Zebrafish embryos were incubated with the primary antibody diluted in blocking solution at 4 °C O.N. and then incubated with the secondary antibody plus DAPI (1:1000, D1036 Invitrogen, ThermoFisher Scientific) diluted in blocking solution for 4 h at RT. After each antibody incubation, embryos were washed 6 times in PBS–T (0.5 % Triton X-100 (#X100, Sigma-Aldrich) in PBS–1x (#137 mM NaCl (53161/60, Fisher Chemical), 2.7 mM KCl (#2676/298, VWR), 10 mM NaHPO4 (#1.06342.0250, Merk), and 1.8 mM KH2PO4 (#1.06355.1000, Merk)) 5 min at RT. Embryos were mounted in 1:1 Glycera/PBS (#BP229-1, Fisher bioreagents, HCL #016342.0250, Merk) and 1.8 mM KH2PO4 (#1.06355.1000, Merk) 5 min at RT. Embryos were imaged with a Leica TCS SPS II confocal microscope (Leica Microsystems, Germany; LAS AF software (v.2.6.3.8173) and processed by Image software (v.1.8.0). Primary antibodies: rabbit anti-Amylase (Abcam #AB159751, VIG, Vector Lab) mouse anti-Aldh3a1 and mouse anti-Aldh3a2 (1:50, #F55A10, DSHB) and mouse anti-Nkx6.1 (1:50, #F55A10, DSHB). Secondary antibodies: goat anti-mouse AlexaFluor647 (#1800, #A-21236 Invitrogen, ThermoFisher Scientific), goat anti-rabbit AlexaFluor568 (#1800, #A-11036 Invitrogen, ThermoFisher Scientific).
The hTERT-HPNE cells were fixed at 48 h after transfection in formaldehyde 4% (Sigma-Aldrich, St Louis, MO) in PBS (137 mM NaCl (G3#316116, Fisher Chemical), 2.7 mM KCl (d294.298, VWR), 10 mM NaHPO4 (d190.290, Merck), and 1.8 mM KH2PO4 (d0853.100, Merck)) for 15 min at RT, permeabilized with 1% Triton X-100 (dX100, Sigma-Aldrich) in PBS and blocked with 2% BSA (dMSb04602, NZYTech) in PBS for 20 min at RT. Incubation with primary antibody (2% BSA/PBS (dMSb04602, NZYTech) was on O/N at 4°C and in secondary antibody (dDy10, Dianova Biotech, Dianova). DAPI (dD1001, Dianova Biotech, Dianova) for 30 min at RT. For immunofluorescence imaging, the ARID1A nuclear staining was measured for each cell GFP + mCherry + and normalized for the average staining of the nucleus of all other cells in the same field. The expression levels were calculated using the Visiopharm software (Visiopharm, Denmark). Statistical analysis. Two-tailed Student’s -test was applied to data with the null hypothesis that the means of two groups are equal. Student’s -test was applied to the paired data with the null hypothesis that the means of two dependent groups are equal. The significance level was set at P < 0.05. Pearson ’ s correlation coefficient was used to determine the strength of the linear relationship between two variables. The significance level was set at P < 0.05. Multiple comparisons were performed using the Tukey’s -test. The results were considered statistically significant if the P value was less than 0.05. HChiP-seq analysis. High-quality raw reads for the two replicates of H3K27ac HChiP-seq (FASTQ v0.11.5.9) were trimmed for adapter sequences using Skewer (d0.2.19)98. We applied the IRD tool to the human dataset from ENCODE project (https://www.encodeproject.org/). GREAT (v.3.0.0)48,49, using the basal plus extension annotation association rule (proximal: 5 kb upstream, 1 kb downstream, plus distal: up to 10 kb). In addition, two public RNA-seq datasets were used (Supplementary Dataset 4g). 4C-seq analysis. 4C-seq libraries were first inspected for quality control using FASTQC (v.0.11.5). Supplementary Dataset 3-5 and default settings of the 4C-seq alignment tool (d2.2.6) were used to map the clean reads to the genome GRCz10 (d095010, Genoscope). Then, the peaks calling was performed using the Irreproducible Discovery Rate (IDR, v.2.0.4) in order to obtain a confident and reproducible set of peaks. The significance level was set at P < 0.05. The alignment reads were mapped to the human dataset from ENCODE project (https://www.encodeproject.org/). UCSC Genome Browser (Fig.1b). The significant peaks were obtained using GREAT (v.3.0.0)48,49, using the basal plus extension annotation association rule (proximal: 5 kb upstream, 1 kb downstream, plus distal: up to 10 kb). The significance level was set at P < 0.05. The alignment reads were filtered using the script “demultiplex.py” from the FourSeq package107, allowing for 1 mismatch in the primer sequence. The filtered reads were mapped back to the genome using the Bowtie2 tool. The mapped reads were converted to reads-per-first-enzyme-frame-end units, and smoothed using a 30 fragment mean window algorithm (Figs. 4a and 5a).
average expression of all genes present in the RNA-seq datasets using R and ggplot for drawing barplots (Fig. 2a, Supplementary Fig. 2c, Supplementary Dataset 3h, Fig. 2a R in https://gitlab.com/rdcamel/pancreasregulome). Identification of Human/zebrafish syntenic blocks. Human/zebrafish syntenic blocks were defined by two aligned regions between both species that kept their relative position among each other. Pre-existing alignments available in the UCSC genome browser were used. Then, enhancers were searched within these blocks in both species.

Conservation between zebrafish and human and PhastCons scores. To obtain the percentage of zebrafish putative active enhancers conserved with human, the coordinates of putative active enhancers from adult zebrafish pancreas and embryos at different development stages (GR10-10danRer10) were used as input to the UCSC genome alignment conversion tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver, liftover (v.1.04.00) to hg19, October 2019) (Fig. 3a). To visualise the conservation of the respective sequences, liftover (v.1.04.00) to hg38 was done and their average PhastCons conservation score plotted (Fig. 3b).

For this, we downloaded PhastCons scores in bigWig format from a 100-way multiple species alignment of 99 vertebrates against human (hg38) (hg38.phastCons100way.hw, October 2019)116 and converted to BedGraph text format using the UCSC’s utility bigWigToBedGraph (v.1.04.00). Then, the BedTools99 suite (v.2.27) was used to intersect and map different putative enhancer clusters in bed format with the conservation scores, storing for each putative enhancer the median and average PhastCons score. To know which of them overlap putative active enhancers in human pancreas, we used the BedTools “intersect” tool with default bp of overlap (Fig. 3b, blue). To calculate the Fold Change (FC) of the graph displayed in Fig. 3c, we have quantified the number of zebrafish H3K27ac positive sequences aligned with the human genome that also showed H3K27ac signal in human pancreas. As a control, we have performed a similar analysis, randomizing the aligned human sequences, quantifying the number of those that also showed H3K27ac signal in human pancreas, repeating this operation 100 times (randomlyH3K27). FC was calculated by the ratio: ZebraHumanK27/average(randomZebraHumanK27) (Supplementary Dataset 3q). This was performed for the different populations of zebrafish enhancers (Pancreas, PSE, DEV, and embryo).

**Transcription factor binding motifs enrichment.** To refine our data, H3K27ac peaks were filtered with the ATAC-seq peaks. Then, the transcription factor binding site (TFBS) predictor program Hypergeometric Optimization of Motif Enrichment (HOMER v.4.11.1) was used to identify conserved sequence motifs enriched117. To evaluate the results, we filtered, using HOMER hypergeometric testing of motifs of interest, different data obtained from: human pancreas, human ventricle, zebra genes homologous to the human disease, and zebra genes homologous to the human disease Ai and in enhancer set Bi; M: number of genes in disease Ai, N: non-disease genes, k: number of shuffled datasets. The analysed data are available on Zenodo (https://doi.org/10.5281/zenodo.6340878).

**References**

1. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosomal topology. Science 361, 1341–1345 (2018).
2. Pasquali, L. et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat. Genet. 46, 136–143 (2014).
3. Klein, A. P. et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat. Commun. 9, 556 (2018).
4. Wolpin, B. M. et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. 46, 994–1000 (2014).
5. Mahajan, A. et al. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat. Genet. 50, 1505–1513 (2018).
6. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 44, 981–990 (2012).
7. Lippi, G. & Mattiuzzi, C. The global burden of pancreatic cancer. Arch. Med. Sci. 16, 820–824 (2020).
8. GBD. Pancreatic Cancer Collaborators. The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol. Hepatol. 4, 934–947 (2019).
9. Huang, J. et al. Worldwide Burden of, Risk Factors for, and Trends in Pancreatic Cancer. Gastroenterology 160, 744–751 (2021).
10. Saeedi, P. et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin. Pr. 157, 107843 (2019).
11. Liscar, N. et al. Type 2 diabetes in adolescents and young adults. Lancet Diabetes Endocrinol. 6, 69–80 (2018).

**Data availability**

All raw sequencing data generated within this study has been submitted to ENA under accession number PRJEB40292. The analysed data are available on USCS browser [http://genome-euro.ucsc.edu/s/VDR_group_public_data/Carrico_et_al_2020_ZebrafishPancreasRegulome] and in Supplementary material.

Other datasets used in this study can be downloaded from ENCODE project (https://www.encodeproject.org/)2: CHIP-seq and ATAC-seq of Human pancreas “ENCSR340GAZ”, ChIP-seq and ATAC-seq of left ventricle “ENCST464TPP”; from Expression Atlas: data (http://www.ebi.ac.uk/gxa/experiments/); RNA-seq of developmental stages of zebrafish “GSE34283”, European Nucleotide Archive (ENA) browser (https://www.ebi.ac.uk/ena); RNA-seq of the pancreatic acinar, alpha, beta and delta cells from zebrafish “PRJEB10410”, RNA-seq of developmental stages of zebrafish “PRJEB12296”,” PRJEB2744,” “PRJEB2988”, ChIP-seq from the DANNIO-CODE consortium to create the blacklist were following (https://dannio-code.zfin.org/ “DCD002894SQ”, “DCD002911SQ”, “DCD0036565SQQ”, “DCD0036564Q”, “DCD0036714Q” and “DCD0002742SQ”). All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request.

**Code availability**

The custom code for analysis of optical action potential traces is available in github (https://gitlab.com/rdcamel/pancreasregulome)118 and in Zenodo (https://doi.org/10.5281/zenodo.6340878).
12. Sinclair, A. et al. Diabetes and global ageing among 65-99-year-old adults: Findings from the International Diabetes Federation Diabetes Atlas, 9th edn. (International Diabetes Federation, 2019).

13. Parker, S. C. J. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. USA 110, 17921–17926 (2013).

14. Khetan, S. et al. Type 2 Diabetes-Associated Genetic Variants Regulate Chromatin Accessibility in Human Islets. Diabetes 67, 2466–2477 (2018).

15. Greenwald, W. W. et al. Pancreatic islet chromatin accessibility and conformation reveals distant enhancer networks of type 2 diabetes risk. Nat. Commun. 10, 2078 (2019).

16. Miguel-Escalada, I. et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. Nat. Genet. 51, 1137–1148 (2019).

17. Gaulton, K. J. et al. A map of open chromatin in human pancreatic islets. Nat. Genet. 42, 255–259 (2010).

18. Roman, T. S. et al. A Type 2 Diabetes-Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the ADCY5 Locus. Diabetes 66, 2521–2532 (2017).

19. Khetan, S. et al. A Common Type 2 Diabetes Risk Variant Potentiates Activity of an Evolutionarily Conserved Islet Stretch Enhancer and Increases C2C4DA and C2C4DB Expression. Am. J. Hum. Genet. 102, 620–635 (2018).

20. Eufrásio, A. et al. In Vivo Reporter Assays Uncover Changes in Enhancer Activity Caused by Type 2 Diabetes-Associated Single Nucleotide Polymorphisms. Diabetes 69, 2794–2805 (2020).

21. Fujitani, Y. et al. Targeted deletion of a cis-regulatory region reveals differential gene dosage requirements for Pdx1 in foregut organ development and pancreas formation. Genes Dev. 20, 253–266 (2006).

22. van Arensbergen, J. et al. A distal intergenic region controls pancreatic endocrine differentiation by acting as a transcriptional enhancer and as a chromatin regulatory protein. J. Clin. Invest. 132, 1521–1531 (2013).

23. Akerman, I. et al. Neonatal diabetes mutations disrupt a chromatin pioneering function that activates the human insulin gene. Cell Rep. 35, 108981 (2021).

24. Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227–229 (2013).

25. Kinkel, M. D. & Prince, V. E. On the diabetic mcin: zebrafish as a model for pancreas development and function. Bioessays 31, 139–152 (2009).

26. Prince, V. E., Anderson, R. M. & Dalgin, G. Zebrafish Pancreas Development and Regeneration: Fishing for Diabetes Therapies. Curr. Top. Dev. Biol. 124, 235–276 (2017).

27. Elgar, G. & Vavouri, T. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. Trends Genet. 24, 344–352 (2008).

28. Prescott, S. L. et al. Enhancer divergence and cis-regulatory evolution in the human and chimpanzee neural tube. Cell 163, 68–83 (2015).

29. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858 (2009).

30. modEncode Consortium, et al. Identification of functional elements and regulatory circuitry across multiple species. Science 330, 1787–1797 (2010).

31. Wittkopp, P. J. & Kalay, G. cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. 13, 59–69 (2011).

32. Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L. & McCallion, A. S. Conservation of REG regulatory function from human to zebrafish without sequence similarity. Science 312, 276–279 (2006).

33. Jones, S. et al. Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. Hum. Mutat. 33, 100–103 (2012).

34. Wu, J. N. & Roberts, C. W. M. ARID1A mutations in cancer: another epigenetic tumor suppressor? Cancer Discov. 3, 35–43 (2013).

35. Weendon, M. N. et al. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat. Genet. 46, 61–64 (2014).

36. Gabbay, M., Ellard, S., De Franco, E. & Moisés, R. S. Pancreatic Agenesis due to Homozygous PTF1A Enhancer Mutations and One with Isolated Pancreas Agenesis. Sci. Rep. 5, 108078 (2020).

37. Shen, J. et al. ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. Cancer Discov. 5, 752–767 (2015).

38. Wang, S. C. et al. SWI/SNF component ARID1A restrains pancreatic neoplasia formation. Gut 68, 1259–1270 (2019).

39. Wang, W. et al. ARID1A, a SWI/SNF subunit, is critical to acinar cell homeostasis and regeneration and is a barrier to transformation and epithelial-mesenchymal transition in the pancreas. Gut 68, 1245–1258 (2019).

40. Pashos, E., Park, J. T., Leach, S. & Fisher, S. Distinct enhancers of ptf1a mediate specification and expansion of ventral pancreas in zebrafish. Dev. Biol. 381, 471–481 (2013).

41. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 46, D1284 (2018).

42. Khouei, P. et al. Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. eLife 6, e28440 (2017).

43. Wang, S. et al. Functionally conserved enhancers with divergent sequences in distant vertebrates. BMC Genomics 16, 882 (2015).

44. Wong, E. S. et al. Deep conservation of the enhancer regulatory code in animals. Science 370, eaa8137 (2020).
75. Snetkova, V. et al. Ultraconserved enhancer function does not require perfect sequence conservation. Nat. Genet. 53, 521–528 (2021).
76. Deplancke, B., Alpern, D. & Gardeux, V. The Genetics of Transcription Factor DNA Binding Variation. Cell 166, 538–554 (2016).
77. Arnott, D. N. & Kulikarni, M. M. Transcriptional enhancers: Intelligible enhancements or flexible billboards? J. Cell Biochem. 94, 890–898 (2005).
78. Bessa, J. et al. A mobile insulator system to detect and disrupt cis-regulatory elements required for macrophage and B cell differentiation to computation. Nucleic Acids Res. 113, E2617–E2626 (2016).
79. Eichenlaub, M. P. & Ettwiller, L. De novo genesis of enhancers in vertebrates. PLoS Biol. 9, e1001188 (2011).
80. Jin, K. & Xiang, M. Transcription factor Ptf1a in development, diseases and reprogramming. Cell Mol. Life Sci. 76, 921–940 (2019).
81. Kwon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development and disease. Nat. Rev. Genet. 22, 334–346 (2021).
82. Ariza-Cosano, A. et al. Differences in enhancer activity in mouse and zebrafish reporter assays are often associated with changes in gene expression. BMC Genomics 13, 713 (2012).
83. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science 346, 1007–1012 (2014).
84. Cooper, G. M. & Brown, C. D. Qualifying the relationship between sequence identities.
85. Pennacchio, L. A. & Visel, A. Limits of sequence and functional conservation.
86. Li, H. et al. The Sequence Alignment/Map format and SAMtools.
87. Li, Q., Brown, J. R., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1779 (2011).
88. Klein, F. A. et al. FourCseq: analysis of 4C sequencing data. Bioinformatics 31, 3085–3091 (2015).
89. Noordermeer, D. et al. The dynamic architecture of Hox gene clusters. Science 334, 222–225 (2011).
90. Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58, 221–230 (2012).
91. Emura, D., Yin, J., Reilly, S. K., Gockley, J. & Noonan, J. P. Origin and evolution of developmental enhancers in the mammalian neocortex. Proc. Natl Acad. Sci. USA 113, E2617–E2626 (2016).
92. Serra, F. et al. Automatic analysis and 3D-modeling of Hi-C data using TADbit reveals structural features of the fly chromatin maps. PLoS Comput. Biol. 13, e1005655 (2017).
93. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat. Methods 9, 1185–1188 (2012).
94. Gordon, A. H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
95. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
96. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050 (2005).
97. MacDonald, P. W., Liang, K. & Janss, A. Dynamic adaptive procedures that control the false discovery rate. Electron. J. Stat. 13, 3009–3024 (2019).
98. Bordeira-Carriço, R. et al. Gibson/Zeno, ed. 10.5281/zenodo.6340878 (2022).
